Engelbert Buxbaum

Fundamentals of Protein Structure and Function

Second Edition

Springer

Fundamentals of Protein Structure and Function

Engelbert Buxbaum

Fundamentals of Protein Structure and Function

Second Edition

Springer

Engelbert Buxbaum
Kevelaer, Germany

ISBN 978-3-319-19919-1 ISBN 978-3-319-19920-7 (eBook)
DOI 10.1007/978-3-319-19920-7

Library of Congress Control Number: 2015946608

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Preface to the First Edition

. . . to describe the things that are, as they are.
(FRIEDRICH II. (1164–1250): On the Art to Hunt with Birds)

In those simple words FRIEDRICH, emperor of the Holy Roman Empire of German Nation, around 1240 described the job of a scientist. He was an avid hunter who particularly loved hunting with birds of prey. He realised that to really enjoy that sport one had to know as much as possible about animals, a knowledge that can be derived only by careful, unbiased observation. Thus he broke with the antique tradition of mixing observation with philosophy, religion, and superstition, and wrote the first truly scientific book in history.

Today science is all-pervading in our lives. We live about twice as long as people did in the days of FRIEDRICH and our lives are much more comfortable, as a result of scientific research and its technical applications.

Proteins are the main actors in living organisms. Although nucleic acids store the information required to make functional proteins, proteins do almost all the real work. So it is not surprising that after several years of emphasis in nucleic acid biochemistry, culminating in the human genome project, attention is now returning to proteins. Modern methods, in particular protein crystallography, have confirmed that "nature is greatest in the smallest things". So apart from its utility, which makes training in science a necessity, understanding nature is, today just as much as in FRIEDRICH's days, also a source of joy and satisfaction. I wish every reader of this book a share in it.

How to Read This Book

I did not intend to write another biochemistry or cell biology textbook, as there are many on the market already. Rather, I wanted to focus on the fascinating world of proteins.

The four parts of this book largely follow the curriculum for biochemistry training of undergraduate medical and science students. The parts on protein structure and enzymology can be covered at the beginning of the course, the special proteins and membrane transport at the end, once molecular biology has been studied.

MEDICAL MATERIAL:
Medical applications are marked.

ADVANCED MATERIAL:
I have added some more advanced material for science students, which could form the basis for seminars, term papers, and the like. This material is not exam-relevant for medical students, but may broaden your view on the more basic topics.

Each chapter contains some example questions, which resemble those that might be asked in exams. The answers are provided together with a short explanation. Special emphasis was put on multiple-choice questions. One might question the didactic value of such questions, but the fact is that they are now in common use.

Acknowledgments

I wish to thank all my students, friends and colleagues who have given me their support and suggestions for this text, and who have gone through the arduous task of proofreading. All remaining errors are, of course, mine. Please report any errors found to me engelbert_buxbaum@web.de.

This text was created with LaTeX using the MikTeX system (www.ctan.org). Chemical structures were created with ISIS-Draw (now Accelrys-Draw, accelrys.com/products/informatics/cheminformatics/draw/no-fee.php). Gnuplot (www.gnuplot.org) was used for plotting mathematical functions. For graphic handling the Gnu Image Manipulation Package (Gimp, www.gimp.org/) was used. Three-dimensional graphics were rendered with the Persistence Of Vision Ray-tracer (POV-Ray, povray.org). Molecular models were calculated from PDB

coordinates obtained from OCA (bip.weizmann.ac.il/oca-bin/ocamain) using DeepView (formerly SwissPDB, http://spdbv.vital-it.ch/).

A big "thank you" to all who made these programs freely available, or who maintain information services on the Internet. Without your generosity this book would not be.

Portsmouth, Commonwealth of Dominica Engelbert Buxbaum

Preface to the Second Edition

In the last couple of years, considerable progress has been made in protein science, which warrants an update to this book. In particular, we now have a unifying concept of protein folding diseases that I have covered in Sect. 10.2. In addition, we now have high-resolution crystallographic data for many membrane proteins that we could not have dreamt of 10 years ago. This has improved our understanding of their function considerably.

My thanks go to all those who have supported me with this book.

Kevelaer, Germany Engelbert Buxbaum

Contents

Part II Enzymes

Part I
Proteins

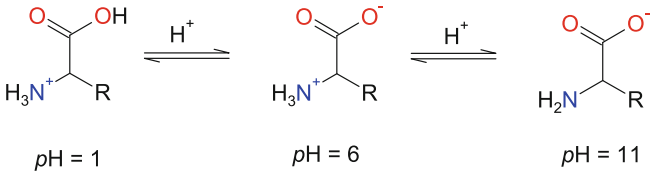
Chapter 1
Amino Acids

Abstract Amino acids are carboxylic acids with an amino-group in the α-position. With the exception of glycine, all amino acids are chiral; usually the L-form is used in living organisms. Twenty-two different amino acids are encoded in genes. Polycondensation of amino acids leads to peptides and proteins. The different side-chains of the various amino acids have different physicochemical properties and allow these amino acids to fulfil different functions inside a protein.

1.1 Basic Structure of Amino Acids

Amino acids contain a carboxy group, an amino group, a hydrogen atom, and a variable side-chain R ("residue"). The simplest amino acid is glycine, where R is a hydrogen atom. Because the α-carbon of Gly carries only three different ligands it is not enantiomeric. Thus glycine is not chiral, unlike all other amino acids (see Fig. 1.1). Only **L-amino acids** are found in proteins. However, D-amino acids are found in the bacterial cell wall and in several antibiotics. In humans, D-Ser is produced by astrocytes to regulate NMDA-receptor responses to Glu and long-term potentiation.

The carboxy group has a pK_a close to 2 whereas the amino group pK_a ranges from 9 to 10. Thus, amino acids can exist in different protonation states:



At neutral *p*H, amino acids exist as **zwitterions**—molecules that possess both a positive and a negative charge.

Several amino acids contain additional acidic or basic groups in their side-chains. The following table gives pK_a values of the side-chains *in water*, the pK_a *inside a protein* depends on the unique environment created by neighbouring side-chains [1]:

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_1

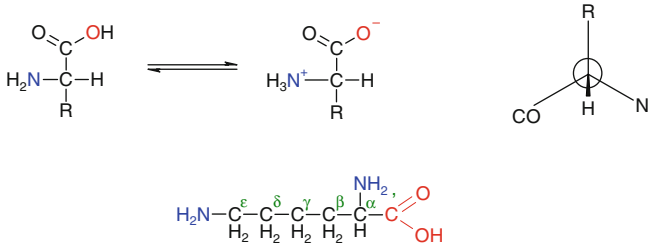


Fig. 1.1 *Top left*: Basic structure of an amino acid. Amino acids can form zwitterions (*Zwitter* (Ger.) = hermaphrodite) with a positive and a negative charge. *Top right*: Because the α-carbon bears four different substituents, it is chiral (exception: glycine where R = H). In L-amino acids if the α-carbon is placed on the paper plane, with the hydrogen facing you, the remaining substituents read "CORN". *Bottom*: Nomenclature of carbon atoms, using lysine as example. The Carboxy-carbon is designated C', the following carbon atoms are labelled with the letters of the Greek alphabet. Sometimes the last C-atom is called ω, irrespective of the chain length

Group	Amino Acid	pK_a in water
Carboxy	Glu, Asp	4.1
Selenol	Sec	5.2
Imidazole	His	6.2
Sulphydryl	Cys	8.3
Phenolic OH	Tyr	10.1
Amino	Lys	10.5
Guanidino	Arg	12.0
Alcoholic OH	Ser, Thr	15.9

You have to be able to recognise the 22 amino acids (see Fig. 1.2). The amino acid side-chains can engage in a number of noncovalent interactions and covalent bonds (Table 1.1):

1.2 The Isoelectric Point

From your chemistry lessons you know how to determine the pK_a of an acid or base, the *p*H at which half of the molecules are charged. A compound like an amino acid, which can act as both acid and base, has another important property: the isoelectric point *p*I, which is the *p*H at which the number of positive charges on the molecule is the same as the number of negative charges. At the *p*I the molecule has no net charge; the molecule's ability to interact with water is lowest, and therefore its solubility.

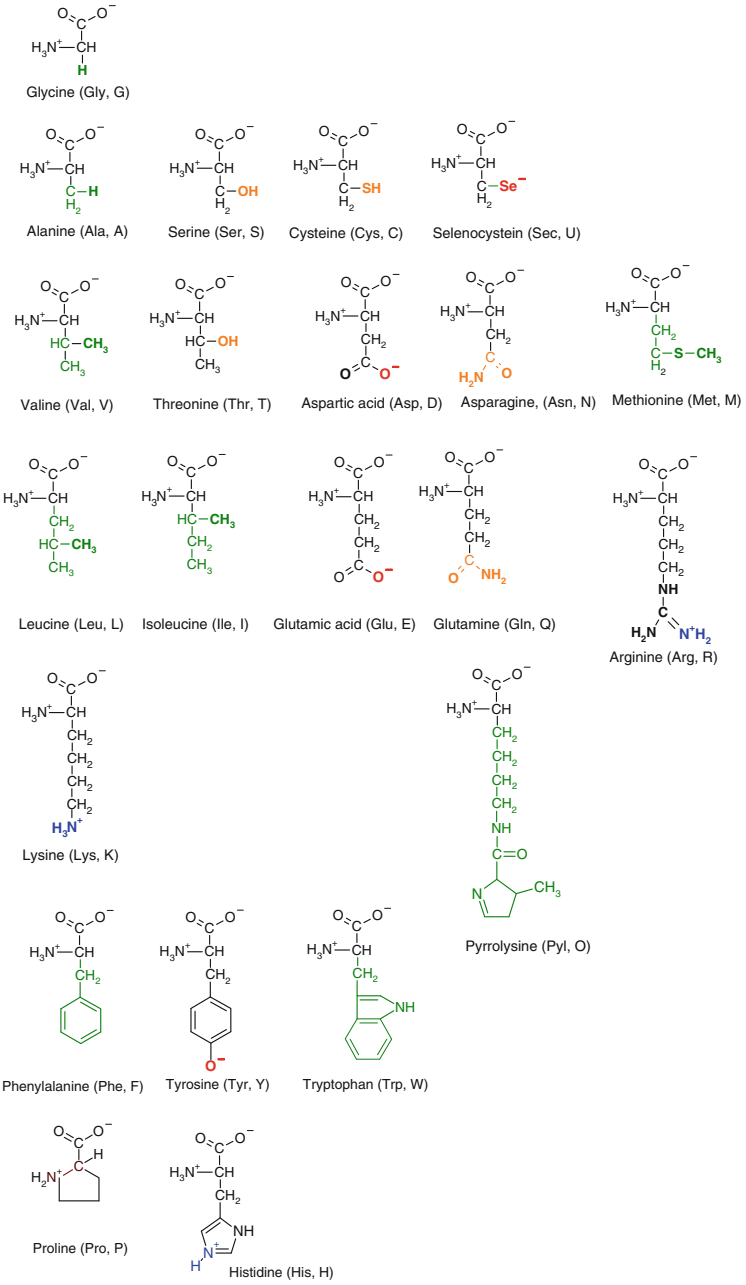


Fig. 1.2 The 22 amino acids encoded by genes. Once incorporated into proteins, amino acids may be further modified. Pyrrolysine has been found only in bacteria; it is encoded by the "amber" stop codon UAG. Selenocysteine is encoded by the UGA "opal" stop-codon (see Fig. D.1 on page 493). Acidic groups are marked *red*, basic groups *blue*, polar groups *orange*, and hydrophobic groups *green*. Note that Thr and Ile have a chiral β- in addition to the α-carbon. Pyl has two chiral carbon atoms in the ring

Table 1.1 Bonds formed by various amino acids. For posttranslational modifications the following abbreviations were used: A = acetyl, F = fatty acid, I = isoprenoid, M = methyl, N = nitrosyl, OH = hydroxyl, P = phosphate, S = sugar, SS = disulphide, U = ubiquitin (-like), Y = AMPylation

Amino Acid	Hydrophobic interaction	Hydrogen bond	Salt bond	Covalent modification
Alanine (Ala, A)	+	–	–	–
Arginine (Arg, R)	+	+++	+++	–
Asparagine (Asn, N)	–	+++	–	S
Aspartate (Asp, D)	–	++	+++	(P), M
Cysteine (Cys, C)	+	(+)	+	SS, M, F, I, N
Glutamine (Gln, Q)	+	++	–	–
Glutamate (Glu, E)	+	++	+++	(P), M
Glycine (Gly, G)	(+)	–	–	F
Histidine (His, H)	–	+++	+	(P)
Lysine (Lys, K)	++	+++	–	OH, A, M, U
Leucine (Leu, L)	+++	–	–	–
Isoleucine (Ile, I)	+++	–	–	–
Methionine (Met, M)	++	–	–	–
Phenylalanine (Phe, F)	+++	–	–	–
Proline (Pro, P)	++	–	–	OH
Pyrrolysine (Pyl, O)	+++	+	–	–
Selenocysteine (Sec, U)	+	++	+	–
Serine (Ser, S)	+	+	–	P, S, F
Threonine (Thr, T)	++	+	–	P, S, Y
Tryptophan (Trp, W)	++	+	–	–
Tyrosine (Tyr, T)	++	+	–	P, Y
Valine (Val, V)	++	–	–	OH

How can we get the isoelectric point? Looking at the titration curve of glycine (see Fig. 1.3, top) we see that below $p\mathrm{K}_1$ most of the molecules bear one positive charge at the amino group; the carboxy group is uncharged.

Above $p\mathrm{K}_2$ it is the other way round: the carboxy group bears a (negative) charge and the amino group is uncharged. Right in the middle between $p\mathrm{K}_1$ and $p\mathrm{K}_2$ there is an inflection point in the titration curve; this is the pI.

Thus we remember: In a chemical that has one acidic and one basic group, pI is the average of the two pK values:

$$p\mathrm{I} = \frac{1}{2} \times (p\mathrm{K}_1 + p\mathrm{K}_2) \tag{1.1}$$

But how do we do it when there are 3 or more ionisable groups?

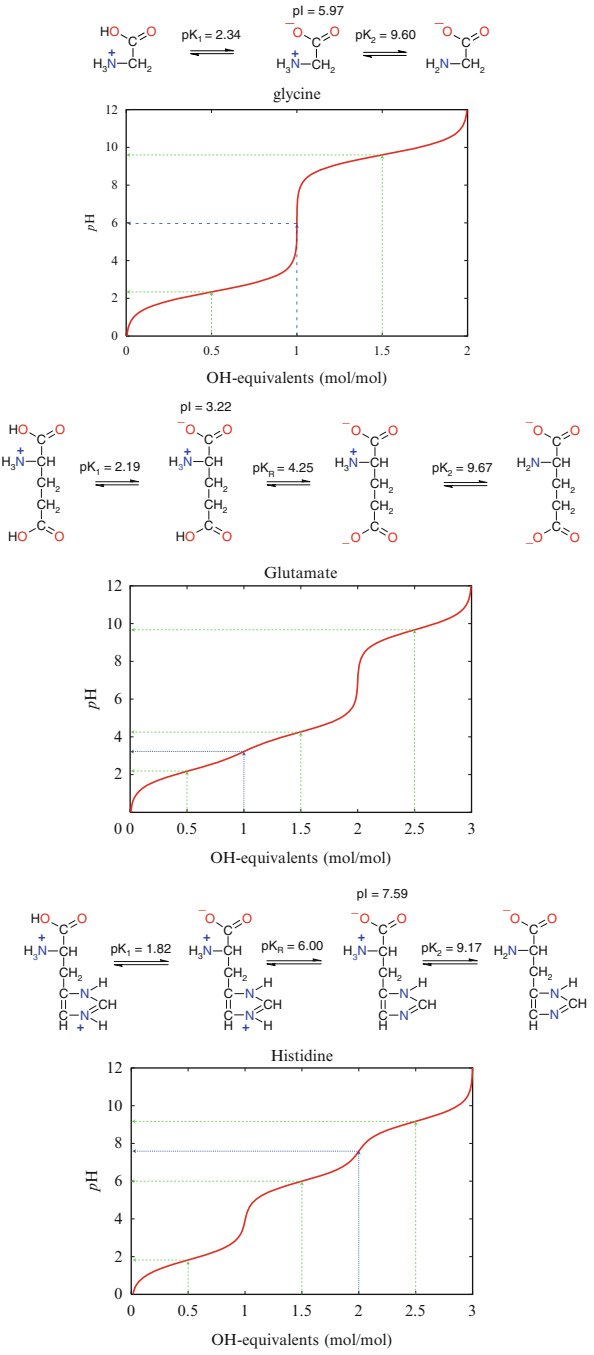


Fig. 1.3 Titration curves of glycine, glutamic acid and histidine. At the isoelectric point amino acids carry an equal number of positive and negative charges, thus they have no net charge. The isoelectric point can be calculated as the average between the pK_a-values on each side of that uncharged form

First we write down the various forms of the molecule, with the corresponding pK_a values between them. Then we count the positive and negative charges on each form and identify the electrically neutral one. The *p*I can be calculated as the average between the pK_a s on either side of it, just as it is done for molecules with only 2 ionisable groups.

Glutamic acid, for example, (see Fig. 1.3, middle) at very low *p*H carries only one positive charge at the amino group (1st form). As the *p*H increases beyond 2.19, more and more protons are lost from the carboxy group on C'. This (2nd) form carries one positive and one negative charge, and is the neutral one. Around *p*H 4.25, a second proton is lost from the terminal carboxy group (3rd form, one positive, two negative charges) and around *p*H 9.67 a further proton is lost from the amino group (4th form, no positive and 2 negative charges). Thus the *p*I is calculated as the average of 2.19 and 4.25, which is 3.22.

The rational with histidine (see Fig. 1.3, bottom) is similar. The electrically neutral form is the 3rd, and the *p*I is the average of 6.00 and 9.17, which is 7.59.

The solubility of amino acids is lowest at the *p*I, as interaction with water is reduced.

This has clinical applications: In the inherited disease cystinuria large amounts of cystine are excreted with the urine. As cystine is poorly soluble at neutral *p*H, this can result in kidney stones. Ensuring a urine *p*H > 8.5—well above the *p*I—makes cystine more soluble and reduces the amount precipitating in the urinary tract.

Solubility of Proteins Is Minimal at the *p*I

Just as for amino acids, the *p*I of a protein is the *p*H at which there are as many positive as negative charges. At this point solubility of the protein is minimal, which can be used to purify a protein by precipitation (see Sect. 3.1.2 on page 66). The behaviour of proteins toward ion exchange columns and during electrophoresis also depends on charge. Because the number of ionic interactions between the amino acids of a protein is smallest at the *p*I, protein stability is reduced.

Naïvely, the *p*I of a protein could be calculated simply from the amino acid composition and the HENDERSON–HASSELBALCH-equation:

$$n_- = \sum_{i=1}^{n} \frac{-1}{1 + 10^{pK_- - pH}} \quad n_+ = \sum_{i=1}^{n} \frac{1}{1 + 10^{pH - pK_+}} \tag{1.2}$$

with pK_- and pK_+ the pK_a-values of positive and negative groups, respectively.

(continued)

Interactions between charged groups within the protein significantly change the pK_a-values [1]. In addition, even the tabulated pK_a-values for amino acids in water vary among different authors by about 0.5 *p*H-units. Thus the determination of protein charge at different *p*H and of the *p*I has to be done in vitro rather than in silico. There are, however, programs available which try to *estimate* the *p*I of a protein, for example, protcalc.sourceforge.net/. On isoelectric.ovh.org/ the issue is discussed further and the algorithms commonly used are described.

1.3 The One-Letter Code

In most cases amino acid names are abbreviated with the first three letters of their names. These abbreviations are easy to remember, however, they use up unnecessary memory in computer databases. The 22 proteinogenic amino acids can be encoded by the 26 letters of the Roman alphabet (leaving space for some rare amino acids), and then each amino acid in a protein sequence uses up only 1 rather than 3 bytes of storage space. Unfortunately, several amino acids start with the same letter (such as Ala, Arg, Asp, and Asn), thus we cannot simply use the first letter to encode them.

The following list should help you to remember one-letter codes:

- Amino acids with a unique first letter: **C**ys, **H**is, **I**le, **M**et, **S**er, **V**al
- Where several amino acids start with the same letter, common amino acids are given preference: **A**la, **G**ly, **L**eu, **P**ro, **T**hr
- Letters other than the firstst letter are used for Asn (asparagi**N**), Arg (a**R**ginine), Tyr (t**Y**rosine)
- Similar sounding names: Asp (aspar**D**ic acid), Glu (glut**E**mate), Gln (**Q**tamine), Phe (**F**enylalanine)
- The remaining amino acids have letters that do not occur in their name: Lys (**K** close to L), Trp (**W** reminds of double ring), Sec (**U**), Pyl (**O**)
- **X** is used as placeholder, meaning "any amino acid". **B** is used for "Asp or Asn", **Z** for "Gln or Glu", **J** for "Ile or Leu". The **-** is used to denote gaps in a protein sequence, e.g., in sequence alignments. **h** is used to denote hydrophobic amino acids (do not confuse with H for His!)

In the medical literature these codes appear mostly to denote mutations: A123Q means that the Ala in position 123 is replaced by Gln. Certain sequences of amino acids occur in several different proteins, where they serve a special function. Such **conserved motives** are usually named after their 1-letter amino acid abbreviations. Thus you may encounter KDEL-motives or DEATH-ATPases.

1.4 Biological Function of Amino Acid Variety

You may now ask why there are so many different amino acids. The answer is that these different molecules have different properties, that let them serve different functions inside proteins (see also Table 1.2).

One difference you have already learned about: there are amino acids whose side-chains can bear positive or negative **charges**, whereas other side-chains are always uncharged. Charged side-chains have different pK_r, which can be can be influenced strongly by neighbouring amino acids, for example, Cys (pK_r = 5–10), His (pK_r = 4–10 and the carboxylic acid group of Glu and Asp (pK_r = 4–7). This is important for proton transfer reactions in the catalytic centre of proteins (acid/base catalysis; see Sect. 5.5 on page 131). Ionisable groups also form the **ionic bonds (salt bridges)** which stabilise protein tertiary structure (see page 32).

Asp, Glu, and His residues can chelate bivalent metal ions including Fe, Zn and Ca. This is important for enzymes with metal cofactors, in hæmoglobin and in some regulatory proteins such as calmodulin.

Some amino acids are **hydrophilic** (= water friendly) because they carry ionised or polar groups ($-COOH, -NH_2, -OH, -SH$). Other amino acids are **hydrophobic** (= water fearing, fat friendly), with long aliphatic (Ile, Leu, Val), or aromatic (Phe, Trp) side-chains. If these residues point into the solution, they force water molecules into a local structure of higher order (i.e., lower **entropy**), which is unfavourable. Burying these residues in the interior of the protein avoids this penalty; this is the molecular basis for **hydrophobic interactions**.

Some amino acids have **small** side-chains (such as glycine), others very big, **bulky** ones (such as tryptophan). The small hydrogen residue of Gly not only fits into tight spaces (see section on collagen (page 324) for an important example), but because it has no β-carbon it can assume secondary structures (see Sect. 2.2 on page 20) that are forbidden for all other amino acids.

Proline has its nitrogen in a ring structure, which makes the molecule very stiff, limiting the flexibility of protein chains.

The SH-group of Cys, the unprotonated His and the OH-group of Ser and Thr are nucleophiles which are essential residues in the active centre of many enzymes.

Some amino acids confer properties to the protein which can be used in the laboratory: Met binds certain heavy metals which are used in X-ray structure determination and reacts with cyanogen bromide ($Br-C\equiv N$) leading to protein cleavage. Cys and Lys are easily labelled with reactive probes. Aromatic amino acids, in particular Trp, absorb UV-light at 280 nm; this can be used to measure protein concentration. In addition they show fluorescence, which can be used to measure distances, and their variation during conformational changes, in proteins.

Table 1.2 Properties of the 21 amino acids encoded in a mammalian genome. Posttranslational modification may change these properties considerably. The helix propensity measures the energy by which an amino acid destabilises a poly-alanine helix

Amino acid	3-letter	1-letter	M_r	pK_1	pK_2	pK_3	*p*I	Hydropathy	Helix propensity	Surface	Volume	Abundance
			(Da)	($-COOH$)	(NH_3^+)	(R)			(kJ/mol)	($Å^2$)	($Å^3$)	(%)
Alanine	Ala	A	89	2.34	9.69	–	6.01	+1.8	0.00	115	67	9.0
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	0.21	225	167	4.7
Asparagine	Asn	N	132	2.02	8.08	–	5.41	−3.5	0.65	160	148	4.4
Aspartic acid	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	0.43	150	67	5.5
Cysteine	Cys	C	121	1.96	8.18	10.28	5.07	+2.5	0.68	135	86	2.8
Glutamic acid	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	0.39	180	114	6.2
Glutamine	Gln	Q	146	2.17	9.13	–	5.65	−3.5	0.16	190	109	3.9
Glycine	Gly	G	75	2.34	9.60	–	5.97	−0.4	1.00	75	48	7.7
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	0.56	195	118	2.1
Isoleucine	Ile	I	131	2.36	9.68	–	6.02	+4.5	0.41	175	124	4.6
Leucine	Leu	L	131	2.36	9.60	–	5.98	+3.8	0.21	170	124	7.5
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	0.26	200	135	7.0
Methionine	Met	M	149	2.28	9.21	–	5.74	+1.9	0.24	185	124	1.7
Phenylalanine	Phe	F	165	1.83	9.13	–	5.48	+2.8	0.54	210	135	3.5
Proline	Pro	P	115	1.99	10.96	–	6.48	−1.6	3.16	145	90	4.6
Selenocysteine	Sec	U	168	2.16	9.40	5.20	3.68					rare
Serine	Ser	S	105	2.21	9.15	13.60	5.68	−0.8	0.50	115	73	7.1
Threonine	Thr	T	119	2.11	9.62	13.60	5.87	−0.7	0.66	140	93	6.0
Tryptophan	Trp	W	204	2.38	9.39	–	5.89	−0.9	0.53	255	163	1.1
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	0.49	230	141	3.5
Valine	Val	V	117	2.32	9.62	–	5.97	+4.2	0.61	155	105	6.9

1.5 Exercises

1.5.1 Problems

1.1. Define the isoelectric point of a compound.

1.2. Connect the following properties with amino acids:

1)	hydrophobic	A)	Tryptophane
2)	positively charged	B)	Serine
3)	small	C)	Lysine
4)	polar	D)	Glycine
5)	aromatic	E)	Glutamine

1.3. Lysine has the pK_a-values 2.18 (carboxy-group), 8.95 (α-amino group) and 10.53 (ϵ-amino group). The *p*I is _____.

1.4. You are working on a research project to elucidate the reaction mechanism of an enzyme. You think that a particular serine residue in the protein is required for catalytic activity. To test this hypothesis you want to genetically replace this amino acid by another, and then test whether the enzyme is still active.

Which amino acid should you choose to replace the Ser?

A Threonine
B Alanine
C Tryptophan
D Glutamic acid
E Histidine

1.5. Which of the following tripeptides would be the most soluble in 1 M NaOH:

A) Phe-Ala-Val
B) Glu-Gly-Asp
C) Gln-Gly-Asn
D) Lys-Arg-His
E) Trp-Lys-Asn

1.5.2 Solutions

1.1 The isoelectric point of a compound is the *p*H, at which it has an equal number of positive and negative charges.

1.2 Trp is aromatic and hydrophobic, Lys positively charged (ϵ-amino group), Gly small (only hydrogen as side-chain), Ser is polar (OH-group). Gln does not fit into any of these categories.

1.3 Below *p*H 2.2 Lys has 2 positive charges from the fully protonated α- and ϵ-amino groups, and no negative charge because the carboxy group will be mostly

protonated as well. Above *p*H 2.2, the carboxy-group will lose its proton, resulting in 1 negative and 2 positive charges, and a net charge of +1. Beyond *p*H 8.95, the α-amino-group will lose its proton and there will be one negative and positive charge, resulting in a net charge of ± 0. Beyond *p*H 10.53, the proton on the ϵ-amino group will be lost as well, resulting in one negative and no positive charges. Thus the *p*I is calculated as $1/2 * (8.95 + 10.53) = 9.74$.

1.4

A Thr also has the OH-group, only has an additional C. It would likely work just as well as Ser, the experiment would not answer the question.

B Ala is Ser minus the $-$OH group which is the catalytically active part of Ser. Thus Ala would be ideal to test the hypothesis.

C Trp is the bulkiest of all amino acids. If the enzyme were no longer active after the replacement, you would not know whether this was because of the lack of the catalytically active OH-group or because of disruption of the 3D-structure of the enzyme.

D Glu is acidic. If you use it to replace a polar but uncharged amino acid the 3D structure of the enzyme would likely be perturbed by salt-bridge formation. If the enzyme didn't work afterwards, you'd not know whether this was because of changed 3D structure or because of the missing OH-group.

E His is basic. If you use it to replace a polar but uncharged amino acid the 3D structure of the enzyme would likely be perturbed by salt-bridge formation. If the enzyme didn't work afterwards, you'd not know whether this was because of changed 3D structure or because of the missing OH-group.

1.5

A) Phe-Ala-Val All three amino acids are hydrophobic, this peptide would be only very sparingly soluble in both water and base.

B) Glu-Gly-Asp Glu + Asp are acidic residues and give additional charges at alkaline *p*H.

C) Gln-Gly-Asn Gln and Asn have acid amide functional groups, which are somewhat polar but do not become charged at high *p*H. Gly is hydrophobic.

D) Lys-Arg-His All three amino acids are basic, at high *p*H they will be uncharged. This peptide would be very soluble in acid, but not in base.

E) Trp-Lys-Asn A bulky hydrophobic, a basic and a somewhat polar amino acid would give low solubility in water and base, but somewhat better solubility in acid.

Reference

1. G.R. Grimsley, J.M. Scholtz, C.N. Pace, A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci. **18**, 247–251 (2009). doi: 10.1002/pro.19

Chapter 2
Protein Structure

Nature is greatest in the smallest things.
(C. LINNÆUS: *Systema Naturae*)

Abstract Peptides and proteins are made by condensation of amino acids, forming peptide bonds. The sequence of amino acids in a protein is called its primary structure. Secondary structure is determined by the dihedral angles ϕ, ψ of the peptide bonds, the tertiary structure by the folding of protein chains in space. Association of folded polypeptide molecules to complex functional proteins results in quaternary structure. Proteins can be further modified by posttranslational addition of small molecules.

2.1 Primary Structure

A peptide bond is formed by the condensation of two amino acids under elimination of water (see Fig. 2.1). Addition of further amino acids to the chain leads to tripeptides, tetrapeptides, and so on. Chains of up to 20 amino acids are called oligopeptides (oligo = few), and longer ones polypeptides (poly = many). **Proteins** are polypeptides with a biological function.

Polypeptides range in size from a few amino acids to thousands; Fig. 2.2 shows aspartam, a dipeptide. Proteins consist either of a single polypeptide chain, or they are formed from separate polypeptide chains called **subunits**. Some proteins contain other covalently bound components, prosthetic groups, and posttranslational modifications (see below).

The sequence of amino acids in a protein is called its **primary structure**. In biochemistry, this is always given starting with the N-terminal and ending with the C-terminal amino acid, because this is the order in which amino acids are added during protein synthesis in the cell (this process is discussed in detail in textbooks of molecular cell biology, e.g., [1, 21]).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_2

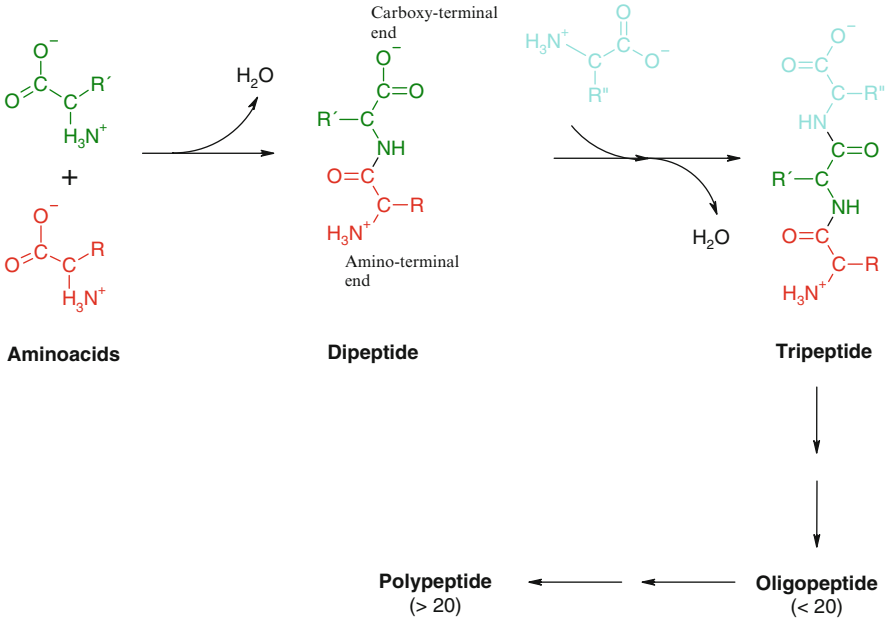
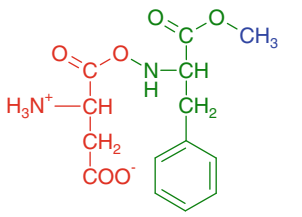


Fig. 2.1 Polycondensation of amino acids to peptides and proteins. Polycondensation is a reaction where organic molecules react with each other via their functional groups, producing small molecules (here: water) in addition to a macromolecule

Fig. 2.2 Aspartam is a peptide used as an artificial sweetener in the food industry



aspartyl-phenylalanine-1-methyl ester (Aspartam®)

C^{α}, C', the nitrogen and the oxygen atom of the peptide bond form a single plane. The bond between C' and N is somewhat shorter than a normal C−N single bond, because of **mesomery** with the C=O double bond (see Fig. 2.3). Since the lone electron pair of N enters into the partial double bond it can no longer accept a proton; the N in a peptide bond is not basic.

Thus the peptide bond has a "partial ($\approx$ 40 %) double bond character" (see Fig. 2.4). Like the C=C bond, it is planar and cannot rotate. The H and O of the peptide bond are in the *trans*-configuration. Formally, we express the same

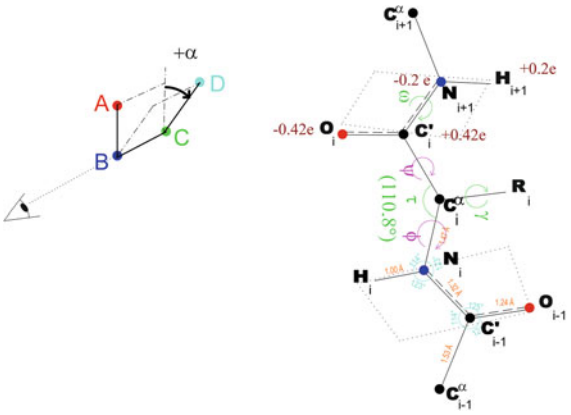


Fig. 2.3 The geometry of the peptide bond. *Left*: The bonds between A−B and B−C define a plane, as do the bonds between B−C and C−D. The angle between these planes is called the dihedral angle of the B−C bond. For the standard way to determine this angle orient that bond into the paper plane, so that the neighbouring atoms (here A and D) point upwards. Then measure the angle formed: clockwise is positive, anticlockwise negative. *Right*: Because of mesomery, the dihedral angle of the bond between the carboxy-carbon (C′) and the nitrogen (ω) is fixed to 180°, with N, H, C and O lying in a single plane. Slight deviations are possible, but rare. The bond angle of C^{α} (τ) is 110.8° ± 2.5° for 86 299 residues investigated in a recent study [41]. Variable are the dihedral angles ϕ and ψ, which determine the secondary structure of a protein (see next section)

idea by saying that the dihedral angle ω of the peptide bond is fixed to 180° (see Fig. 2.4). Because of the bulky R-groups, the *trans*-configuration is more stable with most amino acids (99.7 % probability). The exception is Pro, which occurs in *cis*-configuration much more frequently than other amino acids (5.8 % probability, see Fig. 2.5).

On the other hand, the $N-C_{\alpha}$ and $C_{\alpha}-C'$ bonds are normal single bonds; rotation around those is possible. The angles of rotation are named ϕ and ψ, respectively. Rotation in the peptide chain is limited by two factors. First, at certain angles ϕ and ψ around one amino acid an atom of that amino acid would collide with an atom of the following amino acid (see Fig. 2.6). The angles ϕ, ψ which result in a clash between $C'_n{=}O$ and $N_{n+1}{-}H$ are defined as 0°, 0°. Additionally, size and charge of the R-groups can make certain positions more stable than others.

Thus in a plot of ϕ versus ψ (RAMACHANDRAN-plot [18, 27], see Fig. 2.9 on page 23) there are regions that are sterically forbidden, there are fully allowed regions with no steric hindrance, and there are unfavourable regions which can be assumed by slight bending of bonds.

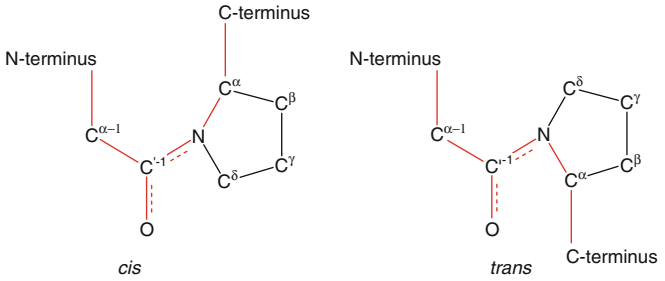


Fig. 2.4 *cis-trans*-isomery around the peptide bond. Because the $C'^{-1}-N$-bond has the character of a partial double bond, rotation around this bond cannot occur and *cis-trans* isomery results. For steric reasons the *trans*-configuration is much more probable than the *cis*-. Pro is unusual in that the *cis*-configuration has a probability of 5–6 %, which is about 100 times higher than with other amino acids

Proline is again a special case because the peptide nitrogen is part of a ring structure; this limits ϕ to values between $-35°$ and $-85°$. As I will show in a moment, this has considerable consequences for protein secondary structure.

Because **glycine** has only a hydrogen as R-group, steric hindrance is much less a problem than with other amino acids. Thus in a RAMACHANDRAN-plot Gly can be found in regions forbidden for other amino acids (Fig. 2.9).

Internet Resources on Protein Structure
Amino acid sequences of proteins and the nucleotide sequences of their genes are stored in databases that are accessible from the Worldwide Web, such as the nonredundant protein sequence database Owl (http://www.bioinf.man.ac.uk/dbbrowser/OWL/) and ExPASy (http://au.expasy.org/). Multiple sequence alignment services are offered by BCM (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html). For taxonomic questions in general, http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html may be consulted. Several discussion groups in the `bionet` hierarchy on Usenet deal with proteins, especially `bionet.molbio.proteins`.

2.1.1 Protein Sequences and Evolution

The great variety of proteins that can be observed today has arisen from a much smaller number of **ancestors** during evolution. This can be shown by comparing the primary structure of proteins; the more similar they are, the closer they are related [10].

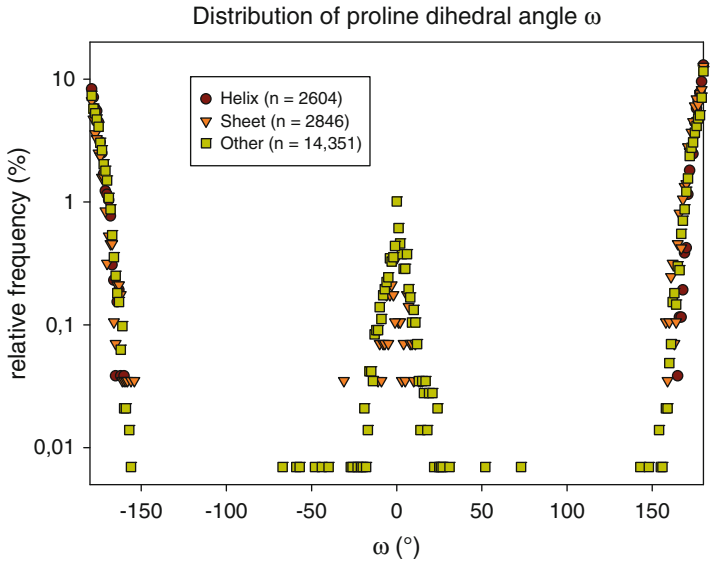


Fig. 2.5 Distribution of dihedral angle ω of Pro in 1453 "non-redundant" proteins whose structure is known with a resolution ≤ 1.5 Å. Of the 431 146 amino acid residues 19 801 were proline. In α-helices, all of the 2604 Pro occurred in *trans* ($\omega \approx 180°$). In β-strands, 73 out of 2846 (2.6 %) Pro occurred in *cis* ($\omega \approx 0°$), in "other" secondary structures 1071 out of 14 351 Pro (7.5 %) were in *cis*. Almost all other Pro residues occurred in *trans*, the number of Pro residues with intermediate ω was 0, 1, and 11, respectively

Because each position in the primary structure can be occupied by any of the 20 common amino acids (and occasionally also by Sec and Pyl), the possible number of combinations is huge. For example, a protein with 100 amino acids has $20^{100} = 1.3 \times 10^{130}$ possible sequences. Given that our universe is about 13.7×10^9 a $\approx 4.32 \times 10^{17}$ s old, creationists have argued that proteins cannot have been created by a process of random mutation and selection. This argument is fallacious, however, because it makes the (unspoken!) assumption that the function of a protein can only be met by one particular amino acid sequence. The existence of isoenzymes – proteins with different structure but the same function – proves this assumption wrong (see Fig. 2.8). Interestingly, the study of protein- and DNA-sequences [10] has confirmed, and in many cases given us additional details on, the tree of life first proposed by C. DARWIN [9].

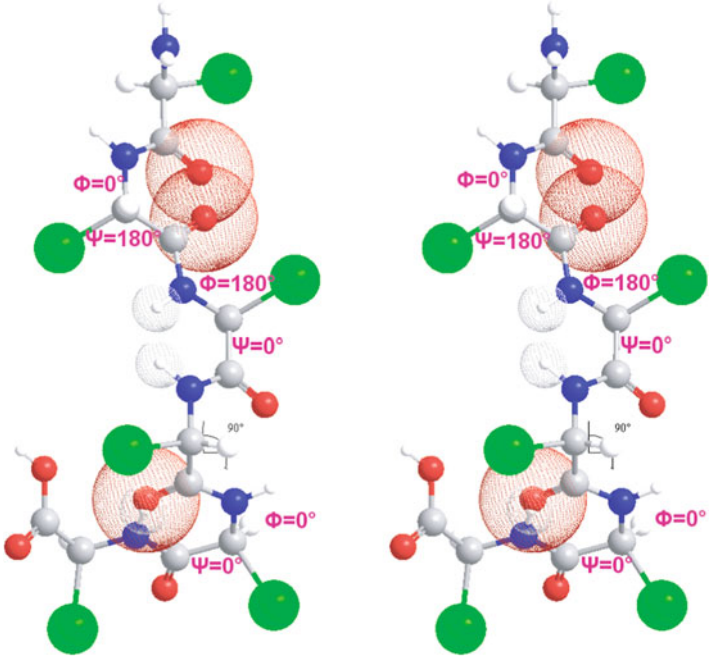


Fig. 2.6 Some ϕ, ψ-combinations lead to collisions between atoms of neighbouring amino acids. Such angles are forbidden; the dihedral angles ϕ, ψ which result in a clash between $C'_n{=}O$ and $N_{n+1}{-}H$ are defined as 0°, 0°. The green balls represent the R-groups. In addition to these next-neighbour effects, some angles also lead to collisions between amino acids further apart. This is a stereogram: if you look at the images cross-eyed, you will see three figures, the middle of which is three-dimensional (this takes some practice and is not a required skill for a physician). Cheap lorgnette-style stereo viewers are available on the Internet to help with this. They are built around two prisms which ensure that each eye sees only the image intended for it (see http://www.shortcourses.com/stereo/stereo1-7.html for examples)

2.2 Secondary Structure

The secondary structure of a protein is any regular, repetitive folding pattern in the molecule. It is stabilised by **hydrogen bonds** (see Fig. 2.13) between the amino- and keto-groups of the peptide bonds, which carry a partial positive and negative charge, respectively (see Fig. 2.3b). Although each hydrogen bond has only a relatively small bond energy ($\approx$ 5 kJ/mol), the sum of the bond energies over all hydrogen bonds in a protein is considerable.

The following structural motives are particularly common [19, 28].

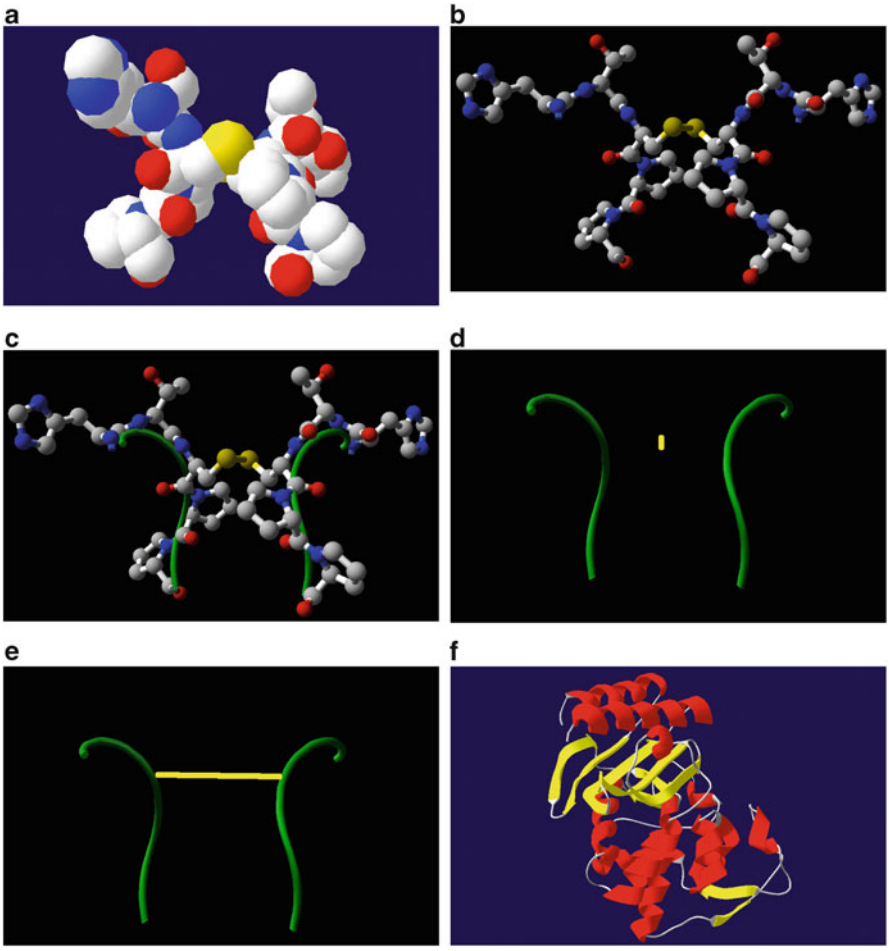


Fig. 2.7 Interpretation of structure diagrams of proteins, using the pentapeptide HTCPP. (**a**) Space-filled diagrams show the true (VAN DER WAALS-) extension of a molecule, but even in short peptides clarity is lost. (**b**) A wire diagram is clearer. The centres of the atoms are connected by thin lines; hydrogens are not drawn. Wire diagrams look a little like structural formulas in organic chemistry, however, atoms are shown in their true three-dimensional arrangement. (**c**) For larger proteins even wire diagrams would be too cluttered. Thus the atoms forming the protein backbone are connected by a thick line, which is used to represent the amino acid chain. (**d**) If the protein contains disulphide bonds, showing only the backbone trace leaves the disulphide bond dangling in free space, thus (**e**) this bond is often shown (incorrectly, but easier to interpret) connecting the backbone traces. (**f**) A further abstraction is achieved by showing elements of secondary structure instead of the backbone trace: alpha helices as red helices and β-strands as yellow arrows, connected by the backbone trace of coils and turns (in *grey*). Other colouring schemes used in this book are N-terminal (*red*) to C-terminal (*purple*), different colours for different chains or "shapely colours" (a quasi-standard in molecular modelling) for different amino acids. Shown here is β-lactamase (PDB-code 1m40, a protein that confers penicillin resistance to bacteria)

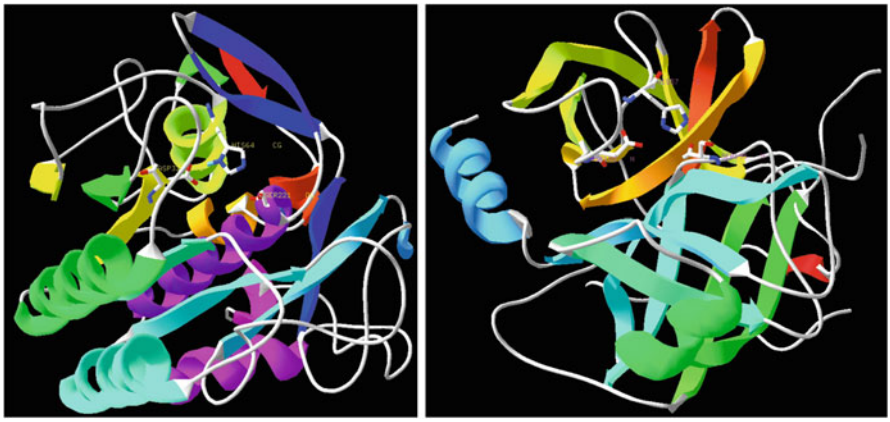


Fig. 2.8 Subtilisin (PDB-code 3VYV, *left*) and Chymotrypsin (PDB-code 1oxg, *right*) are both Ser-proteases that use the classical catalytic triad (Ser, His, Asp, shown as wire diagram) in the catalytic centre to cleave proteins. These amino acids are far apart in the sequence, but close to each other in the folded protein. Both proteins, however, have completely different sequences and secondary structures. This is an example of convergent evolution ("re-inventing the wheel"). The proteins are called *iso*-enzymes (*iso* = the same)

2.2.1 The α-Helix

The polypeptide chain is wound in a counterclockwise spiral around an imaginary axis, with 3.6 amino acids per turn. Such a spiral is called **righthanded**, because if you hold your right hand with the thumb pointing from N- towards C-terminus the fingers curl counterclockwise. Left-handed helices are improbable for L-amino acids, because the β-carbon and carbonyl-oxygen would collide.

Each turn is about 5.4 Å long, the pitch per amino acid is 1.5 Å. The angle between successive residues is about 100°, $\phi = -57°$, $\psi = -47°$. The R-groups point outward. This compact, rod-like structure is maintained by hydrogen bonding between a carboxy-oxygen (partial negative charge) and the amino hydrogen (partial positive charge) 4 amino acids further along the chain. A full turn of the helix has 3.6 amino acids. This "hydrogen bond loop" totals 13 atoms, hence α-helices are sometimes called 3.6_{13}-helices.

Because all N-termini point in the same direction, an α-helix has a dipole moment and can bind to charged molecules. Proline and glycine don't fit well into the α-helix. The α-helix is the most common secondary structure in proteins. It occurs both in many fibrous (long, stretched-out) proteins (such as myosin and keratin), and in many globular (compact-shaped) proteins. Often α-helices have a polar side (facing the outside of a protein) and a nonpolar one which is buried in the interior (**amphipatic** helix).

Proline introduces kinks when it occurs within α-helices, as the secondary amide cannot donate hydrogen bonds. However, it is ideally suited as the N-terminal amino acid in helices, where other amino acids would have a "lonely" backbone-amide.

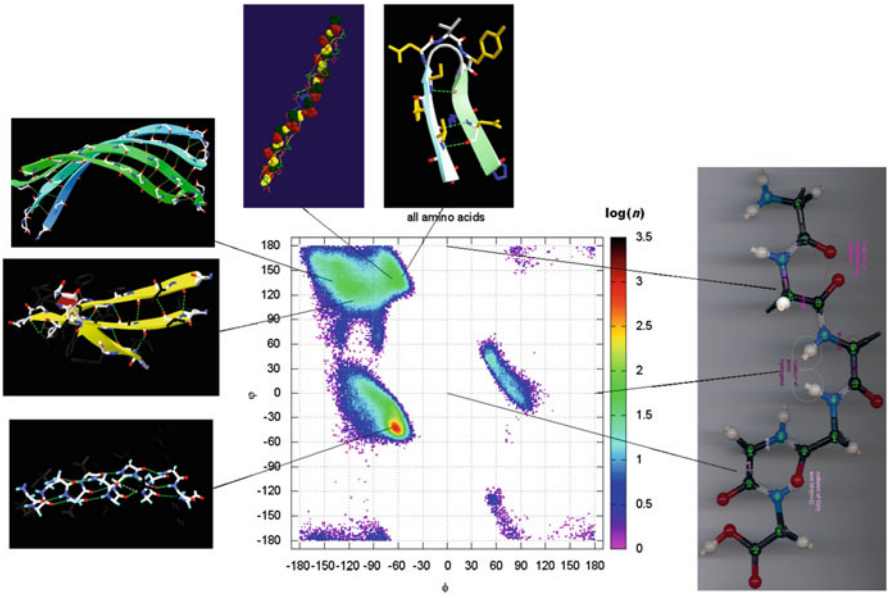


Fig. 2.9 RAMACHANDRAN-plot of "representative/non-redundant" proteins with ≤ 1.5 Å resolution from the PDB database (2014-05-27, 1453 proteins with 431 146 amino acids). For each combination of dihedral angles ϕ and ψ, the frequency was counted (bin size 1°) and represented by a colour of the rainbow (from purple = rare to dark red = frequent). In the top left corner of the diagram are the extended structures (β-strands, P_{II}-helix and turn). The big peak below is at the coordinates of the α-helix; the small peak to the right is the left-handed α_l-helix, a structure that can be only a few amino acids long. Note that certain ϕ, ψ-combinations result in steric clashes, these therefore do not occur in proteins (frequency = 0, represented by white)

Function of α-helices:

- An α-helix of 22 amino acids is long enough to span a double membrane. The part of the helix that is inside the membrane consists of hydrophobic amino acids that can interact with the lipid tails of the membrane. Hydrophilic amino acids on both ends interact with the cytosol and the interstitial fluid, respectively. The cytosolic end has positively charged amino acids and the extracellular end more negatively charged ones, because the potential of a cell is negative inside (−70 mV). Thus the correct orientation of the protein is ensured by the electrical field. At the interface between the membrane and the aqueous environment one finds predominantly aromatic amino acids and Lys.
- Amphipatic α-helices at the N-terminus of a protein serve as recognition sites for the import into mitochondria. Every 4th or 5th amino acid is positively charged, so that all positive charges are in the same quadrant of the helix (see Fig. 2.10 and Fig. 16.10 on page 375).
- Two α-helices wound around each other form a **coiled coil**. Keratin consists of such coiled-coils (see Fig. 2.11). These are held together by disulphide bonds.

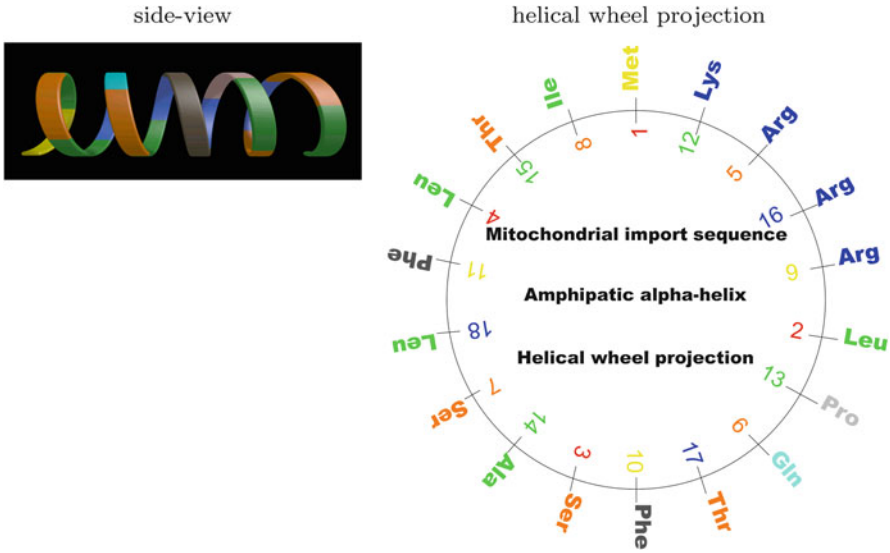


Fig. 2.10 Signal-peptide for import into mitochondria. Most mitochondrial proteins are encoded in the nucleus; they are synthesised in the cytosol and then imported via a transport system that spans both mitochondrial membranes (see Fig. 16.10 on page 375). An amphipatic α-helix serves as the recognition signal for binding of the nascent protein to the transporter. Note that the helical wheel projection is viewed from the N-terminus

Breaking these with thioglycolic acid is the basis of the **permanent wave**.

- Heptad-repeats (Leu-zippers) are α-helices where every 7th amino acid is leucine (Fig. 2.12). Such helices associate because of hydrophobic interactions between the Leu-residues, allowing for specific dimerisation of proteins. Some DNA-binding proteins have this structure.

2.2.2 β-Strand

In the β-strand, the polypeptide backbone is stretched out with $\phi, \psi \approx -120°, 120°$. Several strands are aligned either in a parallel (all carboxy-terminal ends are at the same side) or antiparallel fashion, forming hydrogen bonds between a N−H group of one strand and a C=O-group in a neighbouring strand. This gives rise to a large blanket-like structure, the **β-pleated sheet**. The main difference between the α-helix and β-strand is that in the α-helix hydrogen bonds occur between residues of the same helix, whereas in a β-pleated sheet they occur between residues of neighbouring strands (see Fig. 2.13). Nevertheless, a single β-strand is stable because

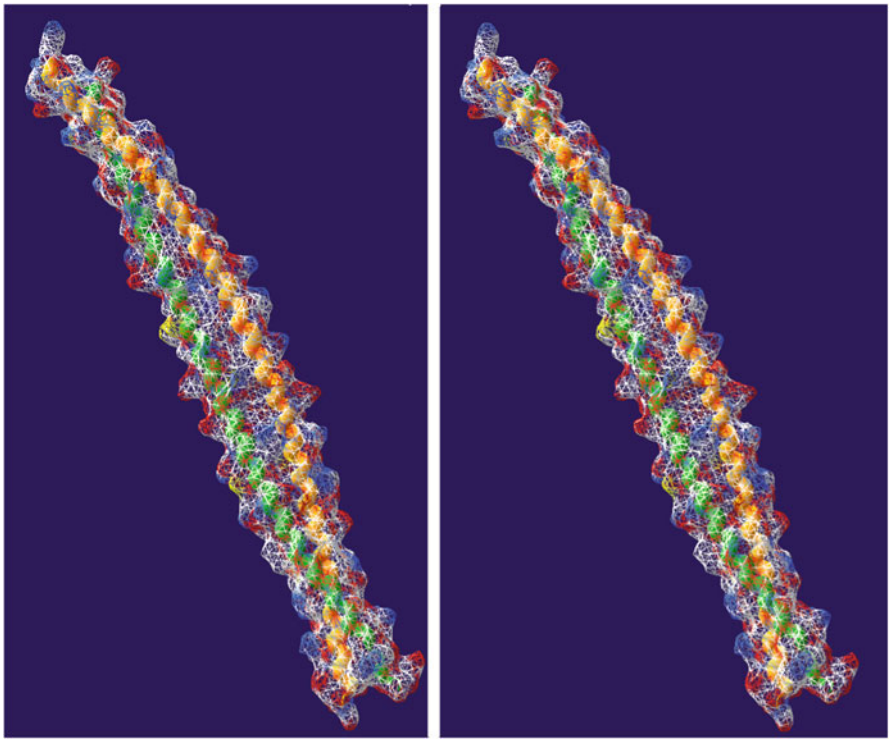


Fig. 2.11 Keratin is a heterodimer that forms coiled-coils. Depicted here is coil 2B of keratin 5 and 14 (PDB-code 3tnu; the structure of the entire keratin molecule has not been solved yet). The helices are shown as green and orange ribbons; in addition the protein surface is shown (blue = basic, red = acidic, yellow = polar and grey = nonpolar)

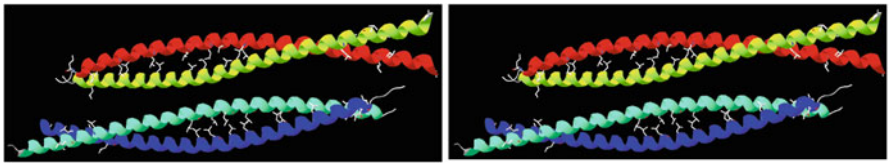


Fig. 2.12 In heptad-repeats (Leu-zipper, here tropomyosin, PDB-code 1ic2) every seventh amino acid is Leu. This leads to specific associations of α-helices by hydrophobic interactions

the amino acids in this extended structure have plenty of "wiggling" space without running into steric hindrance (look up the coordinates in the RAMACHANDRAN plot!), resulting in entropic stabilisation. The R-groups point up- and downwards in turn, making amphipatic sheets with polar and nonpolar or positive and negative faces possible. The entire sheet is rarely flat, but has a right-handed twist, in extreme cases forming a β-barrel (see Fig. 2.13). In schematic diagrams of protein structure each β-strand is drawn as a broad arrow.

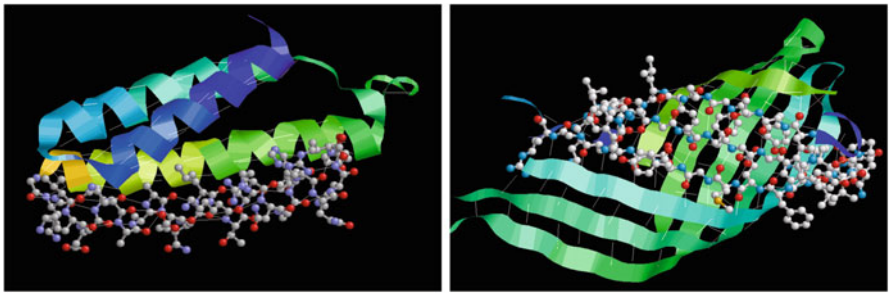


Fig. 2.13 Hydrogen bonding in α-helix (*left*, cytochrome b_{562}, PDB-code 256B) and β-sheets (*right*, *E. coli* OmpA, PDB-code 1QJP). In an α-helix all hydrogen bonds between keto- and amino-groups in the protein backbone occur between neighbouring amino acids of the same helix. In β-sheets, however, all such hydrogen bonds occur between amino acids in different strands, alternating between the right and left neighbour

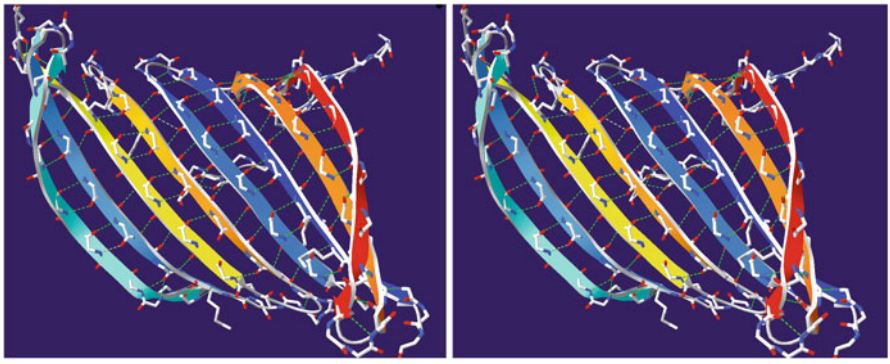


Fig. 2.14 Anti-parallel β-pleated sheet of the silk fibroin N-terminal domain (FibNT) from the silkworm *Bombyx mori* L. at pH 4.7 (PDB-code 3ua0). Two subunits (red to yellow and cyan to blue) form the sheet. Neighbouring strands have alternating directions, and are joined by β-turns. Hydrogen bonds holding the sheet together run at right angle to the strands

2.2.2.1 The Antiparallel β-Sheet

In an antiparallel β-sheet (see Fig. 2.14) the stands point in alternating directions. They are usually joined together by β-turns (see later). Ideal $\phi, \psi = -138°$, 137°. **Silk-protein** is an example for the use of β-sheets in biologically important structures. The amino acids within a β-strand are already in an extend conformation, therefore silk shows little elasticity and has an extremely high tensile strength, as any extension would require breaking covalent bonds. On the other hand, the strands are held together by hydrogen bonds only, giving silk cloth this wonderful soft flow. At neutral *p*H in the silk gland the fibroin protein has soluble random coils; as the silk thread is ejected, acidification leads to β-sheet formation and precipitation of the

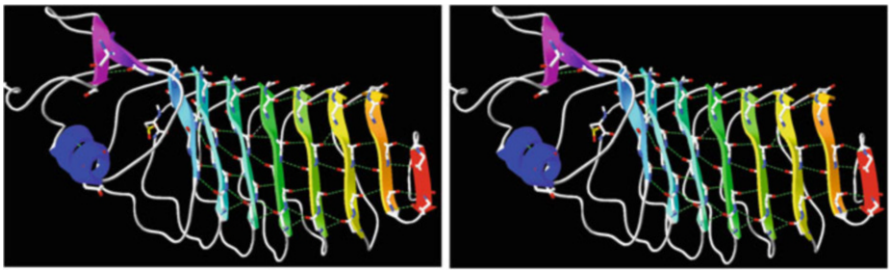


Fig. 2.15 Parallel β-pleated sheet (here PDB-code 2v9s). All strands have the same direction; the "return-legs" are either α-helices or coils. Hydrogen bonds holding the sheet together run obliquely to the strands

protein. Each silkworm cocoon is made from a single silk thread that is 900 m long and has a diameter of 10 µm.

2.2.2.2 The Parallel β-Sheet

In a parallel β-sheet (see Fig. 2.15) the N-termini of all strands point in the same direction, ideal $\phi, \psi = -116°, 111°$. The hydrogen bonds are oblique to the strand direction, hence the parallel β-sheet is less stable than the antiparallel. The strands in a parallel β-sheet are often joined by α-helices, which form the "return-leg".

The ideal parallel and antiparallel β-sheets are characterised by different ϕ, ψ-values. However, because of the twisting of strands in a β-sheet there are no separate peaks for them in the RAMACHANDRAN-plot; rather, they merge into a single big area of extended structure.

2.2.2.3 β-Helix and β-Roll

Parallel β-strands can be wound into right-handed coils (see Fig. 2.16), containing either two (β-roll) or three (β-helix) strands per rung [40]. The β-helix is found in enzymes whose substrates are oligosaccharides (e.g., pectinases); it also occurs in tailspike proteins of bacteriophages and in amyloid aggregates, the cause of several debilitating diseases (see Sect. 10.2 on page 206). In a β-helix one or two of the three β-strands may be replaced by α-helices (see Fig. 2.17). Three β-helices or -rolls can be arranged in coils, similar to the α-helical coiled-coils.

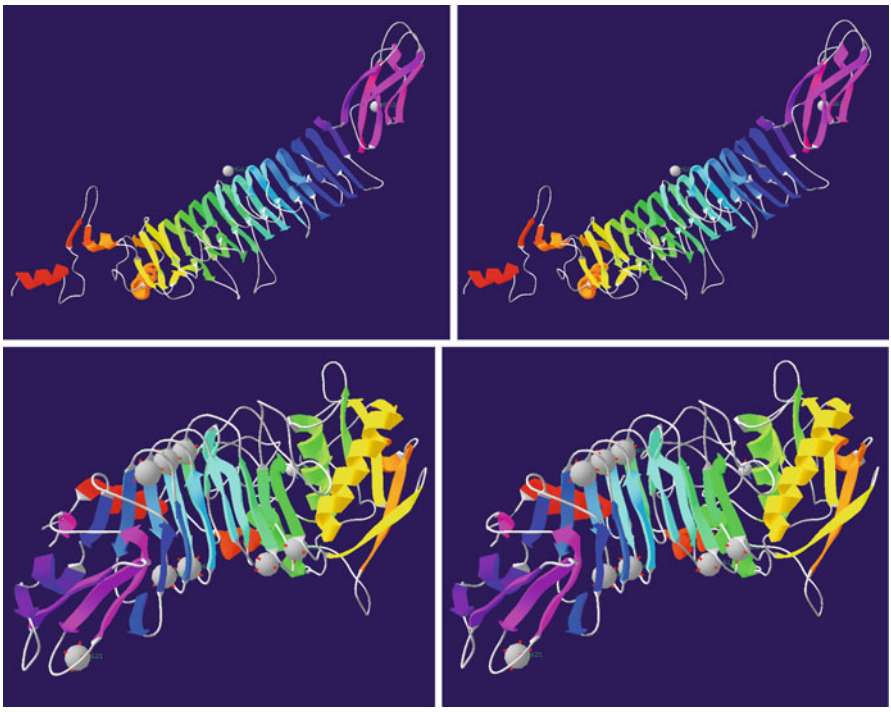


Fig. 2.16 *Top*: The tailspike protein from the bacteriophage Sf6 contains a β-helix with three parallel β-sheets (PDB-code 2vbk). Three such helices are wound around each other, forming a coiled-coil (not shown). *Bottom*: Alkaline protease from *Pseudomonas aeruginosa* (PDB-code 1kap) is an example for a β-roll with two parallel β-sheets. The structure is stabilised by Ca^{2+}-ions

2.2.3 *The P_{II} (syn.: Poly-Pro or Polypeptide II) Helix*

The P_{II} helix is left-handed with three residues per turn and $\phi, \psi = -70°, 140°$. As is the single β-strand, it is stabilised by entropy, not by hydrogen bonds. Pro frequently occurs in this structure, but not all P_{II} helices contain Pro.

2.2.3.1 Collagen

Collagens (see Sect. 14.1.1 on page 324 for a more detailed discussion) are the most important example for the P_{II}-helix, they consist of three P_{II}-helices wound around each other (hetero- or homo-oligomer, see Fig. 2.18). The human genome contains 42 collagen genes, which encode for 28 known collagen types. Of these types I, II, and III are the most important. Each of the three molecules in collagen has 1050 amino acids, with the sequence Gly-X-Pro. The angle of the Pro peptide bond (amino group part of a ring) allows the sharp turn in the molecule [3], and the

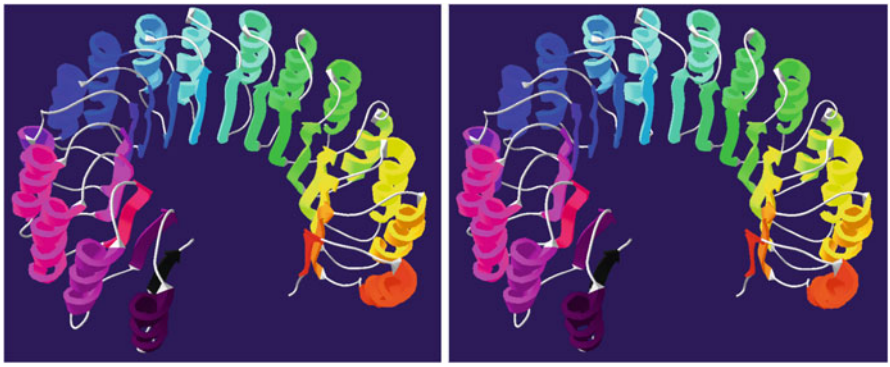


Fig. 2.17 In porcine ribonuclease inhibitor (PDB-code 2bnh) α-helices and β-strands form a α/β coil. To prevent side-chain interference, the α-helices have to twist, resulting in a circular, rather than helical, structure

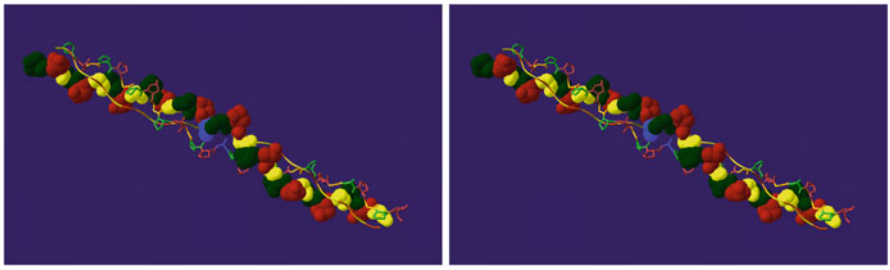


Fig. 2.18 Collagen (PDB-code 1cag). To make the tight association between the three strands clearer, one each is drawn space-filling, as wire diagram and as carbon-backbone. Note the repeating Gly-X-Pro (yellow, green, brown; with X often hydroxy-Pro) sequence. Marked in blue is a Gly→Ala mutation that prevents a close fit and destabilises the molecule. Such mutations cause, for example, EHLERS-DANLOS-syndrome

small R-residue of Gly (only a H) allows the three protein molecules to wrap tightly around each other. The resulting "rope" has a tensile strength higher than steel. If only a single one of the Gly-residues in one of the collagen chains is mutated, wrapping is no longer possible, leading to **osteogenesis imperfecta** (brittle bone disease, collagen I), to EHLERS-DANLOS**-syndrome** (collagen I, III, or V), with too brittle or too elastic ligaments and death by vascular or organ rupture, **epidermolysis bullosa** (blistering of skin, collagen XVII), or to ALPORT**-syndrome** (collagen IV, kidney, and hearing defects).

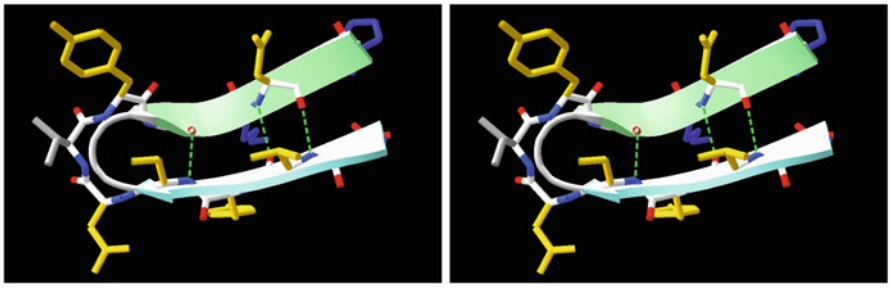


Fig. 2.19 β-turns (here in PDB-code 1qiv) are most common between the strands of an anti-parallel β-sheet

Heating turns collagen into **gelatine**. The dissociation temperature of collagen is influenced by the hydroxylation of proline. **Vitamin C** (ascorbic acid) is required for the correct function of Pro-hydroxylase (see Fig. 14.4 on page 328). In **scurvy**, the dissociation temperature of collagen drops below the body temperature of 37 °C, explaining the connective tissue weakness typical for this condition.

Apart from collagen, P_{II}-helices also occur in **SH3-domains**, which occur in proteins involved in signal transduction.

2.2.4 Hairpin Turns

Hairpin Turns allow the protein to fold back onto itself in a 180° angle. They are important, for example, between the different strands of an antiparallel β-sheet. Because the C=O- and N−H-groups of a turn are not all involved in hydrogen bond formation within the protein, they are often surface-exposed and interact with water. They may also occur in the catalytic centre of enzymes, where they are involved in substrate binding.

Turns can contain 4 amino acid residues (β-turn, frequent, with a hydrogen bond between $N{-}H_i \cdots O{=}C_{i-3}$, see Fig. 2.19) or 3 (γ-turn, rarer). Turns with 2 (δ), 5 (α) or 6 (π) amino acids have been described, but are very rare. Turns often contain Gly (smallest amino acid) or Pro residues, the latter because of its specific value of ϕ; in addition the C^{α} and C^{δ} of Pro can undergo $C{-}H{\cdots}\pi$ interactions with neighbouring aromatic amino acids, which—although not as strong as regular hydrogen bonds—can stabilise the turn [3]. Turns may also be found in the catalytic centre of an enzyme (with a $C{-}H{\cdots}\pi$ bond between Pro and an aromatic substrate). The different types of turns can be distinguished by the ϕ, ψ-values of their peptide bonds [39].

γ-Turns and Cardiovascular Disease
γ-turns occur in fibrinogen; these are recognised by the GPIIb-IIIa receptor on platelets, leading to platelet aggregation and thrombus formation. Mimetics that structurally resemble γ-turns can block that receptor and are candidates for drugs that reduce blood clotting in patients at risk for strokes.

2.2.5 Rare Structures

In addition to the α-helix there are two other, much rarer helical conformations

3_{10}-helix 3 residues per turn and a hydrogen bond between residues i and i+3 ($\phi, \psi = -50°, -25°$). It occurs only at the C-terminal end of α-helices, and can be only 4–5 residues long.

π-helix with 5 residues per turn and a hydrogen bond between residues i and i+5. The π-helix is usually only 7–10 amino acids long and flanked on both ends by α-helices. In effect, it introduces a kink into a long α-helix (see Fig. 2.20). Evolutionary, they are created by insertion of an amino acid into an α-helix. They occur at least once in about 15 % of proteins, often in the active site of an enzyme. π-helices have variable ϕ, ψ-values.

There are some other structures which occur in only a few proteins, but have important functional roles. We discuss those when we talk about some special proteins. Each secondary structure can be characterised by the ϕ, ψ angles in the protein backbone (see Fig. 2.9).

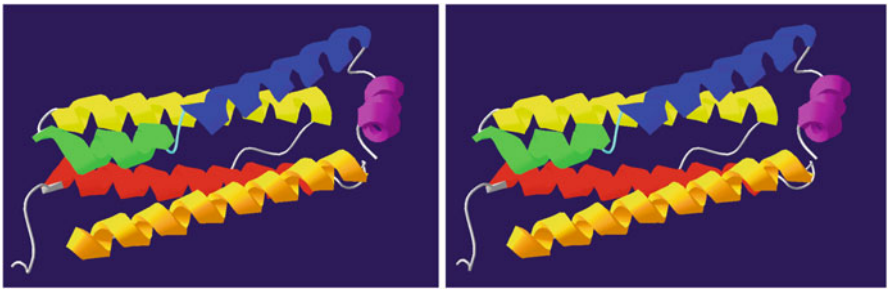


Fig. 2.20 π-helix (*cyan*) introduces a kink between the flanking α-helices, here human ferritin (PDB-code 3ajo)

2.2.6 *Coils*

Coils are any structure except those mentioned above. Note that amino acids in coils still have a defined position within the structure of a protein, thus the terms "random coil" or "unordered", sometimes found in the literature, are misleading. These areas have an important function too, because they add flexibility to the protein and allow conformational changes, for example, during enzymatic turnover. Their peptide bonds are not involved in intra-protein hydrogen bonding, therefore they are often exposed to interact with water, small ligands, or with other proteins. Coils tend to tolerate mutations better than other structures and are therefore hotspots for evolution. In Chap. 11 on page 225 we will see that it is coils that give **antibodies** their specific binding properties.

2.3 Tertiary Structure

Tertiary structure describes the global conformation of a protein, in other words, the way in which the elements of its secondary structure are arranged in space. Tertiary structure is determined by

Hydrophobic interactions of amino acid side-chains. Typical globular proteins have a core of hydrophobic side-chains, whereas hydrophilic side-chains are on the surface where they interact with water or with other proteins. If hydrophobic residues were exposed to water, the water would have to form an ordered cage (so-called clathrate) around them, which would decrease the entropy of the system.

VAN DER WAALS-interactions are fluctuating dipole interactions with a bond energy of 4–17 kJ/mol. The bond length is ≤ 4 Å.

Hydrogen bonds are interactions between permanent partial charges. The bond length is about 3 Å; the bond energy is 2–6 kJ/mol if both partners are partially charged and up to 21 kJ/mol if one partner is fully charged. If the distance between the partners is too large, an **indirect hydrogen bond** may be formed where water acts as a bridge ($\delta^{-}\cdots \mathrm{H_2O}\cdots{}^{+}\delta$)

Salt bridges are interactions between fully charged groups. The bond length is 2.8 Å. The bond energy is 10–30 kJ/mol in an aqueous environment, but can be significantly higher if both groups are buried in a hydrophobic core.

Disulphide bonds are formed between two Cys residues after folding of the protein into its higher-order structure ($\mathrm{R{-}SH} + \mathrm{HS{-}R'} \rightarrow \mathrm{R{-}S{-}S{-}R'} + 2[\mathrm{H}]$). This is an oxydation (removal of hydrogen), which will not normally occur in the reducing environment of the cytosol. However, the environment inside the ER is oxydising. Thus disulphide bridges are found

more frequently in the cell surface and secreted proteins than in cytosolic ones. They may occur between two Cys residues in the same polypeptide (intrachain), or between different polypeptides (interchain). Bond length is 2.2 Å and bond energy 167 kJ/mol.

Coordination around cofactors Several amino acids in a protein can be involved in the coordination of metal ions (Ca, Zn, Fe, Mg, Na, K) or prosthetic groups such as **hæme** or FAD.

In transmembrane segments, hydrophilic amino acids are in contact with water (**snorkelling effect**) and hydrophobic amino acids are found in contact with the fatty acid tails of the lipids (**anti-snorkelling effect**). The lipid/water interface is formed by three amino acids with special properties: Trp, Tyr, and Lys (the so-called **aromatic belt**; see Fig. 2.21). They have in common relatively long molecules which are hydrophobic, but have a hydrophilic (polarised or ionised) end and can make contact with lipids and water at the same time. Thus a transmembrane segment has a well-defined position within the membrane and cannot bob up and down [15].

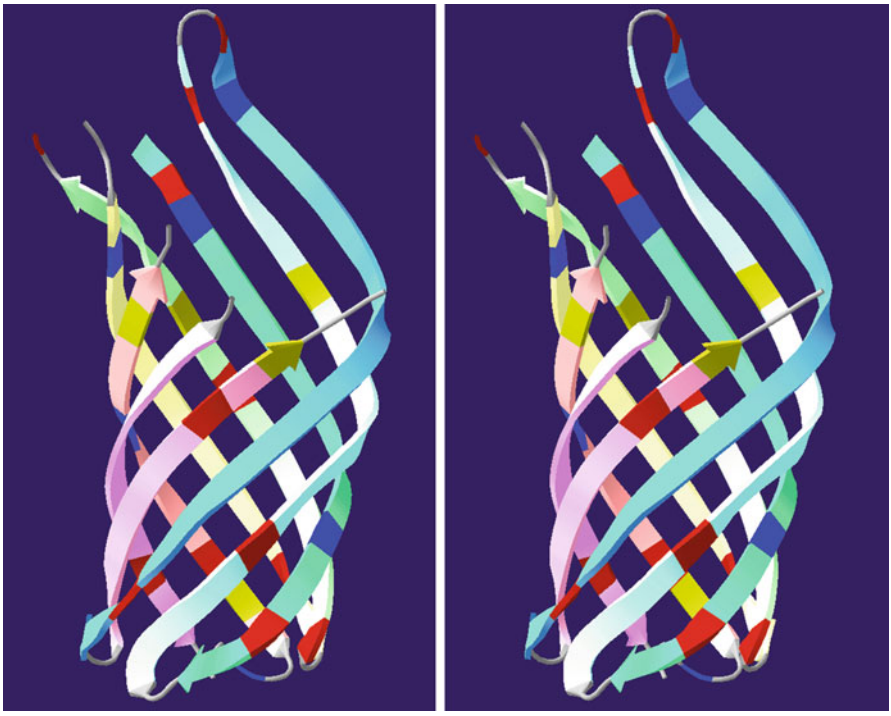


Fig. 2.21 Aromatic belt in a transmembrane protein, here outer membrane protein A (PDB-code 1qjp). Trp (*olive*), Tyr (*brown*) and Lys (*blue*) are marked; they occur preferentially at the membrane surface

Some proteins have several **domains**, that is, individually folding regions connected by short segments. These individual domains can be isolated by gentle proteolysis; they may maintain not only their structure, but even their catalytic function. For example, the chaperone Hsc70 (see Fig. 15.1 on page 346) has three domains: an ATPase-, a peptide-binding-, and a regulatory domain. Gentle treatment with chymotrypsin will digest the links between those domains. The isolated ATPase domain can still hydrolyse ATP.

2.3.1 Classification of Proteins by Folding Pattern

If one looks at many different proteins, one will find certain patterns in the way elements of secondary structures are arranged. These folding patterns are called **motives**. It is interesting to note that motives are much more stable during evolution than amino acid sequences. In other words, some proteins can be shown to be homologous by their folding patterns, even though they no longer have significant similarity in their amino acid sequence (e.g., the muscle protein actin, the enzyme hexokinase, and the chaperone Hsc70).

According to their folding pattern, protein domains may be hierarchically classified into groups. Because such classification is somewhat subjective, different schemes have been suggested. One commonly used scheme is the **Structural Classification of Proteins (SCOP)** database (since 2014 SCOP2, http://scop2.mrc-lmb.cam.ac.uk/). Classification at present cannot be done automatically, but requires expert knowledge. The following taxa are used in SCOP (see also figs. 2.22 and 2.23):

Class Coarse classification according to the relative content of α-helix and β-strand.

Fold Major structural similarity, the proteins have identical secondary structure elements (at least in part) and the same topological connections. However, there may be considerable variation in peripheral regions of a domain. Similarities may arise from common origin or from convergent evolution.

Superfamily Domains have a common folding pattern and their functions are similar, but sequence identity may be low. The ATPase domains of actin, hexokinase, and Hsc70 are an example for a superfamily. Common evolutionary origin is probable.

Family Proteins with high sequence homology (> 30 % identity) and/or similar function. Proteins clearly have an evolutionary relationship. Identical proteins are subclassified by species.

For a basic understanding of protein folding only the class is relevant:

all-α Proteins which contain only α-helices, or where the content of β-strands is at least insignificant.

all-β Proteins which contain only β-strands, or where the content of α-helices is at least insignificant.

α/β Proteins which contain alternating or interspersed α-helices and β-strands. Mainly parallel β-sheets (β-α-β units).

$\alpha + \beta$ Proteins which contain segregated α-helices and β-strands. Mainly anti-parallel β-sheets.

Small proteins Usually dominated by metal ligand, hæme, and/or disulphide bridges

intrinsically disordered proteins See Sect. 10.1 on page 203 for a description of this class.

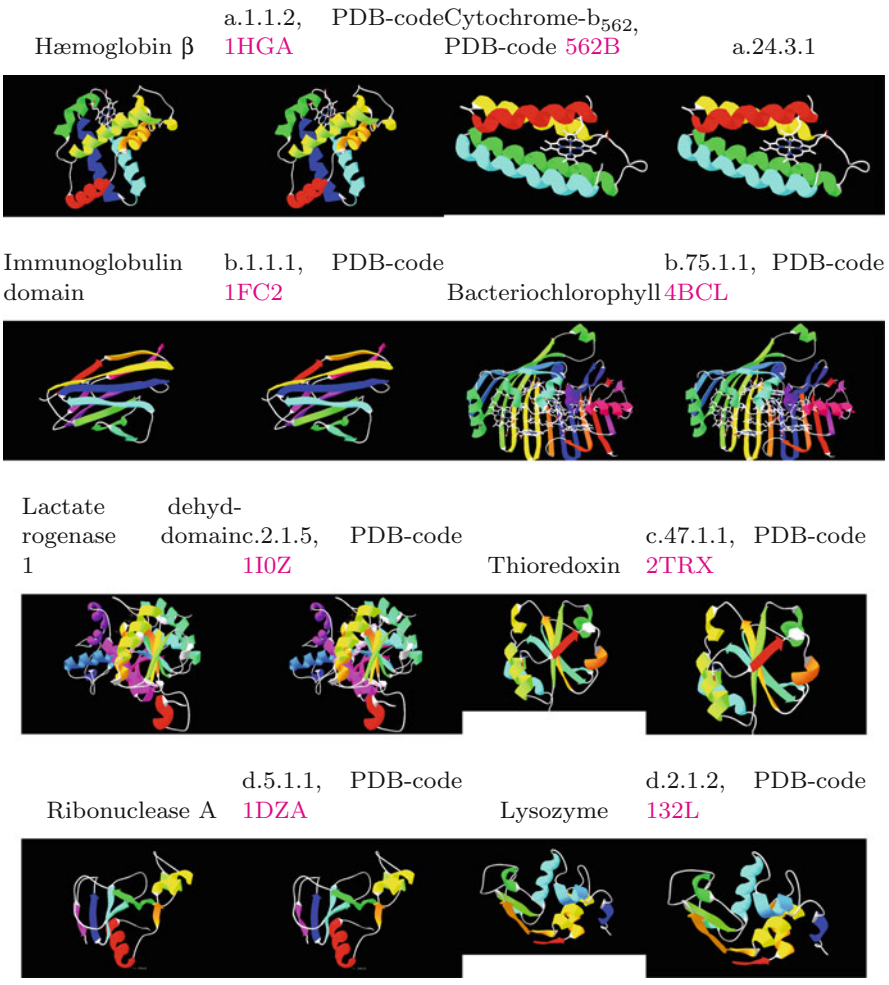


Fig. 2.22 Examples for $\alpha, \beta, \alpha/\beta$ and $\alpha + \beta$ protein structures

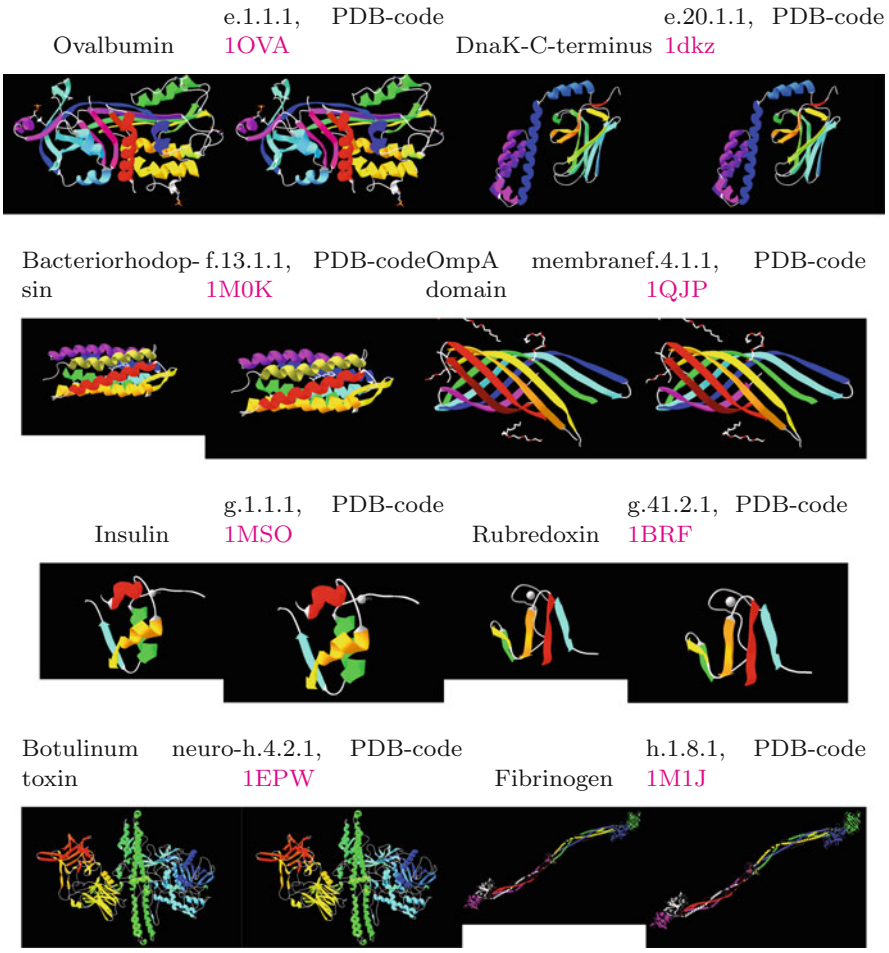


Fig. 2.23 Stereo views of multidomain and membrane proteins, small proteins, and coiled-coils

Apart from SCOP there are also other approaches for protein classification, in particular CATH (http://www.cathdb.info) and FSSP (http://ekhidna.biocenter.helsinki.fi/dali/), but these yield largely similar results [17].

Proteins can be described by a **set of concise classification strings (sccs)** according to their structure, for example, b.2.1.1 (class b = all β, fold 2 = $NAD(P)^+$-binding ROSSMANN-fold domains, superfamily 1 = Alcohol dehydrogenase-like and family 1 = Alcohol dehydrogenase). Within families, proteins are sorted by species and isoform (Table 2.1).

Table 2.1 Number of entities in the SCOP-database (02/2009, entering data into SCOP2 is not yet complete)

Class	Folds	Superfamilies	Families
a) All alpha proteins	284	507	871
b) All beta proteins	174	354	742
c) Alpha and beta proteins (a/b)	147	244	803
d) Alpha and beta proteins (a+b)	376	552	1055
e) Multidomain proteins	66	66	89
f) Membrane and cell surface proteins	58	110	123
g) Small proteins	90	129	219
Total	1195	1962	3902

Internet Resources
Protein structures are stored in the Brookhaven Protein Data Bank (PDB) in a unified format that can be used by modelling software such as DeepView (formerly known as Swiss-PDB, http://www.expasy.ch/spdbv/mainpage.html). Coordinates may be obtained from PDBlite (http://oca.ebi.ac.uk/oca-bin/pdblite), OCA(http://bip.weizmann.ac.il/oca-bin/ocamain), PDBsum (http://www.ebi.ac.uk/pdbsum/), or, if the EC-number is known, from http://www.ebi.ac.uk/thornton-srv/databases/enzymes/.

PDBTM (http://pdbtm.enzim.hu/?) deals with membrane proteins. Three-dimensional structures of nucleic acids may be retrieved from NDB http://ndbserver.rutgers.edu/.

The protein structures presented in this book were created with DeepView using data files obtained from OCA.

2.4 Quaternary Structure

Quaternary structure describes how several polypeptide chains come together to form a single functional protein. As with tertiary structure it is determined by ionic and hydrophobic interactions between amino acid R-groups. Many proteins consist of several subunits. Depending on the number of subunits, we speak of monomers, dimers, trimers, and so on. Depending on whether these subunits are identical, we put homo- or hetero- in front. Thus a heterodimer is a protein consisting of two different polypeptide chains. As we will see later, association of several subunits

into a protein has important consequences for its function, which is often lost if the subunits are separated (see chapter 7 on page 163).

In some proteins several polypeptides come together to form a subunit, which repeats several times. Such subunits are called **protomers**. For example, hæmoglobin is a diprotomer; each protomer consists of an α- and a β-chain (see Fig. 7.2 on page 166).

2.5 Further Aspects of Protein Structure

2.5.1 LEVINTHAL*'s paradox:*

Assume a protein with 100 peptide bonds, each of which can assume 6 stable conformations (α-helix, $\beta_{\uparrow\uparrow}$-sheet, $\beta_{\uparrow\downarrow}$-sheet, P_{II}-helix, turn, coil).

Because each of these states is characterised by a ϕ, ψ-angle pair, this results in $2^6 = 64$ possible angles per peptide bond and $100^{64} = 10^{128}$ for the entire protein (note that this is an underestimate!).

Rotation around a σ-bond takes about 10^{-13} s, thus folding by random testing of all possible angles would take $10^{128} \times 10^{-13}\,\text{s} = 10^{128-13}\text{s} = 10^{115}\text{s}$. Our best estimate for the age of the universe is 13.7×10^9 a $\approx 4.32 \times 10^{17}$ s. Proteins therefore should never fold. You now also understand why it is so difficult to calculate protein structures *ab initio*.

In reality, folding is a rapid process; in *E. coli* at 37 °C a 100 amino acid protein folds in about 5 s. During folding hydrophobic residues are buried in the interior and hydrophilic residues appear on the outside of the protein, resulting in a compact "**molten globule**" structure. This brings amino acids so close to each other that the formation of hydrogen bonds between peptide bonds gives rise to secondary structure.

The conformational freedom for protein folding is, however, much smaller than it might appear at first sight. We have already discussed steric hindrance between the atoms of neighbouring amino acids, which lead to large nonpermissible areas in the RAMACHANDRAN plot. But steric hindrance is also possible between amino acids farther along the protein chain, unless the protein is in an extended conformation (upper left hand quadrant in the RAMACHANDRAN-plot) [24]. Thus it is not possible to have a β-strand directly following an α-helix (or *vice versa*) without an intervening coil. These restrictions no doubt explain the limited number of structural motives found in proteins.

Thermodynamics of LEVINTHAL's Paradox

ZWANZIG *et al.* have looked at LEVINTHAL's paradox in thermodynamic terms [43]. If one assumes that folding is the process of converting n bonds between $n + 1$ amino acids from an incorrect (i) to the correct (c) conformation, then the reaction $\mathsf{c} \underset{k_1}{\overset{k_0}{\rightleftharpoons}} \mathsf{i}$ is described by the following differential equation:

$$\frac{d[\mathsf{c}]}{dt} = -k_0[\mathsf{c}] + k_1[\mathsf{i}], \qquad [\mathsf{c}] + [\mathsf{i}] = \mathsf{n} \tag{2.1}$$

k_0 and k_1 are the rate constants for unfolding and folding, respectively; these rate constants are related by the law of mass action:

$$\frac{[\mathsf{i}]_{eq}}{[\mathsf{c}]_{eq}} = \frac{k_0}{k_1} = K \tag{2.2}$$

Note that K is a thermodynamic property and completely independent of the processes that lead to folding and unfolding.

If there are ν incorrect states of a bond (the number of correct states is of course 1) and if a correctly folded bond has a free energy of ϵ and an incorrectly folded bond a free energy of $\epsilon + U$, it can be shown from statistical mechanics that

$$K = \frac{k_0}{k_1} = \nu e^{-\frac{U}{kT}} \tag{2.3}$$

If the energy penalty for the misfolded state $U = 0$, then the average mean first passage time τ required to arrive for the first time at $[\mathsf{c}] = \mathsf{n}$ (all bonds correctly folded) becomes

$$\tau = \frac{1}{\mathsf{n}k_0}(\nu + 1)^{\mathsf{n}} \tag{2.4}$$

which is a mathematical statement of LEVINTHAL's paradox since $(\nu + 1)^{\mathsf{n}}$ is the number of possible states of the protein. If, however, there is an energy penalty for misfolding, then the first passage time becomes much shorter. Even for $U = 2kT$ per bond (where k is BOLTZMANN's constant and T the absolute temperature, kT is the average kinetic energy of atoms at that temperature) τ is on the order of seconds!

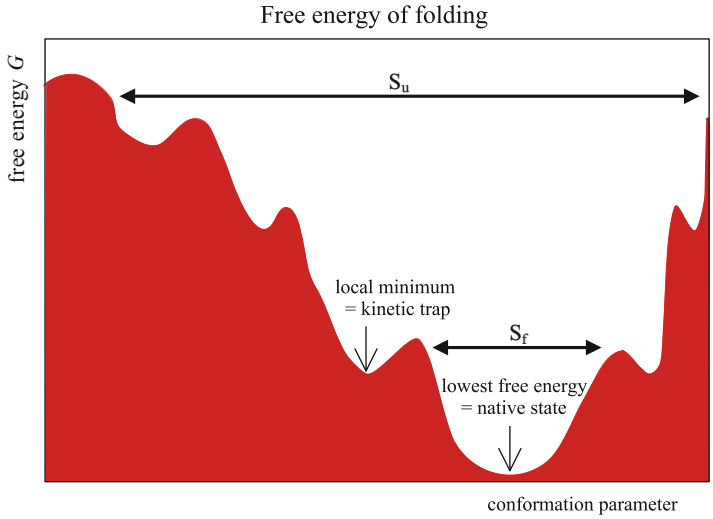


Fig. 2.24 Protein folding reduces free energy (G). The native structure is the one with the lowest free energy. However, proteins may get kinetically trapped in local minima of the energy landscape. During folding the entropy of the unfolded protein (S_u) is reduced to that of the folded (S_f), symbolised by the width of the funnel. This entropy reduction (more orderly, less probable state) reduces the overall change in free energy of the folding process. This is shown here for a two-dimensional reaction, but in protein folding each amino acid can adjust at least ϕ and ψ, so the number of dimensions is impressive and it is not surprising that no way to calculate the 3D-structure of a given protein sequence from first principles is known

2.5.2 *Energetics and Kinetics of Protein Folding*

Amino acids in a folding protein have a choice of undergoing interactions with either other amino acids or with water. Thus only the difference $\Delta G_{\text{folding}} = \sum \Delta G_{a-a} - \sum \Delta G_{a-w}$ is available to stabilise the native structure of proteins (see Fig. 2.24). Although folding decreases the enthalpy (H) (which stabilises native structure), it also decreases the entropy (S), which tends to destabilise the native structure. Thus protein folding is a compromise between forces, and the actual stabilisation energy is only about 20–40 kJ/mol, about 10× the thermal energy at room temperature (kT for individual molecules or $RT \approx 2.5$ kJ/mol) [43]. This **marginal stability** of proteins has a good side, however: it allows protein flexibility required for ligand binding and enzymatic activity (see fig. 2.25).

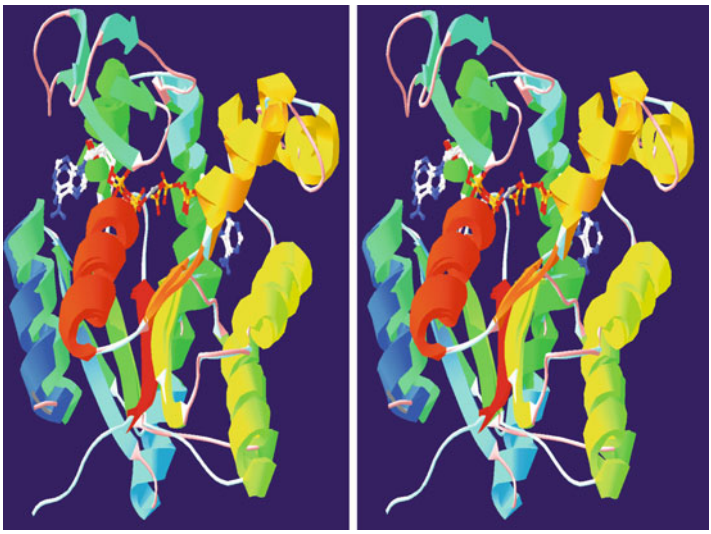
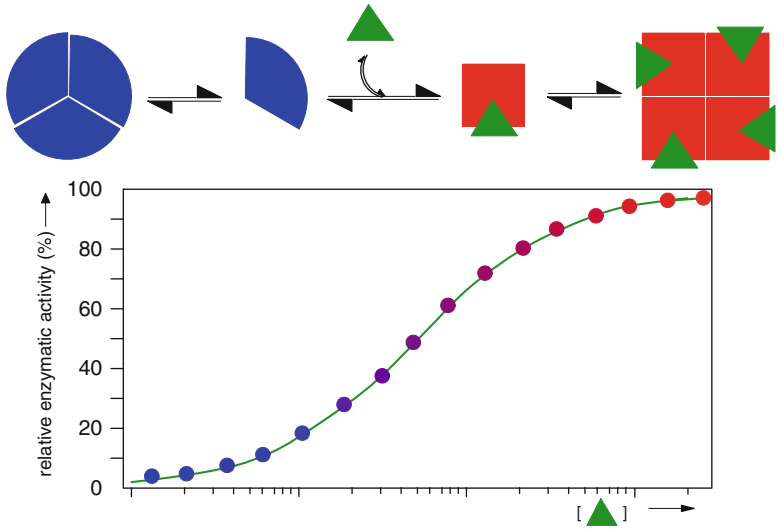


Fig. 2.25 Movement in adenylate kinase during substrate binding [23]. Note how α-helices and β-sheets provide an overall stable structure, with coils acting as hinges. Viewing this video (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3369945/bin/pcbi.1002555.s012.mov) requires a connection to the Internet. The still image shows a superposition of the enzyme with either 2 ADP (PDB-code 2cdn) or with AMP and the nonhydrolysable ATP-analog AMP-PNP (PDB-code 1ank) bound

2.5.3 Morpheeins

Morpheeins are proteins that form different homo-oligomers according to the following schema:



These oligomers may have different enzymatic properties. DOSS **(PBGS)-porphyria** is an example of a disease caused by a mutation that shifts the equilibrium between oligomers of an enzyme. Such diseases may in future be treated by small molecules, which increase the likelihood that the protein occurs in the desired conformation. Do not mix up morpheeins (proteins that morph) with the drug morphine (from Gr. MORPHEUS = god of sleep).

2.5.4 Molecular Chaperones and Chaperonins

Although protein folding is a spontaneous process driven by thermodynamic forces and hence, given enough time, all proteins will eventually arrive at their native structure, kinetically the process can be trapped in local minima in the energy landscape (see Fig. 2.24). Such **metastable** intermediates would expose hydrophobic patches on their surface, which leads to protein aggregation. As opposed to quaternary structure, such aggregates have no reproducible structure and no biological function. On the contrary, they may interfere with cellular function (see Sect. 10.2 on page 206). Note that proteins in the cytosol are very closely packed at about 300–400 mg/ml. The average distance of protein molecules is just one protein diameter, and the space between them is filled with water, salts, and metabolites. This close packing increases the probability of unwanted interactions.

Cells have two lines of defence against misfolded proteins:

Molecular chaperones bind to unfolded proteins and prevent their aggregation until these proteins can achieve folding [37]. Binding/unbinding cycles of chaperones may or may not require the hydrolysis of ATP. Examples: Hsp90, Hsc70, crystallins.

Molecular chaperonins use the energy of ATP-hydrolysis to unfold misfolded proteins actively, giving them a second chance to arrive at the proper fold. Example: GroES/GroEL. Unfolding occurs in a "beaker" formed by the chaperonin [32], in which the client protein can try to refold without disturbance from other proteins (at "infinite dilution"). The beaker has a diameter of about 45 Å, enough to contain proteins (or protein domains) of up to 60 kDa.

Note: Neither chaperones nor chaperonins actively fold proteins; they merely protect them against aggregation during the folding process. For a more detailed discussion of this topic see Chap. 15 on page 343.

2.5.5 Protein Denaturation

Secondary, tertiary, and quaternary structures of a protein are determined by relatively weak interactions, such as hydrogen bonds or hydrophilic interactions. The energy of a hydrogen bond is only about 4 kJ/mol, which is low compared to

the energy of thermal motion (RT, at room temperature $\approx$ 2.5 kJ/mol). Changes in environmental conditions can break such bonds, leading to the denaturation of proteins.

Strong acids and bases denature proteins by disrupting ionic interactions.

Organic Solvents can denature proteins by disrupting hydrophobic interactions. Proteins are not soluble in organic solvents. More water soluble solvents (e.g. ethanol or acetone) bind water and thus reduce the concentration of water available to the protein.

Detergents disrupt hydrophobic interactions. They can denature proteins without precipitating them.

Salts precipitate proteins because they reduce the concentration of water available to maintain protein structure.

Small hydrophilic substances such as urea can denature proteins when they are present in high concentration, both by binding water and by binding to the protein.

Heavy metal ions (lead, mercury) bind to carboxylate or sulphydryl groups of proteins. That's why they are toxic!

Heat An increase in temperature leads to increased molecular motion. This can result in breaking hydrogen bonds. If some hydrogen bonds break, the structure of the protein (say, an α-helix) is weakened; that is, other hydrogen bonds become easier to break. Denaturation by increasing temperature therefore is a process that starts quite suddenly at a certain critical temperature and is completed at a temperature only marginally higher. Renaturation is sometimes possible with small proteins (ribonuclease, lysozyme) under laboratory conditions, but denaturation is irreversible in the real world (boiled egg).

> Humans die if their core body temperature exceeds 42 °C, when key proteins lose their function. For a useful application of thermal protein denaturation see Fig. 2.26.

The covalent bonds in proteins are more robust, but peptide bonds are hydrolysed by heating in strong acids and bases, and by proteolytic enzymes. Disulphide bonds are cleaved by reducing agents; oxydizing agents can form disulphide bonds from −SH-groups.

2.5.6 Protein Folding

Some proteins, such as ribonuclease, can be denatured completely by heat or the addition of denaturants. If the solution is cooled, or if the denaturant is removed by dialysis, the protein spontaneously goes back to its native conformation, and

Fig. 2.26 Lionfish (here *Pterois volitans* L. 1758) produce a toxic slime on the tip of the rays of their fins; touching them is intensely painful. The active ingredient is a mixture of proteins that also causes high blood pressure, vomiting, and interferes with breathing. Envenomation with lion fish toxin is treated by exposing the affected body part to hot water (45–50 °C, so hot that it is just bearable). The water is frequently changed for about 30 min. This denatures the toxic proteins, bringing almost immediate relief

enzymatic activity is restored. These experiments, carried out first by C. ANFINSEN in the 1950s [2], prove that the secondary and tertiary structure of a protein is determined solely by its amino acid sequence. No external information is necessary.

This has an important consequence: A mutation of only a single amino acid may interfere with protein folding, resulting in loss of function. For example, **cystic fibrosis** is a disease caused by lack of a transmembrane chloride channel (cystic fibrosis transmembrane conductance regulator, CFTR). See page 449 for a discussion of the pathomechanism of cystic fibrosis.

Another example would be collagen, where single amino acid mutations lead to osteogenesis imperfecta or EHLERS-DANLOS-syndrome (see page 324). Protein folding problems as cause of disease have been reviewed in [34].

There are essentially three models to describe protein folding:

Framework model Protein adapt their secondary structure first, using the information contained in their sequence. Once the secondary structure is established, long-range interactions between amino acids stabilise the tertiary structure.

Hydrophobic collapse After synthesis proteins adapt a conformation where hydrophobic amino acids are buried inside and hydrophilic amino acids are on the surface of the protein. This structure is called "**molten globule**". Inside the

molten globule long-range interactions between amino acids are established, allowing the tertiary structure to form. The secondary structure is established last.

Nucleation/Condensation This model is a combination of the other two: proteins collapse under formation of bonds between key residues. This reduction in the distance of amino acids leads to the formation of long-range (tertiary structure) and short-range (secondary structure) interactions at the same time. As a consequence, folding occurs in one big step without folding intermediates. For many proteins, especially those with < 100 amino acids, this agrees well with experimental evidence. Large proteins often consist of independently folding modules (**domains**), whose folding rates can vary considerably. These modules then have to find together in the final structure of the protein.

Folding is a fairly fast process, synthesis and folding of a 100 amino acid protein in an *E. coli* cell is complete in less than 5 s (at 37 °C).

Folding starts even while a protein is synthesised on a ribosome, that is when only a part of the protein molecule is available for the folding reaction. But even when an entire protein is totally unfolded and then allowed to refold, some strong interactions form first and then determine which other interactions may form later. As LEVINTHAL [20] already noted, such specific folding pathways may lead to native states that are not at the absolute minimum of the free energy function (i.e., not at equilibrium), but are rather metastable states in local energy minima. As long as the activation energy for going to the absolute minimum is high enough, such metastable states can live long enough to be biologically functional.

If the activation energy is lowered, rapid misfolding may result, leading to **folding diseases**. This can be caused by mutations that increase the rate of formation of the lower energy states (e.g., HUNTINGTON chorea). These diseases are discussed in Sect. 10.2 on page 206.

Protein Structure and the Development of Pharmaceuticals

Knowledge of the structure of enzymes and receptors is essential for the design of pharmaceuticals. For example, human immunodeficiency virus (HIV) protease is required for splitting an inactive precursor protein in the virus envelope into two active proteins. Without these proteins, a virus particle cannot bind to its target cells and would be noninfectious. HIV-protease is a homodimer, each subunit is 99 amino acids long. They are held together by an antiparallel β-sheet formed by amino acids 1–4 and 96–99 of both subunits (see Fig. 2.27).

(continued)

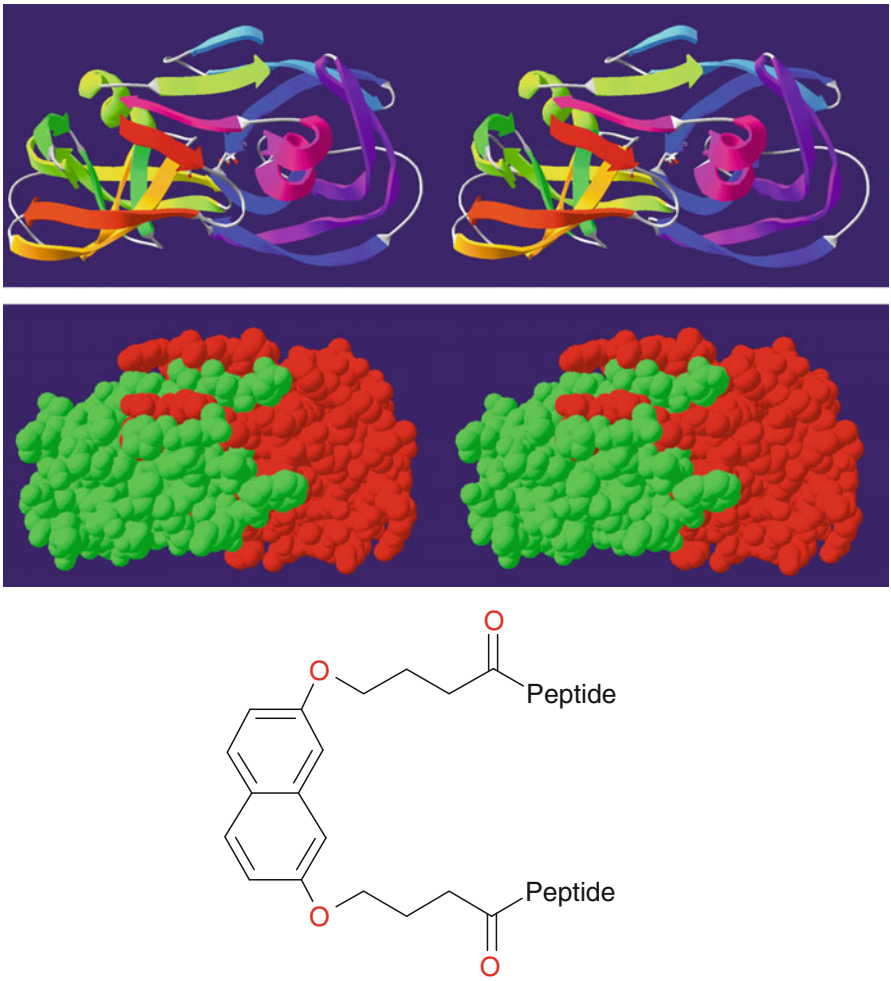


Fig. 2.27 *Top*: Dimerisation of HIV-protease (PDB-code 1DAZ) occurs by the formation of an antiparallel β-sheet from the ends of both subunits. This brings the two catalytic aspartate residues (D^{25} and $D^{25'}$) close together. *Middle*: Space filling model of HIV-Protease. The β-sheet formed by the N- and C-terminal ends of both subunits is clearly visible. *Bottom*: Substances with two peptides linked by a stiff backbone can interdigitate into the dimerisation site of a monomer and prevent dimerisation

This brings the catalytic aspartate residues (D^{25} in each subunit) together, thus forming the catalytic site of the enzyme. Several pharmaceuticals are on the market which bind to the catalytic centre of the enzyme, but these are beginning to lose their effectiveness due to the development of resistant virus strains. Also they are very hydrophobic compounds, which makes their pharmaceutical use difficult. A new class of protease inhibitors binds to the dimerisation site and prevents the formation of the active enzyme. Development of such substances requires an intimate understanding of the structure and function of an enzyme (see Fig. 2.27 and [4] for further details).

2.6 Posttranslational Modifications of Proteins

The human genome contains ≈ 23 000 genes [7, 8, 35]. mRNA-processing (alternative splicing, mRNA editing etc.) results in ≈ 3 mRNAs per gene (Fig. 2.28). Posttranslational modification of the proteins produced from them creates ≈ 10 different protein species from each mRNA. Thus the human proteome consists of ≈ 10^6 proteins, with different functions, regulation, destruction...

The properties of proteins can be changed by posttranslational modification; in some cases this can be done (or undone) quickly in response to environmental stimuli, for example, exposure to hormones. Such modifications can switch enzymes between active and inactive states and are required for the proper targeting of a protein to subcellular structures. The following reactions are of particular importance:

2.6.1 Glycosylation

Glycosylation is the process of enzymatic transfer of oligosaccharide (sugar) trees to proteins (see page 379 for a more detailed discussion of the process). They are affixed either to the OH-groups of Ser or Thr (**O-linked**) or to the acid amide group of Asn (**N-linked**). Other amino acids (Arg, Tyr, Trp, Hyl, Hyp) are involved much less frequently, e.g., in collagen. Addition occurs in the ER and the GOLGI-apparatus to the extracellular domain of membrane proteins and to secreted proteins. Cytosolic proteins are rarely glycosylated. In bacteria, glycosylation occurs in the periplasm.

Glycosylation is required for proper protein folding. Glycosylation inhibitors (see Fig. 2.29) are used as **antiviral drugs** (e.g. nojirimycin or desoxynojirimycin). Sugar trees are also required as "address labels" in the intracellular transport

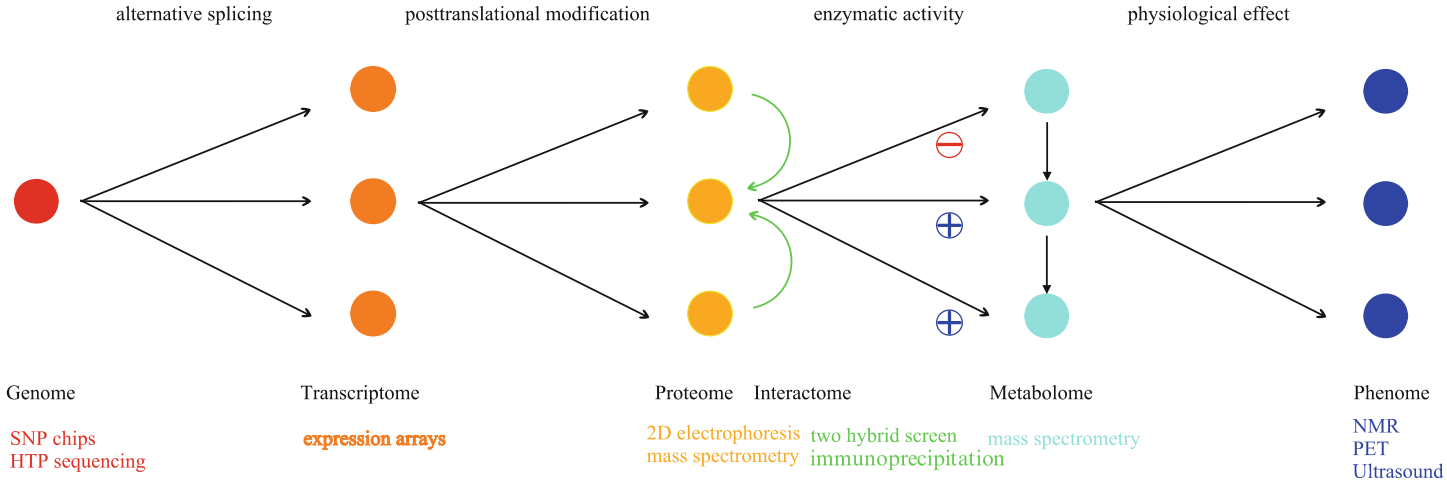


Fig. 2.28 The pathway from genome to phenome is studied at different levels with different methods. The resulting complex data can only be handled by advanced computing techniques

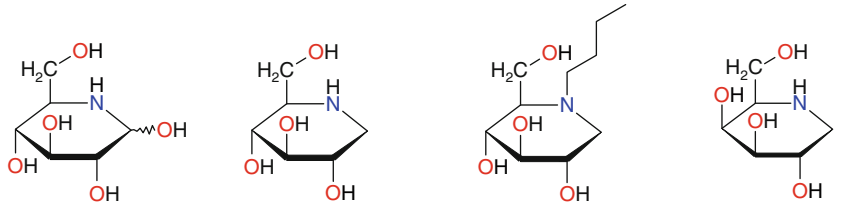


Fig. 2.29 Sugar-analogues where the aldehyde-group is replaced by $-NH_2$ act as glycosylation inhibitors. They can be used as antiviral drugs, and also in some inherited diseases (mucopolysaccharidoses), where the enzymes that degrade glycoproteins in the lysosomes do not work properly

of proteins between compartments. For example, in **I-cell disease** transport of lysosomal proteins into this compartment fails because an enzyme which transfers the sugar mannose-6-phosphate to them is defective. They are secreted into the bloodstream instead and, as a consequence, the lysosomes are nonfunctional.

On the cell surface, the sugar trees of membrane proteins serve as recognition sites for cell–cell-interactions, as immunological determinants and—since everything has to have a downside too—as docking sites for bacteria and virus.

2.6.2 Glucation

In glycosylation sugars are enzymatically transferred to proteins in a carefully orchestrated process. Glucation, however, is a spontaneous reaction between sugar aldehydes (and some other reducing compounds) and amino groups (SCHIFF-base formation, see Fig. 2.30) in proteins, nucleic acids, and lipids.

The velocity of glucation depends on the concentration of glucose in the blood. This has a direct medical application: In **diabetics**, the concentration of glucated hæmoglobin (**HbA_{1c}**) depends on the average blood glucose concentration during the lifespan of an erythrocyte (about 3 mo).

Glucated proteins can react further to **Advanced glucation end products (AGE)**; this is thought to be involved in **aging** and in long-term diabetic damage. There is a PAMP-receptor for AGEs, called RAGE, stimulation of which is proinflammatory, procoagulant, and promitotic. This is thought to be responsible for at least some of the damage (see, e.g., [12, 22, 26, 29–31]).

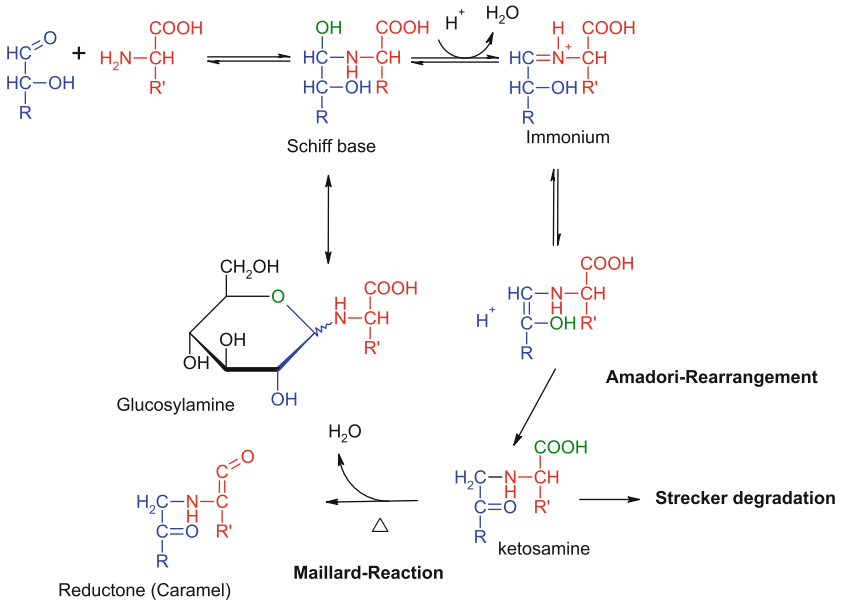


Fig. 2.30 Glucation of proteins by the aldehyde group of glucose proceeds via an unstable SCHIFF-base and AMADORI-rearrangement to a stable ketosamine. During roasting, this is converted into caramels via the MAILLARD-reaction. These are responsible for the taste of cooked food. Ketosamine may also be converted to Advanced glucation end products (AGE) by STRECKER-degradation

2.6.3 Disulphide Bond Formation

Oxydation of the SH-groups of two cysteine residues leads to the formation of a covalent bond. The cell uses the tripeptide glutathione (see Fig. 2.31) as reducing agent: $-SH + HS- + GSSG \rightleftharpoons -S-S- + 2GSH$. Disulphide bond formation does not happen in the reducing environment of the cytosol, but in the ER (or the bacterial periplasm) which is oxydising. Special enzymes, **Protein disulphide isomerase (PDI)**, make sure that the right Cys residues undergo disulphide bond formation.

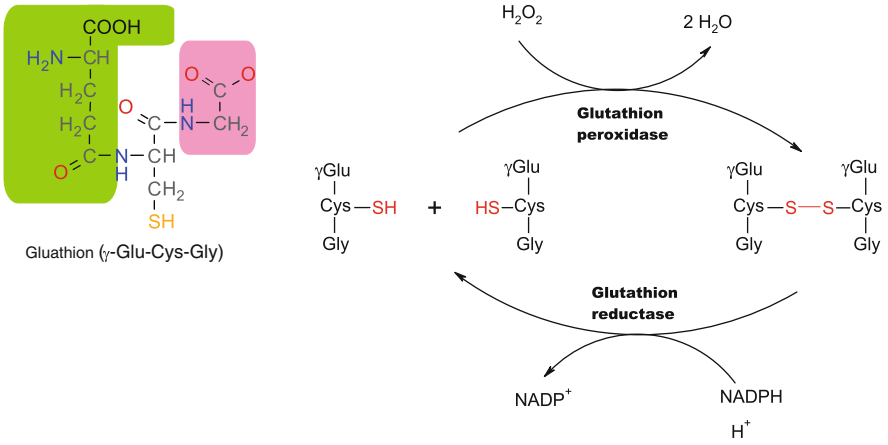


Fig. 2.31 The tripeptide glutathione serves as a redox-coupler in our cells. *Left*: Structure of glutathione. *Right*: Coupling of detoxification of reactive oxygen species (ROS, here H_2O_2) and consumption of NADPH + H^+ by glutathione

When cytosolic proteins are used in the laboratory one has to make sure that their SH-groups are not oxydised by air oxygen, which would lead to inactivation. The buffers therefore usually contain an antioxydant such as β-mercaptoethanol or dithiotreitol.

Bacterially expressed eukaryotic proteins are often misfolded and precipitate as **inclusion bodies** because bacteria are less active in disulphide formation than eukaryotes. However, bacteria do have an enzyme operon (Dsb, short for disulphide bond) for formation and isomerization of protein disulphide bonds in their periplasm.

Some mucolytic pharmaceuticals, including **N-acetylcysteine (ACC)**, work by breaking S–S-bonds in mucus proteins, decreasing the viscosity of mucus and making it easier to clear it from the airways.

2.6.4 Proteolysis

Proteolysis is involved in the activation of proenzymes, for example, in the digestive system. Digestive enzymes (e.g. trypsin, chymotrypsin) are produced as inactive precursors (zymogens), so that they cannot harm the cells secreting them. Once

released into the intestine, they are activated by cleaving off a part of the enzyme that was blocking the active site. Cascades of proteolytic enzymes make up our blood-clotting and complement system (see Sect. 11.3 on page 249). Prohormones (e.g. insulin) are activated in a similar manner. On the other hand, proteins no longer needed can be inactivated by proteolysis (e.g., cyclins in cell cycle).

Proteolysis may also be used to remove signal peptides. For example, some proteins destined for the intermembrane space of mitochondria carry a signal sequence for mitochondrial import (see Fig. 2.10) which leads to their import into the mitochondrial matrix. There the signal peptide is cleaved off by matrix protease, exposing a second signal directing the protein's export into the intermembrane space through a different transporter.

A special form of proteolysis is **protein splicing** [25]. This reaction is carried out by a protease within the protein itself, the **intein**. This protease cuts itself out of the protein and rejoins the flanking segments (**exteins**), and all this without requiring any external proteins, cofactors or sources of energy such as ATP! The intein protein, once cut out of the host protein, has endonuclease activity. Intein genes are mobile elements (parasitic DNA); the corresponding mRNA can be used to direct the synthesis of cDNA by reverse transcriptases encoded by retrovirus inside the cellular DNA. This cDNA then is integrated into genes of other proteins by the endonuclease activity of the intein. As the intein cuts itself out of that protein, this insertion has little negative consequences for the host. Inteins are therefore the smallest possible parasites [13, 14]. They are now used as self-cleaving affinity-tags to make protein pharmaceuticals.

2.6.5 Hydroxylation

Protein hydroxylation occurs on Pro and Lys residues. We have already discussed the importance of Pro-hydroxylation for collagen formation (see page 28).

Proteins regulated by Pro-hydroxylation are the **hypoxia induced transcription factors (HIF)** [16]. These consist of two subunits, α and β. In the presence of oxygen, the α-subunit is hydroxylated on P402 and P564 by HIF-prolyl hydroxylases (PHD-1, -2 and -3, EC 1.14.11.29), leading to their proteasomal destruction. In the absence of oxygen, the α-subunits accumulate and form a complex with β, which binds to **hypoxia response elements** in the cellular DNA. As a consequence, oxygen consumption of the cell is down-regulated; it can survive a low oxygen supply for a longer time.

This mechanism may one day be exploited to increase the survival time of organs in infarct or transplantation, e.g., with inhibitors of PHD-1 (currently available inhibitors produce too many side effects due to concomitant inhibition of PHD-2 and -3).

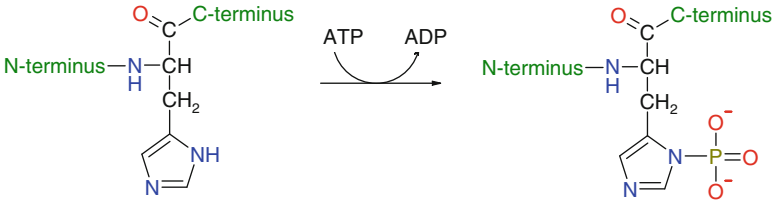


Fig. 2.32 Phosphorylation of histidine residues in His-kinases. Although mammals do not use regulatory phosphorylation of His, bacteria and fungi regulate the expression of pathogenicity factors that way. Thus, His-kinases may become important drug targets

2.6.6 Phosphorylation/Dephosphorylation

The transfer of phosphoryl groups from ATP to the hydroxy groups of Ser, Thr, and Tyr (rarely onto His-nitrogen or Asp and Glu COO^-) is important for the **reversible** regulation of enzyme activity. The transfer is catalysed by **protein kinases**, and the removal by **protein phosphatases**. Thus the reaction is rapidly reversible at minimal expense for the cell (a single high energy phosphate bond). One-third of all proteins in the cell undergo regulatory phosphorylation/dephosphorylation cycles (Fig. 2.32).

2.6.7 Acetylation/Deacetylation

Transfer of acetyl groups from acetyl-CoA onto the ϵ-amino group of Lys by **protein acetylases**, and their removal by **protein deacetylases**, are also used for regulation of enzymatic activity. The human acetylome, containing 1750 proteins, has recently been determined [5]. Many DNA-binding proteins are regulated by acetylation, because the acetylated Lys is much less likely to be protonated, hence less likely to bind to the negative charges on DNA. In addition, the change in protonation also affects the binding of transcription factors. The activity of metabolic enzymes may also be regulated by Lys-acetylation, for example, the glycolytic activity of Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is increased by acetylation; gluconeogensis is stimulated by deacetylation.

> This is interesting as the metabolism of cancer cells is shifted towards glycolysis (WARBURG effect); it makes acetylases and deacetylases targets for anticancer drugs.

Three classes of deacetylases are known: class I and II hydrolyse the bond with water, whereas class III deacetylases (**sirtuins**) use NAD^+ (see Fig. 2.33), thus their activity depends on the nutritional status of the cell [11, 33, 38, 42]. This is probably the mechanism behind the observation that mild **caloric restriction** prolongs the life expectancy of lab animals.

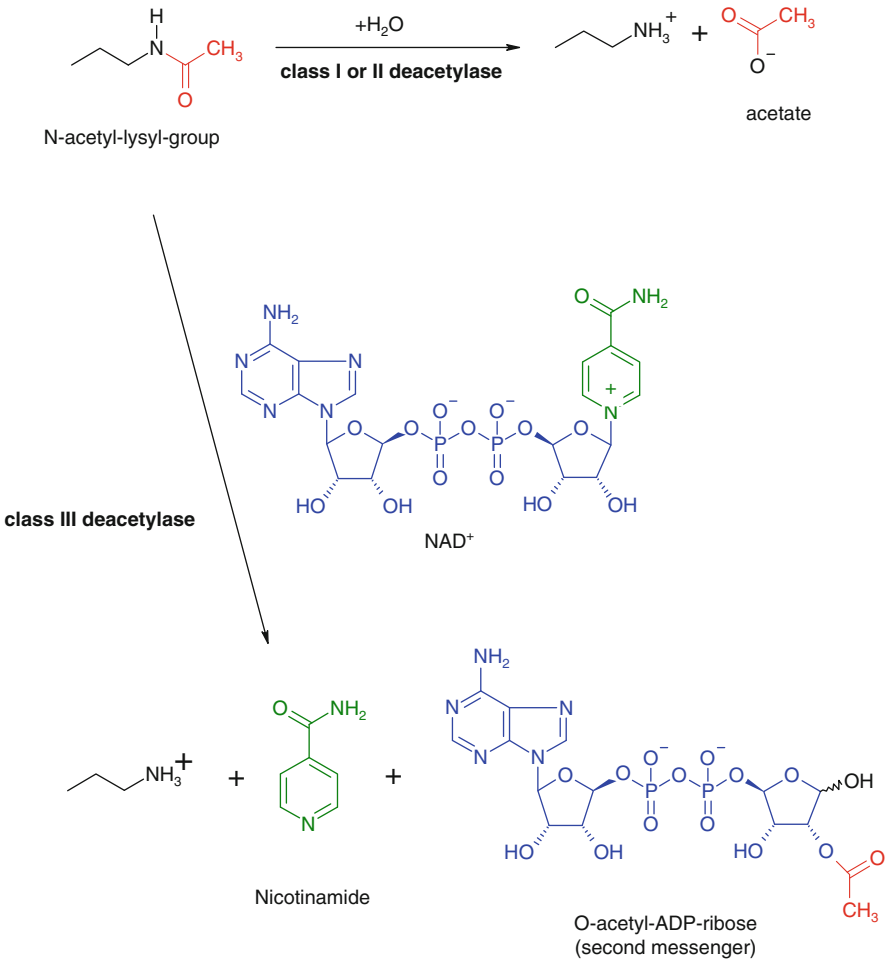


Fig. 2.33 There are three classes of deacetylases. Enzymes of classes I and II simply hydrolyse the amid bond. However, enzymes of class III (sirtuins) use NAD^+, which is available in the fasting state, but converted to $NADH + H^+$ in the fed state. Thus sirtuins regulate gene expression depending on the energy available to the cell

2.6.8 Methylation/Demethylation

Transfer of methyl groups from S-adenosyl methionine (SAM) onto proteins may also serve regulatory purposes, but we know very little about it [6]. Transfer can be to

carboxyl groups, forming methyl esters. This reaction is used to mark damaged proteins for destruction, but also in signal cascades of unknown function.

amino groups, forming methyl amines. The function is unknown. No N-demethylases are known, so the modification is probably permanent.

thiol groups, forming thioesters. The function is unknown.

Unlike phosphorylation, methylation has never been observed to occur on hydroxyl groups.

2.6.9 Addition/Removal of Hydrophobic Tails

Addition of

palmitoyl- (fatty acid) groups to internal Cys or Ser
myristoyl- (fatty acid) groups to N-terminal Gly
farnesyl- or geranylgeranyl (isoprenoid) groups to C-terminal Cys

converts cytosolic enzymes to membrane-bound (cytosolic leaflet). Because this is required for the activation of some enzymes, the transferases make possible drug targets (e.g., anticancer drugs).

2.6.10 S-Nitrosylation

S-Nitrosylation (see Fig. 2.34) occurs on Cys-residues: $R{-}SH + NO\cdot \rightleftharpoons R{-}S{-}^{\cdot}N{-}OH$. The Cys must be positioned between a basic and an acidic residue (either in primary, tertiary, or quaternary structure) because of acid-base catalysis. Nitrosylation serves as an additional pathway for $NO\cdot$ regulation besides cGMP-dependent kinases. This pathway too has not been fully explored.

2.6.11 ADP-Ribosylation

ADP-Ribosylation (see Fig. 2.35) is used by some bacterial toxins (*Vibrio cholerae, Bordetella pertussis*) to inactivate cellular proteins. This is the starting point of the patho-mechanism of the diseases associated with these bacteria (**cholera** and **whooping cough**, respectively).

2.6.12 Deamidation

Deamidation is the removal of the acid amide group from Gln or Asn, forming Glu and Asp, respectively. It may be followed by racemization (formation of D-amino acids).

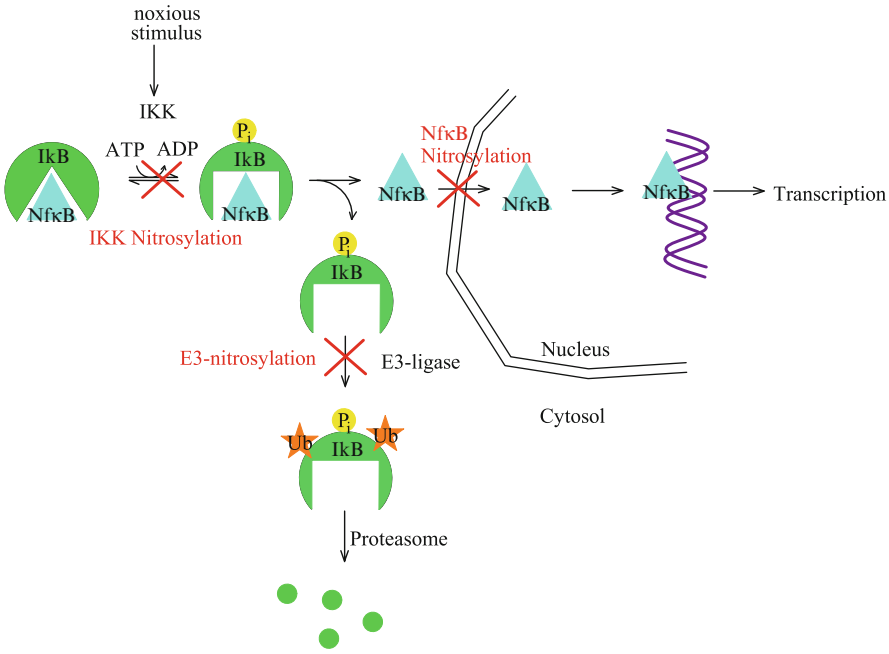


Fig. 2.34 Regulation of a pathway by multiple posttranslational modification events. Upon noxious stimuli, the **Inhibitor of NF-κ B (Iκ B)** protein is phosphorylated by **Inhibitor of NF-κ B kinase (IKK)** and releases **Nuclear factor κ B (NF-κ B)**. The latter enters the nucleus and changes DNA expression, whereas the phosphorylated Iκ B is ubiquitinated and thus marked for destruction by the proteasome. Nitrosylation of IKK prevents the phosphorylation of Iκ B, nitrosylation of NF-κ B its transport into the nucleus, and nitrosylation of a ubiquitin ligase subunit the transfer of ubiquitin to phosphorylated Iκ B

This has medical implications:

catalytic removal by bacterial pathogenicity factors (**cytotoxic necrotising factors**) on heterotrimeric G-proteins and small GTPases → GTPase activity is inhibited; the protein cannot go from the active GTP-bound to the inactive GDP-bound form.

spontaneous (Asn faster than Gln): age determination in

forensic science long lived, low turnover proteins (bone, teeth)

archaeology rate constant depends on temperature, humidity, soil *p*H etc, hence better suited for determination of relative age within a series of finds.

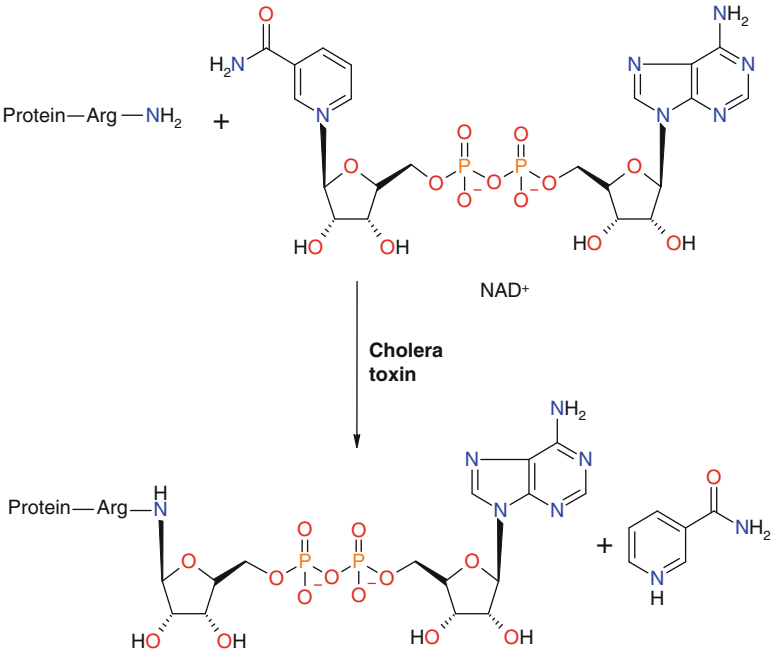


Fig. 2.35 ADP-ribosylation of proteins

2.6.13 AMPylation (Adenylylation)

AMPylation (Adenylylation) is performed from ATP onto critical Thr hydroxy groups of Rho GTPases and other important regulatory proteins by *Vibrio parahaemolyticus* **Vibrio outer protein S (VopS)**. This results in depolymerisation of the actin cytoskeleton, affected cells round up. As in ADP-ribosylation the bacterial toxin uses a readily available energy-rich substrate to disable critical host proteins. Just as in the ADP-ribosylation from NAD^+ or glycation from UDP-glucose AMPylation from ATP is performed by a bacterial **A/B toxin** to interfere with host defences. AMPylation, however, is (at least in bacteria) also used for control of metabolism: the glutamine synthetase of *E. coli* is controlled in part by AMPylation of Tyr-397; the adenylyl transferase (adenylic acid = old name for AMP) responsible in turn is controlled by **uridilylylation**.

2.6.14 Transfer of Peptides

Ubiquitin, a 8.6 kDa protein (see Fig. 3.17 on page 88) is transferred to proteins by a group of ubiquitin-ligases, of which three classes exist. **E1-ligase** (UbA1) (see Fig. 2.36) forms a thioester bond with the C-terminal glycine residue of ubiquitin in an ATP-dependent reaction. This activated ubiquitin is then transferred to an

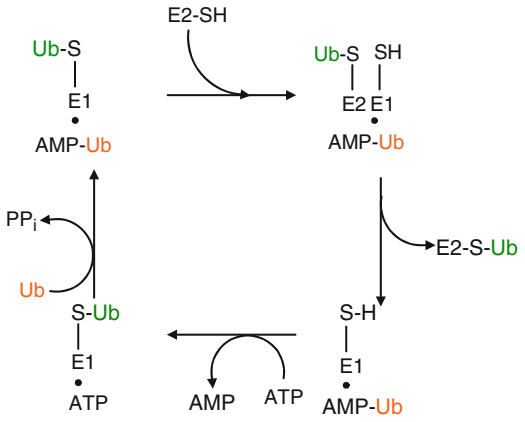


Fig. 2.36 The E1-ubiquitin ligase binds two molecules of ubiquitin: one is covalently bound to a thiol residue of the enzyme, the second is bound to AMP. The first ubiquitin molecule is transferred to the E2-ubiquitin ligase, then the second ubiquitin is moved to the SH-group, and the AMP is exchanged for ATP. Then another ubiquitin is bound to the nucleotide, pyrophosphate is released in the process, and the enzyme is ready for a new cycle

E2-ligase (UbCs) and from there to an **E3-ligase**. All three ligases bind ubiquitin as thioester. There is only one (or, in some species, a few) E1 (UbA1) but several UbCs and many E3-ligases, which are often specific for a single target. Ubiquitin is usually transferred to the ϵ-amino group of a lysine, forming an isopeptide bond. Binding to the N-terminus, or to Cys, Ser, and Thr has been described, but is rare. Ubiquitin contains seven functionally distinct Lys-residues, whose ϵ-amino groups form isopeptide bonds with the C-terminal Gly of other ubiquitin molecules, resulting in long chains of poly-ubiquitin. Protein degradation in the proteasome, for example, results from poly-ubiquitination at Lys-11 and/or Lys-48. For the discovery of ubiquitin A. CIECHANOVER, A. HERSHKO & I. ROSE received the Nobel Prize for Chemistry in 2004.

There is a whole family of **ubiquitin-like modifiers** which are transferred in a similar manner, but whose function we are only beginning to understand (see, e.g., [36] for a recent review). Transfer is often by an E2-ligase directly; E3-ligases are required for ubiquitin presumably because of the large number of different proteins labelled with this marker. These **ubiquitin-like modifiers** (UbLs) are involved in the regulation of endocytosis, apoptosis, cell cycle, DNA repair and other processes. There are even ubiquitin-like proteins (Isg15 and Fat10), which are regulated by interferon and modulate immune response. The mechanisms involved in these regulatory pathways are, however, poorly understood.

Ubiquitin and ubiquitin-like modifiers are removed by deconjugases that cleave the isopeptide bond. These are often highly specific for both the tag and the modified target.

2.7 The Relationship Between Protein Structure and Function: Green Fluorescent Protein

GFP is produced by the cœlenterate *Aequorea victoria* (Hydrozoa, MURBACH & SHEARER, 1902). It accepts energy from a chemiluminescent protein (which would otherwise produce blue light) and translates it into green light by **bioluminescence resonance energy transfer (BRET)**. The function of bioluminescence in these animals is unknown. Because of its intensive green fluorescence GFP has become a favourite **marker** in molecular biology. The genetic information for GFP is attached to the gene for the protein under investigation, so that a fusion product is generated. Both the amount of target protein produced and its subcellular localization can then be studied by fluorescence video microscopy.

The fluorophore of GFP is produced by oxydation from 3 neighbouring amino acids (see Fig. 2.37); this process can proceed spontaneously in the absence of other proteins and cofactors except oxygen. Thus GFP can be used as marker in any aerobic cell. The reaction increases the system of conjugated double bonds (π-system) compared to Tyr and shifts the absorbtion maximum from 280 to 395 nm.

GFP consists mainly of antiparallel β-strands, which together form a β-barrel. An α-helix with the fluorophore runs in the centre of the barrel (see Fig. 2.38), where the fluorophore is protected from collisions with water and, in particular, oxygen. Any such collision would prevent fluorescence by taking away energy from the excited fluorophore. The structure thus explains the high quantum efficiency of GFP (≈ 0.8 green photons produced per blue photon absorbed).

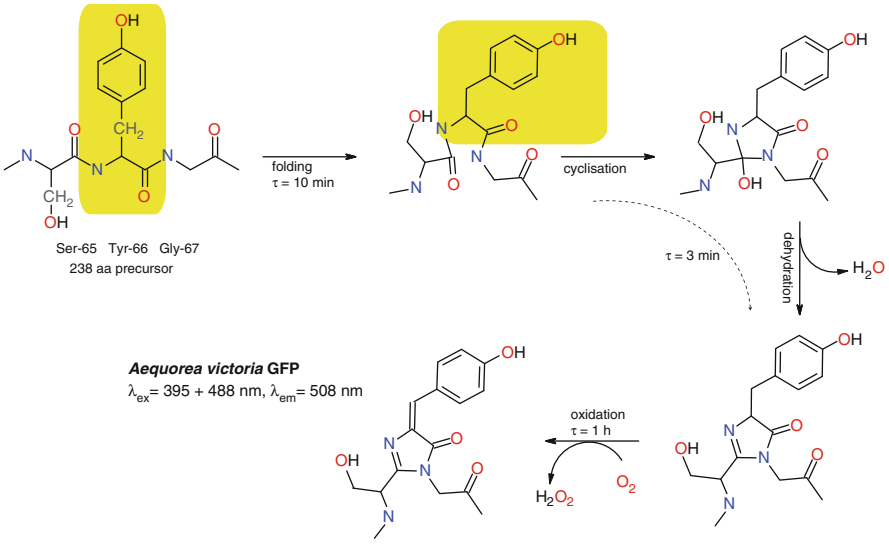


Fig. 2.37 Maturation of the fluorophore in GFP. The reaction does not require enzymes or cofactors except molecular oxygen

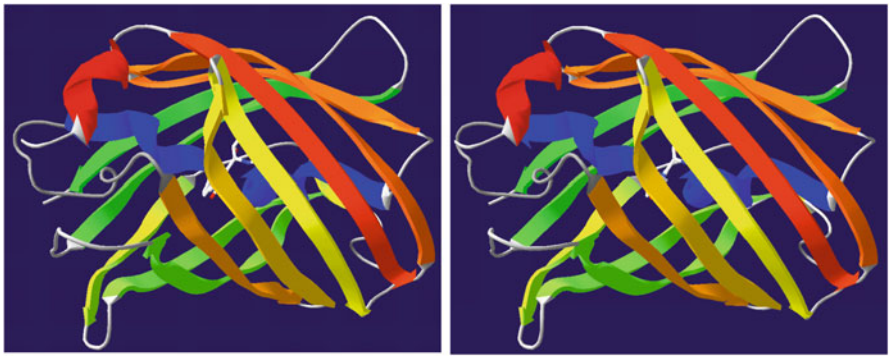


Fig. 2.38 Stereo representation of the crystal structure of GFP (PDB-code 1ema)

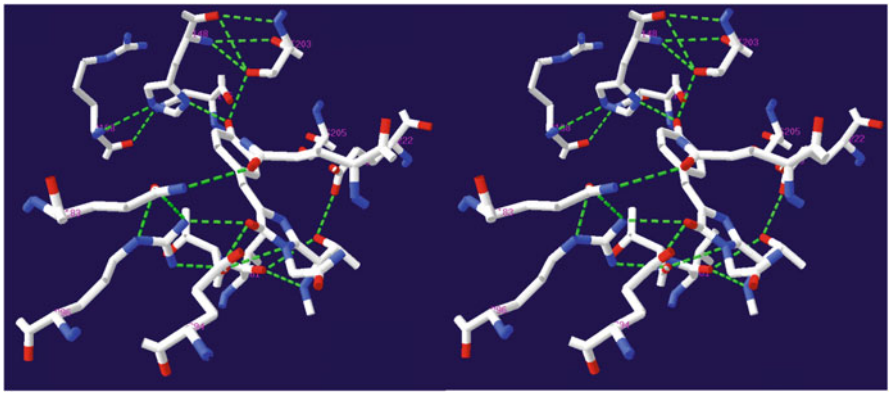


Fig. 2.39 Amino acids that interact with the fluorophore in the active centre of GFP

GFP has 2 absorption maxima, that of the nonionised fluorophore (phenol) at 395 nm (UV), and that of the ionised (phenolate) at 488 nm (blue). Ser-65 donates a hydrogen bond to Glu-222 which makes deprotonation of Glu-222 easier (see Fig. 2.39). The negative charge on this residue then prevents deprotonation of the phenyl-group of the fluorophore. If Ser-65 is mutated to Ala, ionization of the fluorophore becomes easier, the absorbtion maximum at 395 nm is reduced, and that at 488 nm becomes stronger. The phenolate ion forms a hydrogen bond with Thr-203; if this is mutated to Ile the fluorophore is stabilised in the phenol-form and the absorbtion maximum at 395 nm becomes stronger at the expense of that at 488 nm. If Thr-203 is mutated to an aromatic amino acid such as Tyr, stacking of the π-systems leads to a red-shift of both absorbtion and emission maxima by 20 nm because of the reduced exited state energy (yellow fluorescent protein). If Tyr-66 is replaced by Trp or His, the maxima are blue-shifted to 436/476 nm (cyan fluorescent protein) and 390/450 nm (blue fluorescent protein), respectively.

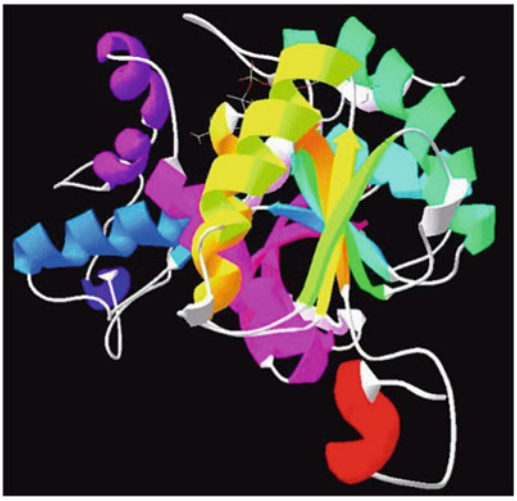
2.8 Exercises

2.8.1 Problems

2.1. Which of the following amino acids cannot be posttranslationally modified?

A Cysteine
B Serine
C Tyrosine
D Alanine
E Asparagine

2.2. Why can the isoelectric point of a protein not be predicted from its amino acid composition and the known isoelectric points of the amino acids?



2.3. The above picture shows a cartoon of the structure of lactate dehydrogenase. To which SCOP class of proteins does this enzyme belong?

A all α
B all β
C α /β
D α +β
E coiled coil

2.4. Which of the following amino acids would you expect to find in the core of a protein?

A Lys
B Arg
C Glu
D Leu
E Asp

2.8.2 Solutions

2.1 Cys is involved in the formation of disulphide bonds; Ser can be phosphorylated and can carry O-linked sugar trees. Tyr can also be phosphorylated and Asn carries N-linked sugar trees. However, Ala has a rather unreactive methyl group as side-chain, which does not participate in posttranslational modifications.

2.2 Because the isoelectric points of amino acids were measured in aqueous solutions. Inside a protein interactions between the ionisable groups of the amino acids can modulate their pK_a-values and hence the tabulated *p*I-values do not apply.

2.3 Note the parallel β-sheet resulting from alternating α-helices and β-strands. Thus the structure is α /β.

2.4 Soluble proteins are held together by hydrophobic interactions between amino acids in the core, whereas hydrophilic amino acids on the surface allow interactions with the environment. Leu with its hydrocarbon side-chain is apolar.

References

1. B. Albers, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, J. Wilson, T. Hunt, *Molecular Biology of the Cell*, 5th edn. (Garland Science, New York, 2008). ISBN 0-8153-4105-9
2. C. Anfinsen, Principles that govern the folding of protein chains. Science **181**, 223–230 (1973). doi: 10.1126/science.181.4096.223
3. R. Bhattacharyya, P. Chakrabarti, Stereospecific interaction of proline residues in protein structures and complexes. J. Mol. Biol. **331**, 925–940 (2003). doi: 10.1016/S0022-2836(03)00759-9
4. N. Boggetto, M. Reboud-Ravaux, Dimerization inhibitors of HIV-1 protease. Biol. Chem. **383**, 1321–1324 (2002). doi: 10.1515/BC.2002.150
5. C. Choudhary, C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, M. Mann, Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science **325**(5942), 834–840 (2009). doi: 10.1126/science.1175371
6. S. Clarke. Protein methylation. Curr. Opin. Cell Biol., **5**, 977–983 (1993). doi: 10.1016/0955-0674(93)90080-A
7. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. Nature **409**, 860–921 (2001). doi: 10.1038/35057062
8. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. Nature **431**, 931–945 (2004). doi: 10.1038/nature03001
9. C. Darwin, *On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life* (John Murray, London, facsimile edition, 1859). ISBN 978-0-6746-3752-8
10. M.O. Dayhoff, Computer analysis of protein evolution. Sci. Am. **221**(1), 86–95 (1969). doi: 10.1038/scientificamerican0769-86
11. J.M. Denu, Linking chromatin function with metabolic networks. Sir-2 family of NAD^+-dependent deacetylases. TIBS **28**, 41–48 (2003). doi: 10.1016/S0968-0004(02)00005-1
12. J.M. Forbes, M.E. Cooper, Mechanisms of diabetic complications. Physiol. Rev. **93**(1), 137–188 (2013). doi: 10.1152/physrev.00045.2011
13. F.S. Gimble, Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol. Let. **185**(2), 99–107 (2000). doi: 10.1111/j.1574-6968.2000.tb09046.x

14. J.P. Gogarten, A.G. Senejani, O. Zhaxybayeva, L. Olendzenski, E. Hilario, Inteins: structure, function, and evolution. Annu. Rev. Microbiol. **56**(1), 263–287 (2002). doi: 10.1146/annurev.micro.56.012302.160741
15. E. Granseth, G. van Heijne, A. Elofsson, A study of the membrane-water interface region of membrane proteins. J. Mol. Biol. **346**, 377–385 (2005). doi: 10.1016/j.jmb.2004.11.036
16. V.H. Haase, Regulation of erythropoiesis by hypoxia-inducible factors. Blood Rev. **27**(1), 41–43 (2013). doi: 10.1016/j.blre.2012.12.003
17. C. Hadley, D.T. Jones, A systematic comparison of protein structure classifications: SCOP, CATH and FSSP. Structure **7**(9), 1099–1112 (1999). doi: 10.1016/S0969-2126(99)80177-4
18. B.K. Ho, A. Thomas, R. Brasseur, Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics and H-bonding in the α-helix. Protein Sci. **12**(11), 2508–2522 (2003). doi: 10.1110/ps.03235203
19. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577–2637 (1983). doi: 10.1002/bip.360221211. URL http://zhanglab.ccmb.med.umich.edu/literature/dssp.pdf
20. C. Levinthal, How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, ed. by J.T.P. DeBrunner, E. Munck, pp. 22–24 (University of Illinois Press, Urbana, IL, 1969). URL http://www-wales.ch.cam.ac.uk/~mark/levinthal/levinthal.html
21. H. Lodish et al., *Molecular Cell Biology*, 7th edn. (W.H. Freeman and Company, New York, 2012). ISBN 978-1-4292-3413-9
22. C. Luevano-Contreras, K. Chapman-Novakofski, Dietary advanced glycation end products and aging. Nutrients **2**(12), 1247–1265 (2010). doi: 10.3390/nu2121247
23. Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, A. Kidera, Minimum free energy path of ligand-induced transition in adenylate kinase. PLoS Comput. Biol. **8**(6), e1002555 (2012). doi: 10.1371/journal.pcbi.1002555
24. R.V. Pappu, R. Srinivasan, G.D. Rose, The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding. Proc. Natl. Acad. Sci. USA **97**, 12565–12570 (2000). doi: 10.1073/pnas.97.23.12565
25. H. Paulus, Protein splicing and related forms of protein autoprocessing. Annu. Rev. Biochem. **69**, 447–496 (2000). doi: 10.1146/annurev.biochem.69.1.447
26. F. Piarulli, G. Sartore, A. Lapolla, Glyco-oxidation and cardiovascular complications in type 2 diabetes: a clinical update. Acta Diabetol. **50**(2), 101–110 (2013). doi: 10.1007/s00592-012-0412-3
27. G.N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and proteins. Adv. Protein Chem. **23**, 283–437 (1968). doi: 10.1016/S0065-3233(08)60402-7
28. J.S. Richardson, The anatomy and taxonomy of protein structures. Adv. Protein Chem. **34**, 167–339 (1981). doi: 10.1016/S0065-3233(08)60520-3
29. D.R. Sell, V.M. Monnier, Molecular basis of arterial stiffening: Role of glycation. Gerontology **58**(3), 227–237 (2012). doi: 10.1159/000334668
30. J.G. Snedeker, A. Gautieri, The role of collagen crosslinks in ageing and diabetes - the good, the bad, and the ugly. Muscles Ligaments Tendons J. **4**(3), 303–308 (2014). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4241420/pdf/303-308.pdf
31. G. Sorci, F. Riuzzi, I. Giambanco, R. Donato, RAGE in tissue homeostasis, repair and regeneration. Biochim. Biophys. Acta **1833**(1), 101–109 (2013). doi: 10.1016/j.bbamcr.2012.10.021
32. F. Takagi, N. Koga, S. Takada, How protein thermodynamics and folding mechanism are altered by the chaperonin cage: Molecular simulations. Proc. Natl. Acad. Sci. USA **100**, 11367–11372 (2003). doi: 10.1073/pnas.1831920100
33. H.A. Tissenbaum, L. Guarente, Increased dosage of sir-2 gene extends lifespan in *Caenorabditis elegans*. Nature **410**, 227–230 (2001). doi: 10.1038/35065638
34. J.S. Valastyan, S. Lindquist, Mechanisms of protein-folding diseases at a glance. Dis. Models Mech. **7**(1), 9–14 (2014). doi: 10.1242/dmm.013474
35. J.C. Venter et al., The sequence of the human genome. Science **291**, 1304–1351 (2001). doi: 10.1126/science.1058040

36. R.D. Vierstra, The expanding universe of ubiquitin and ubiquitin-like modifiers. Plant Physiol. **160**(1), 2–14 (2012). doi: 10.1104/pp.112.200667
37. S. Walter, J. Buchner, Molecular chaperones – cellular machines for protein folding. Angew. Chemie Int. Ed. **41**, 1098–1113 (2002). doi: 10.1002/1521-3773(20020402)41:7%3C1098::AID-ANIE1098%3E3.0.CO;2-9
38. Q. Wang, Y. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. Yao, Z.-B. Ning, R. Zeng, Y. Xiong, K.-L. Guan, S. Zhao, G.-P. Zhao, Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. Science **237**, 1004–1007 (2010). doi: 10.1126/science.1179687
39. C.M. Wilmot, J.M. Thornton, Analysis and prediction of the different types of β-turns in proteins. J. Mol. Biol. **203**, 221–232 (1988). doi: 10.1016/0022-2836(88)90103-9
40. M.D. Yoder, F. Jurnak, The parallel β helix and other coiled folds. FASEB J. **9**(5), 335–342 (1995). URL http://www.fasebj.org/content/9/5/335.abstract
41. A.Q. Zhou, C.S. O'Hern, L. Regan, Revisiting the Ramachandran plot from a new angle. Protein Sci. **20**(7), 1166–1171 (2011). doi: 10.1002/pro.644
42. L. Zhou, Y. Zeng, H. Li, Y. Li, J. Shi, W. An, S.M. Hancock, F. He, L. Qin, J. Chin, P. Yang, X. Chen, Q. Lei, Y. Xiong, K.-L. Guan, Regulation of cellular metabolism by protein lysine acetylation. Science **237**, 1000–1004 (2010). doi: 10.1126/science.1179689
43. R. Zwanzig, A. Szabo, B. Bagchi, Levinthal's paradox. Proc. Natl. Acad. Sci. USA **89**, 20–22 (1992). doi: 10.1073/pnas.89.1.20

Chapter 3
Proteins in the Lab

Abstract Before proteins can be studied, they need to be isolated from all other proteins present in a cell or organism. Differences in size, shape, charge, stability, or binding properties between proteins are utilised, using chromatographical, electrophoretic, or precipitation methods. Before membrane proteins can be purified, they need to be solubilised out of the membrane by detergents. Once purified, the primary, secondary, tertiary, and quaternary structures of a protein need to be determined. Finally, the protein's function is characterised.

3.1 Protein Purification

Each cell contains several thousand different proteins. If we want to find out about their properties, we first have to purify them. How else could we be sure that a particular reaction is really caused by a particular protein?

The basic principles behind protein purification (and only those can be covered here) are easily understood, however, it is difficult if not impossible to predict which methods are most suitable for the purification of a given protein. Very rarely is it possible to purify a protein by a single method, usually the judicious combination of several purification steps is required. Protein purification is therefore much more art than science. More detailed explanations of this material may be found in [5, 14].

3.1.1 Homogenisation and Fractionisation of Cells and Tissues

If you want to study, say, the enzymes present in the liver, the first thing you have to do is to obtain fresh liver tissue. This is done at the abattoir, immediately after an animal has been slaughtered. The tissue is transported into the laboratory on ice, to minimise proteolytic damage.

Once in the lab, the tissue needs to be disrupted. This is a critical step: cells should be broken open, but cell organelles should remain intact. Usually the tissue is minced first by hand, then cut into a fine pulp by rotating knives (e.g., in a blender such as those used to make milkshakes in the kitchen) and finally homogenised by the application of shearing forces in specialised equipment (POTTER-ELVEHJEM- or

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_3

DOUNCE-homogeniser, French press). All these steps are performed on ice. Buffer solutions are used to keep the *p*H at the required value; they usually also contain protease inhibitors, antioxydants, and sucrose or mannitol to keep the osmotic pressure in the solutions at the same level as in the cell ($\approx$ 300–350 mosm). Ions such as Na^+, K^+, Ca^{2+} or Mg^{2+} are added as required by the enzyme to be isolated, some very sensitive enzymes also require the addition of their substrate to stabilise them.

Then the various cell organelles need to be separated from each other. This is done by fractionated centrifugation [11]. First connective tissue, undamaged cells, and other debris are removed by a brief spin at low speed (10 min at 500 g[1]). In the next step nucleï and plasma membranes are spun down (10 min at 3000 g). Mitochondria, plastids, and heavy microsomes require about 30 min at 20 000 g to be spun down; 1 h at 100 000 g is required for light microsomes. The remaining supernatant contains the cytosol.

These crude preparations are then subjected to further purification steps.

3.1.2 Precipitation Methods

Protein precipitation is a very crude method for purification and rarely achieves enrichment by more than a factor of 2–3. However, it is quick and cheap. If applied to crude homogenisates, it may remove material that would interfere with later purification steps.

3.1.2.1 Salts

Proteins interact strongly with water, and only the hydrated form is soluble. If the available water concentration is reduced by the addition of salts or water-miscible organic solvents (methanol, ethanol or aceton), proteins precipitate out of solution. It is essential to perform these reactions in the cold; proteins are rapidly denatured by precipitating agents at room temperature.

The precipitating agent is slowly added to the well-stirred protein solution until it reaches a concentration where the desired protein is just soluble. Precipitated material is removed by centrifugation, then more precipitant is added, until all desired material has precipitated. It is separated from the solution by centrifugation.

Proteins can be separated into two groups depending on their behaviour towards salt:

[1]The centrifugal acceleration is usually measured as a multiple of the gravitational acceleration on earth ($g = 9.81\ m/s^2$).

Globulins are more or less globular proteins; their electrical charge is evenly distributed across their surface area. Globulins are soluble in distilled water, and can be precipitated by high salt concentrations.

Albumins have molecules with an elongated, rod-like shape with unsymmetrical distribution of electrically charged groups. As a result, albumins are insoluble in distilled water, because the molecules form head-to-tail aggregates held together by electrical forces. Low salt concentrations neutralise these charged groups and allow the albumins to go into solution; high salt concentrations precipitate them again.

From the salts ammonium sulphate is most often used, as it is cheap, nontoxic, highly water soluble, and strongly ionised. Purified proteins can be crystallised by slow addition of ammonium sulphate; such crystals, suspended in the mother liquor, tend to be very stable if kept refrigerated. Many enzymes are sold in this form by the suppliers. Salts are removed from proteins by dialysis or gel filtration.

3.1.2.2 Solvents

Organic solvents (ethanol, methanol, or acetone) need to be used even more carefully than salts to prevent irreversible denaturation of proteins; precipitation is usually done at subzero temperatures. Particular attention needs to be given to the fact that mixing of solvents with water generates heat. After precipitation the solvent is usually removed by lyophilisation (freeze-drying). The resulting powders are quite stable. Lab suppliers offer "acetone powders" produced from various organs of various species as a crude source of enzymes.

Polyethylene glycol (PEG) may also be used, it is less denaturing and produces no heat when mixed with water.

3.1.2.3 Heat

Proteins are irreversibly denatured by heat and precipitate out of solution. However, some proteins are more resistant than others, especially in the presence of their ligands. If crude protein extracts are heated to a carefully chosen temperature for a carefully chosen time, some of the extraneous proteins precipitate, whilst the relevant one stays intact. Precipitated material is removed by centrifugation.

3.1.3 Chromatography

Chromatography was invented by the Russian botanist MIKHAIL SEMJONOWITSCH TSWETT for the separation of leaf pigments [34, 35]. The method was extended to

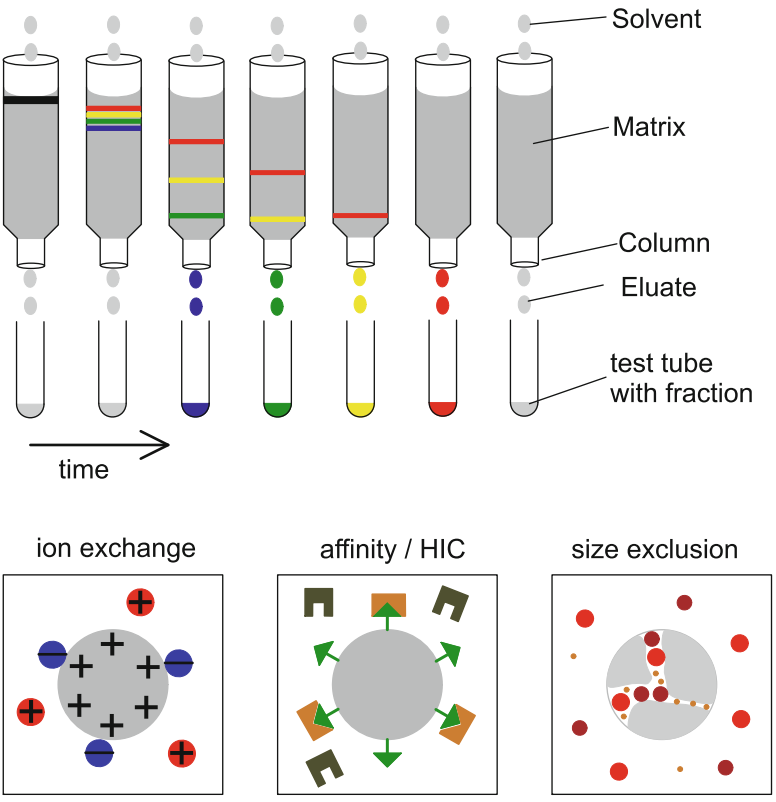


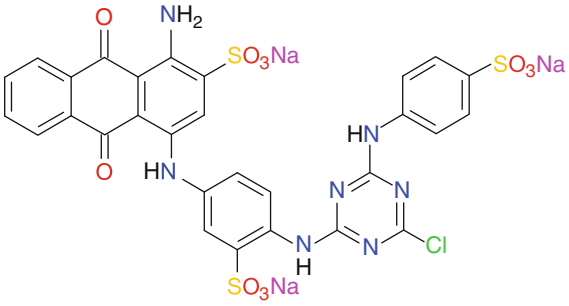
Fig. 3.1 Principles of chromatography. The sample is moved by a solvent (mobile phase) past a matrix (stationary phase). Different sample molecules have different partition coefficients between mobile and stationary phase and are delayed differently. Separation can be by ionic interactions, specific interactions with a ligand, and size exclusion. For details see text

proteins by RICHARD WILLSTÄTTER. Proteins are bound to a solid support and then specifically eluted. Several types of interaction can be used (see also Fig. 3.1):

- The electrical charge of proteins depends on their amino acid composition and the *p*H of the medium. If the *p*H is lower than the isoelectric point of a protein, it will have a positive net charge; above the *p*I the net charge will be negative. If a protein is passed over a support with charged groups (e.g., sulphopropyl $-CH_2-CH_2-CH_2-SO_3^-$ or diethylaminoethan $-CH_2-CH_2-NH^+(CH_2-CH_3)_2$) it may or may not bind to it, depending on *p*H. Binding strength depends on the number of charges on the protein; weakly bound proteins can be eluted with salt solutions of low concentration, whereas for strongly bound proteins high salt concentrations are required (10 mM to 1 M). This method is called **ion exchange chromatography**; it is probably the most

often used method for protein purification. It is also possible to use a *p*H gradient for elution (**chromatofoccusing**); this method has higher resolving power but is much more expensive and rarely used.

- **Gel filtration**, also called gel chromatography or size exclusion chromatography (SEC), is based on the different size and shape (STOKES-radius) of protein molecules. The matrix contains pores of different sizes. Small proteins can diffuse into all of them and are delayed, whilst the largest proteins do not fit into any pores and pass the gel without delay. Because no binding of proteins to the gel occurs, this method is very gentle, however, it is applicable only to small volumes of concentrated samples.
- Proteins contain a variable amount of hydrophobic amino acids which leads to different binding strength to hydrophobic groups (e.g., butyl- or phenyl-groups) attached to a support matrix. For **hydrophobic interaction** chromatography proteins are passed over such a column dissolved in a salt solution (0.5 to several M), to maximise interactions with the support. They are eluted with a gradient of decreasing salt concentration, until even very hydrophobic proteins leave the column. In organic chemistry columns with octadecyl-groups (C_{18}, "**reverse phase chromatography (RPC)**") are used, but it would not be possible to elute proteins from such columns without denaturing them. C_{18}-columns are useful for small peptides, however, which can be eluted with water/acetonitril mixtures.
- Proteins show specific interactions with some other molecules, enzymes with their substrates, receptors with their ligands, antibodies to antigens, or glycoproteins with lectins. If such molecules are chemically bound to a support, those proteins that interact with them will be retained on the column, whereas all other proteins pass the column unhindered. The bound proteins can then be eluted with a ligand solution or by changing *p*H or ionic strength to reduce protein-ligand interactions. Because of the specificity of ligand-protein interactions, **affinity chromatography** can sometimes lead to one-step purification protocols. Biological ligands are often expensive and unstable, however, organic chemistry has produced a lot of compounds, which may to a protein look like its ligand, even though the structure may be quite different. Such compounds can be used with advantage for affinity chromatography, the use of Cibacron blue (see Fig. 3.2) for the purification of NAD^+ or $NADP^+$ dependent enzymes is probably the most well known example.
- Immobilised ion affinity chromatography (IMAC) uses columns where metal ions (mostly Ni, but also Co and some others) are held by chelating groups such as nitrilotriacetate. Surface-exposed His-groups on proteins bind to these immobilised metal ions and can be eluted either by competition with increasing concentrations of imidazole or by stripping bound protein and metal ions from the column with strong complex-forming agents such as EDTA. This method is used chiefly with proteins genetically engineered to have a poly-His tail at either the N- or C-terminal end ("**His-tag**"), but can also be useful for natural proteins.
- The theoretical basis for **hydroxyapatite chromatography** is much less understood than that for the above methods. Hydroxyapatite is a crystalline form of calcium phosphate also found in teeth and bones. Depending on chromatographic



Cibacron Blue F3G-A
(Procion blue)

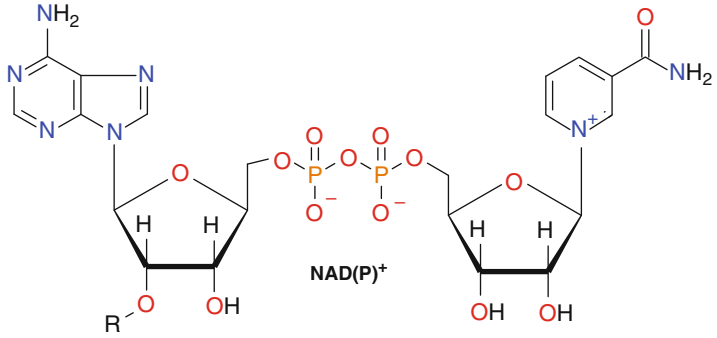


Fig. 3.2 Cibacron Blue is used for the affinity purification of $NAD(P)^+$ dependent enzymes. Because unlike $NAD(P)^+$ the dye is very stable, the column can be used for the isolation of such enzymes repeatedly

conditions the surface of these crystals bears an excess either of positively charged calcium or negatively charged phosphate ions. HA-chromatography is probably a combination of ion exchange and affinity chromatography. Bound proteins are eluted with a potassium (or sodium-) phosphate solution of increasing concentration (10–500 mM). Although results with HA chromatography are even more difficult to predict than with other methods, it is an essential tool for the isolation of some proteins.

Chromatographic methods are the workhorses in protein purification. Columns can be constructed for sample volumes from a few µl for analytical applications to several L for industrial-scale preparative purification. The equipment for protein chromatography needs to be constructed from biocompatible materials (glass, certain plastics); steel may release heavy metal ions and is unsuitable (whilst in common use in organic chemistry). Sample vials, columns, and fraction collectors should be kept at 4 °C during chromatographic runs. Fouling of the columns is prevented by filtering all samples and buffers through low-protein-binding membrane filters of 0.45 or 0.22 µm pore size.

3.1.4 Electrophoresis

Because proteins are charged molecules (if *p*H $\neq$ *p*I), they move in an electrical field. Direction and magnitude of the electric force acting on the protein molecules depends on their amino acid composition and the *p*H of the buffer, and the friction experienced depends on the pore size of the medium and the size and shape of the protein. Thus chromatographic separations can be designed to separate by **charge**, **isoelectric point** or by STOKES **radius**.

Electrophoretic methods are normally used for analytical separations, as it is difficult to remove the heat produced by the electrical current from larger gels. However, equipment for small-scale preparative work (up to a few mg) is also available.

Originally electrophoresis was carried out on paper or nitrocellulose strips soaked in buffer. Nowadays this method is limited to special applications; instead gels of polyacrylamide are commonly used. The pore size of these gels can be adjusted by the polyacrylamide concentration. If very large pores are required, agarose or starch gels are used instead.

3.1.4.1 Native Electrophoresis

Native electrophoresis [26] separates by charge (controlled by the *p*H of the buffer), size, and shape of the protein molecules. Because of this, native gels are sometimes difficult to interpret.

A special case is the separation on paper- or nitrocellulose stripes: there are no pores to interact with as the proteins move on the surface. Thus separation is by charge only. Serum proteins are separated by this method in clinical laboratories.

3.1.4.2 Denaturing Electrophoresis

Most common is the method described by LAEMMLI [19]. The protein is heated in a sample buffer which contains a mercapto-compound (β-mercaptoethanol or Dithiotreitol (DTT)) to open −S−S− bridges in the protein, destroying its tertiary structure. The buffer also contains the anionic detergent Sodium dodecylsulphate (SDS), which destroys the secondary structure of proteins and binds to them at a fairly constant ratio of about 1 molecule of SDS per 3 amino acids. For this reason the charge-to-weight ratio of all proteins is almost constant, and proteins experience constant acceleration in an electrical field, independent of their composition. However, the pores of the gel slow down big proteins more than small ones. Separation is therefore by size only. If the migration distance of a protein is compared to that of proteins of known size, its molecular mass can be estimated (see Fig. 3.3).

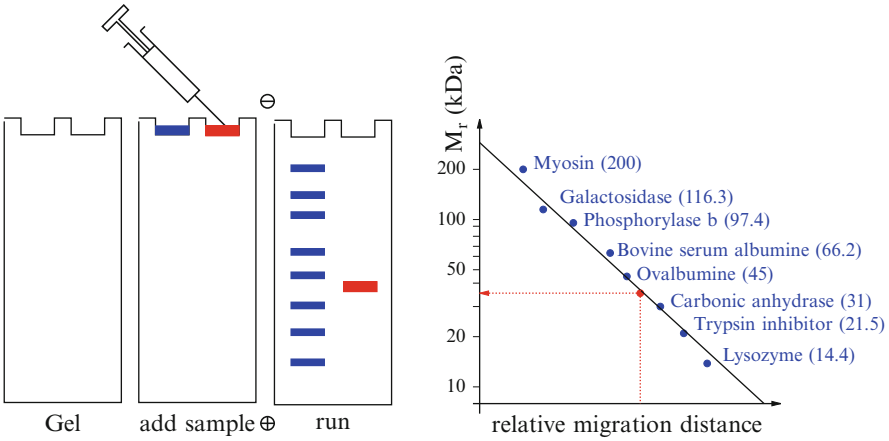


Fig. 3.3 Gel-Electrophoresis. Proteins of known molecular mass are used to establish a standard curve that allows the mass of sample proteins to be determined

A LAEMMLI-gel consists of two separate zones, which differ in their *p*H. In the *p*H 6.8 stacking gel the proteins are focussed into narrow bands by **isotachophoresis**, the *p*H 8 separating gel then separates the focussed bands (discontinuous electrophoresis, [26]). This method is called SDS-polyacrylamide gel electrophoresis (SDS-PAGE). It is quick, relatively cheap, and reproducible, and is the most often used type of electrophoresis in a biochemical research laboratory. In some cases the positively charged detergent CTAB may be used with advantage instead of the negatively charged SDS [4].

Native blue electrophoresis [31] is intermediate between denaturing and native electrophoresis. Instead of detergent, the anionic dye Coomassie Brilliant blue R250 (CBB-R250) is used for solubilisation of proteins, in the absence of reducing agents. This method is used to isolate protein complexes, for example, from mitochondria. The complexes may then be resolved into their constituent proteins by a second, denaturing, electrophoresis.

3.1.4.3 Isoelectric Focusing and 2D-Electrophoresis

If the gel contains several thousand different buffer molecules, each with slightly different pK_a-value, these buffers can be sorted by an electrical field according to pK_a. This results in a continuous *p*H-gradient across the gel. If proteins are added to the gel, they move in the *p*H-gradient until they reach their *p*I, where they are uncharged and hence no longer move. As a result proteins are sorted by *p*I (see fig. 3.4). This method can separate proteins which differ only by a single charge, resulting in a *p*I-difference of 0.001. IEF can be performed with native proteins or with proteins whose tertiary and secondary structure has been destroyed with (nonionic) caotropes like urea.

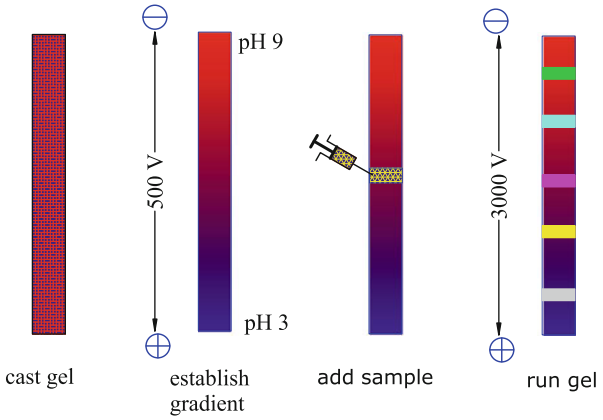


Fig. 3.4 Isoelectric focussing. A *p*H-gradient is established across the gel by electrophoresis at relatively low voltages ($\approx$ 500 V). The sample is applied to this gradient and run at high voltages until all proteins have achieved their equilibrium position

If after separation the IEF-gel is mounted across an SDS-PAGE-gel, the proteins which have already been separated by *p*I can be separated again by molecular mass (see fig. 3.5). This "two-dimensional" electrophoresis can separate about 10 000 different proteins from a cell lysate [18].

Comparing the protein expression profiles from cells of different developmental stages, or of healthy and diseased cells, can turn up proteins involved in developmental regulation or in disease processes. With luck such a protein may be a useful drug target. Thus the field of **proteomics** (the proteome of a cell is the collection of all proteins it contains, just as the genome is the collection of all its genes [17]) has attracted considerable attention in the pharmaceutical industry.

3.1.5 Membrane Proteins

Before membrane proteins can be separated from each other, they need to be taken out of the membrane. This process is called **solubilisation**. Proteins which are only attached to a membrane (peripheral membrane proteins) may be solubilised by elevated *p*H or salt concentration [9, 12, 16], or by caotropic agents such as urea. However, proteins with transmembrane segments (integral membrane proteins) can be solubilised only by detergents. Detergents form soluble complexes with membrane proteins; any attempt to remove the detergent will usually lead to aggregation and precipitation of the proteins, which are not by themselves water soluble.

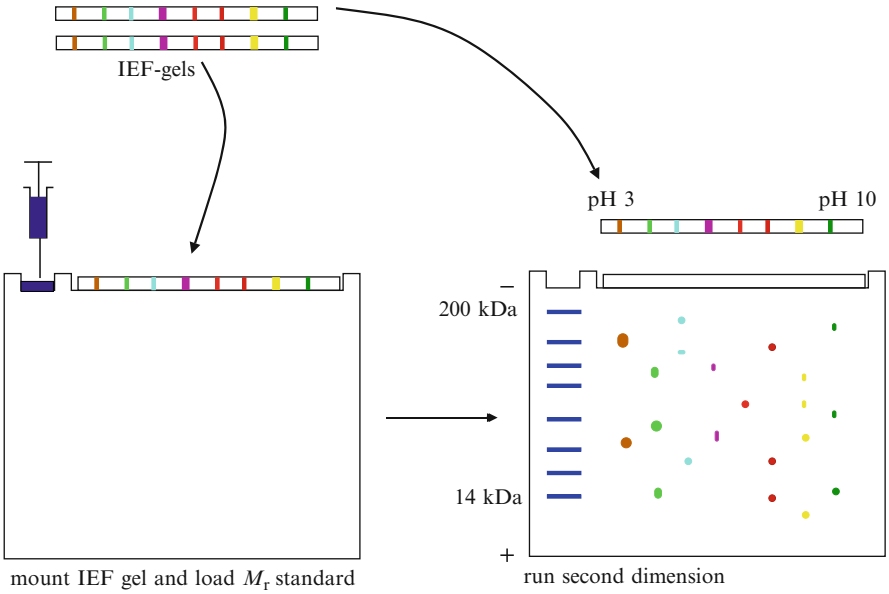


Fig. 3.5 Two-dimensional electrophoresis. A protein mixture is first separated on an IEF-gel, which is then mounted across a LAEMMLI-gel. The second electrophoresis separates the bands of proteins with identical *p*I by molecular mass. About 10 000 different proteins can be distinguished on a 2d-gel of mammalian cells

We distinguish nonionic, cationic, and anionic detergents (see Fig. 3.6). Of these the nonionic detergents tend to be the mildest, which is good for the preservation of protein function, but on the other hand their ability to dissociate protein complexes, a requirement for the purification of proteins, is limited.

Purification of solubilised membrane proteins is in principle done in a similar way as with soluble proteins. Care must be taken however to have detergent present in all solutions, lest proteins aggregate (Fig. 3.7).

The detergent/protein and detergent/lipid ratios are critical during solubilisation. Membrane proteins are surrounded by a ring of tightly bound lipids (**annular** as opposed to **bulk** lipids; see Fig. 3.8). If annular lipids are replaced by detergent, irreversible loss of protein function occurs, as annular lipids adapt their conformation to the irregular shape of the protein surface; many detergent molecules can't do this. If the detergent/protein ratio is too small, several proteins will occupy a single detergent micelle and purification will not be possible. If the detergent/lipid ratio is too high, annular lipids will be stripped from the proteins and inactivation will occur. This target conflict can be solved by adding extraneous lipid during solubilisation.

Once a pure membrane protein has been obtained, the protein needs to be reinserted into a membrane, so that its function may be studied. This is done by slowly removing the detergent in the presence of lipids [29]. Under suitable conditions, the lipids form closed membrane vesicles (**liposomes**), and the proteins

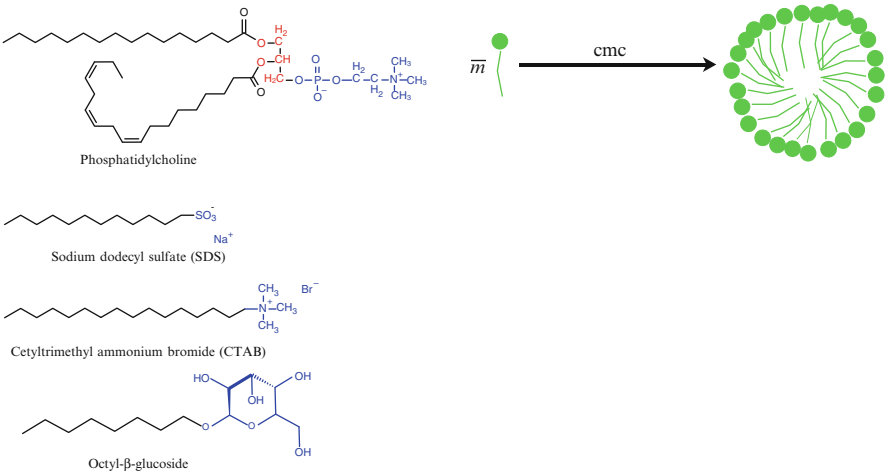


Fig. 3.6 Lipids and detergents. *Left*: A typical lipid consists of the alcohol glycerol (*red*) esterified with two fatty acids (*black*). The remaining hydroxy-group is esterified with phosphoric acid, which carries a hydrophilic side-chain, here phosphocholine (*blue*). All detergents have a lipophilic tail and a hydrophilic headgroup. In the case of SDS, the sulphonic acid group bears a negative charge at physiological *p*H (anionic), whereas CTAB is positively charged (cationic). OG is an example for a nonionic detergent; the hydrophilic headgroup is a glucose molecule. *Right*: Detergents form aggregates, so-called micelles, above a certain concentration. In a micelle the lipophilic tails point to the centre, where they are shielded from water; the hydrophilic headgroups are on the surface and interact with water. The concentration at which micelles form (**critical micellar concentration (cmc)**) and the aggregation number (size of the micelles $\bar{m}$) depend on the headgroup and the tail length of the detergent

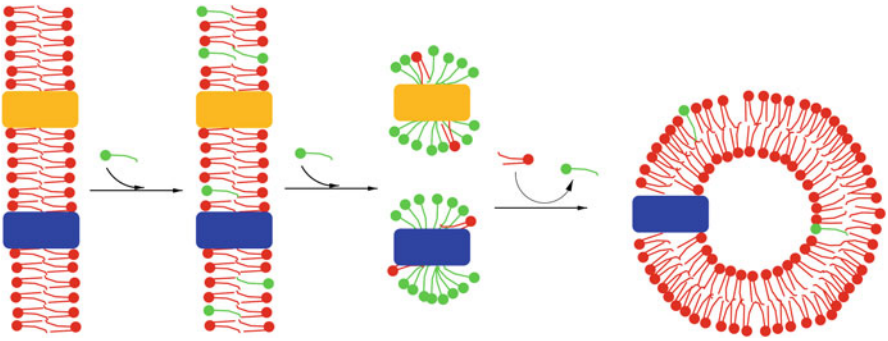


Fig. 3.7 Solubilisation of membrane proteins by detergents (*green*). If detergent is slowly added to a membrane suspension, detergent molecules dissolve in the membrane plane until saturation is reached. Then mixed detergent-lipid-protein micelles form. Proteins may stay active in such micelles unless the detergent replaces closely bound lipid molecules, which membrane proteins need to maintain their structure. Such mixed micelles contain only a single protein, and protein purification is possible. Once this has been achieved, proteins are reconstituted into liposomes by addition of lipid and removal of detergent

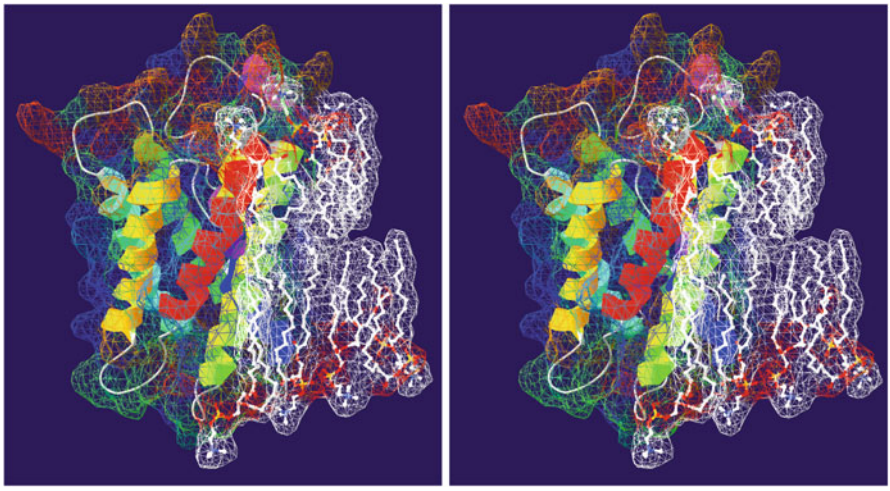


Fig. 3.8 Annular lipids on aquaporin-0 (dimyristoyl-glycerophosphocholine, DMPC). These lipids are bound by the protein, are held in an ordered conformation, and can thus be characterised by electron crystallography (PDB-code 2b6o). EM crystallography does not require the use of detergents and is therefore better suited to investigate bound lipids than X-ray crystallography

insert into them. This process is called reconstitution and is probably the blackest of all arts in protein science.

3.2 Determination of Protein Concentration

One should think that the determination of protein concentration is a very basic task, and that reliable methods were available to do this reproducibly. Alas, this is not the case. In most cases the best result one can obtain is "the same colour as x mg/mL of standard protein". Because, however, proteins—depending on their amino acid composition—give different colour yield with most methods, the same protein sample may with different methods and standard proteins give results that differ by an order of magnitude. In addition, chemicals commonly used in the protein lab interfere with some assays.

The following methods are in common use:

Absorbance Proteins do not absorb visible light (380–760 nm) and are uncoloured unless they contain a coloured prosthetic group (flavine, hæme, Cu,...). However, almost all proteins contain Phe, Tyr, and/or Trp. The aromatic rings of these amino acids absorb UV-light, with a maximum at 280 nm [37].

The biuret assay In alkaline solution, copper salts (Cu^{2+}) are reduced by the protein to Cu^{+}, which forms a violet complex with substances containing two or more peptide bonds (see Fig. 3.9).

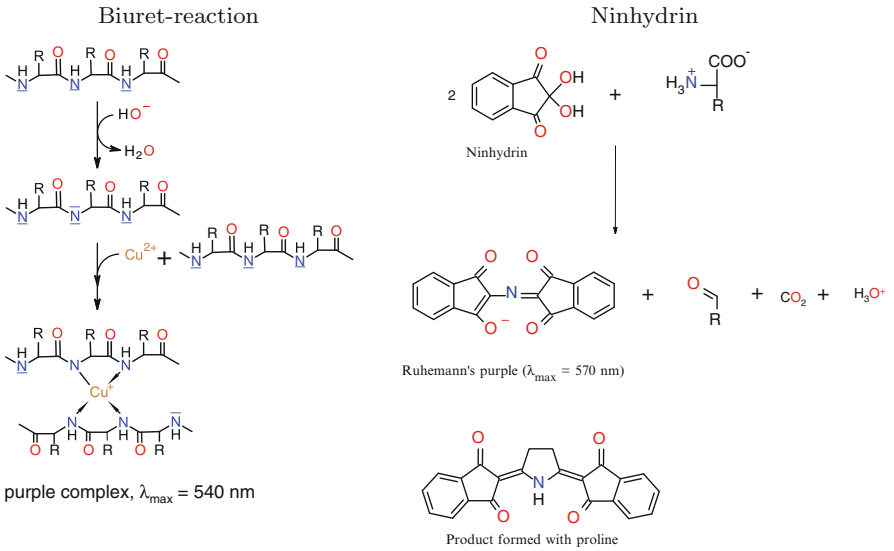


Fig. 3.9 *Left*: Biuret reaction of proteins. The Cu^{2+} is reduced by the protein in alkaline solution to Cu^{+}, which forms a purple complex with the protein. The Cu^{+} can undergo further reactions, which are the basis of the BCA and LOWRY-tests. *Right*: Reaction of amino acids and other primary amines with ninhydrin. Note that Pro is a secondary amine and gives a different reaction, turning yellow instead of purple

The LOWRY Method [22] Similar to biuret, but more sensitive. The Cu^{+} produced in the biuret reaction and the tyrosine residues in the protein react with molybdophosphoric acid (Mo^{6+}), forming molybdenum blue (Mo^{4+} and Mo^{5+}). This assay is very commonly used, but time-consuming. Colour yield depends on the Tyr-content of the protein. It is interfered with by complex-forming and reducing agents, detergents, and many other common chemicals.

BCA-reaction The Cu^{+} formed in the biuret reaction forms an intensively purple complex with bichinchonic acid (BCA) [32]. There is interference by complex-forming and reducing chemicals, but not by detergents.

The ninhydrin reaction Ninhydrin reacts with primary amino groups to yield a purple product. It is used for free amino acids (see Fig. 3.9), also proteins after hydrolysis.

Fluorescent amine reagents such as OPA and fluoram react with primary amino groups in proteins. They are used for detection of amino acids after chromatography.

The BRADFORD-assay [3] is based on the fact that proteins bind hydrophobic dyes such as CBB-G250, which have different colours in aqueous and hydrophobic environments. Colour yield depends on the properties of the protein. There is interference by detergents, but not by reducing or complex-forming agents.

3.3 Protein Sequencing

An important step in the characterisation of a new protein is the determination of its primary structure. If the protein is simply hydrolysed, the amino acids can be separated by chromatography and quantified. This gives the amino acid composition, but not the sequence. For this, the amino acids need to be removed one at a time.

3.3.1 Edman Degradation

The classical way to determine the primary structure of a protein is EDMAN-degradation (see Fig. 3.10).

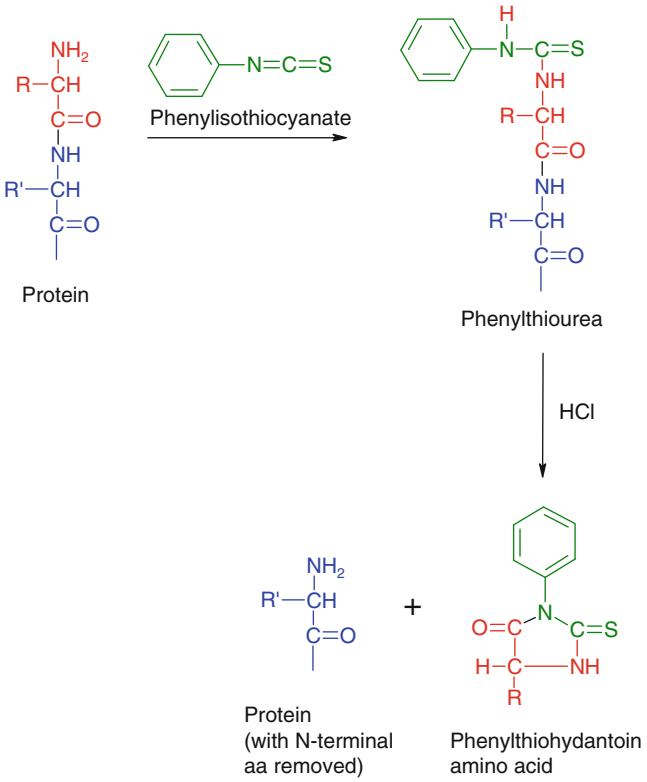


Fig. 3.10 EDMAN-degradation of a protein. In each cycle one amino acid is cleaved off from the N-terminus and identified. The remaining protein can go directly to the next cycle

The protein is bound to a matrix and is N-terminal amino acid converted to a Phenyl isothiocyanate (PITC) derivative, which is cleaved off with HCl; this is done automatically (in modern equipment the reactants are in the gas phase rather than in solution). Because the efficiency of the reaction is only about 80–95 %, the signal diminishes each cycle. Under optimal conditions, up to 50 amino acids can be sequenced, about 100 pmol of purified protein is required.

Larger proteins are split into peptides which are sequenced, using different proteases and reagents to create overlapping sequences. These sequences are put together like a jigsaw puzzle. This can be a very time-consuming process. Such digestion may also be used if the N-terminus of the protein is "blocked" by chemical modification:

Cyanogen bromide (CNBr) in 70 % formic acid splits C-terminal of Met. Careful: highly toxic!
Controlled hydrolysis by 250 mM acetic acid 8 h at 110 °C splits proteins C-terminal of Asp.
Iodosobenzoic acid splits C-terminal of Trp.
Trypsin splits proteins C-terminal of Arg and Lys.
Chymotrypsin cuts after Tyr, Phe and Trp.
V8-protease splits C-terminal of Glu.

Today protein sequences are usually determined by sequencing their genes, which is quicker and easier. Indeed, great effort has been spent in the last couple of years in sequencing the entire genome of bacteria, yeasts, plants, animals, and humans [7, 36]. With these data available it is often sufficient to sequence the first 10 or 20 amino acids of a protein and then look for the gene sequence in a computer database.

With all the data coming in from gene sequencing attention has now shifted to identify not only the proteins expressed in a cell and their posttranslational modifications, but also how they interact with each other (the **interactom**). It appears that proteins form well-defined, cell-type specific networks with each other. These can be studied, for example, by immunofluorescent microscopy of whole cells. Appearance of a protein in an unusual spot can be a valuable diagnostic marker. In such cases the total amount of protein may not change, only the interaction of the protein with its neighbours. Such interactions are lost when the cells are homogenised.

3.3.2 Mass Spectrometry

In Mass spectrometry (MS), atoms or molecules are ionised and accelerated in an electrical field (ion source). The ions are then separated by their mass/charge (m/z) ratio (analyser) and detected (detector and recorder). Various types of ion sources and analysers are available, so that a suitable pair for the task at hand can be selected.

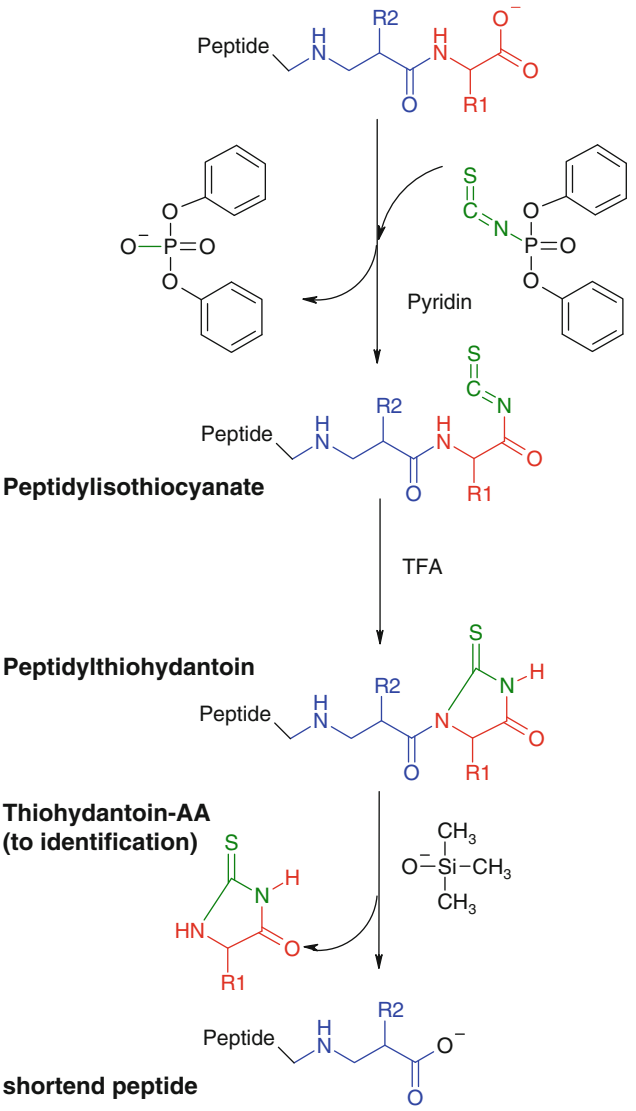


Fig. 3.11 Similar to N-terminal sequencing by the EDMAN-procedure it is also possible to sequence from the C-terminal end [15]. Once the protein has been sequenced at both ends it is possible to find the entire sequence by molecular biology methods

Ion mobility spectrometers (sniffers, see Fig. 3.12) are used for real-time analysis under field conditions.

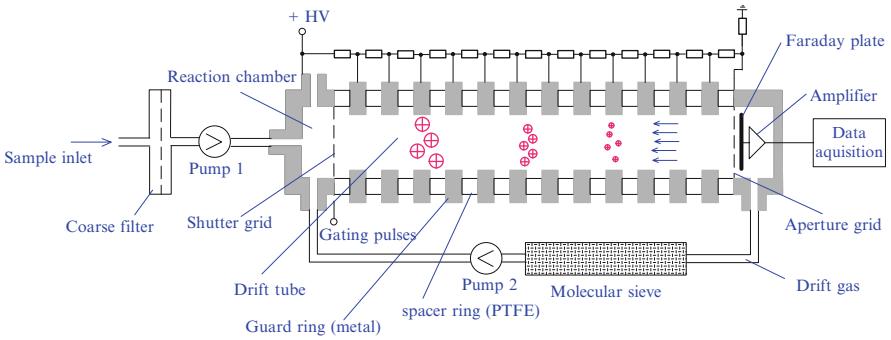


Fig. 3.12 The ion mobility spectrometer is used to detect and identify contaminants in air, for example, agents of biological or chemical warfare. Sample molecules are ionised and move against a stream of drift gas. Small molecules will experience less resistance and arrive at the FARADAY plate earlier than large molecules. The entire system (computer, gas supply etc.) can be mounted onto a small vehicle and used for real-time analysis, say, after a terrorist attack or industrial accident

3.3.2.1 Tandem Mass Spectrometry

If different compounds have the same m/z ratio, they cannot be distinguished by simple mass spectrometry. Tandem mass spectrometry solves this problem.

All ions of a given m/z ratio are fed into a collision cell, where a "collision gas" (usually helium or argon under reduced pressure) collides with them. The collisions induce breaks in the sample molecules; i.e., each species present in the sample is disintegrated into a set of fragments, which are fed into a second mass spectrometer. There the fragments are analysed for their m/z ratio. Each compound under these conditions will give a specific set of fragments (i.e., a specific spectrum), as breakage occurs preferentially on certain bonds. The combination of molecular mass and fragmentation spectrum uniquely identifies a substance, which can be used for forensic purposes.

Tandem mass spectrometry can be used, for example, in newborn screening for inherited diseases. Blood or urine samples from the babies are analysed for unusual accumulation of metabolites (see Fig. 3.13). Similarly, toxic substances can be identified and quantified in samples of emergency room patients quickly, so that intervention is still possible.

3.3.2.2 Protein Sequencing by Tandem MS

The important point about tandem MS is that proteins tend to fracture at peptide bonds. So the fragment spectrum of a protein will consist of two series of peaks (N- and C-terminal), each differing from its neighbour by the molecular mass of one amino acid. Because all amino acids except Leu/Ile have different molecular mass, the sequence of the protein can be determined by tandem-MS. Indeed, because the

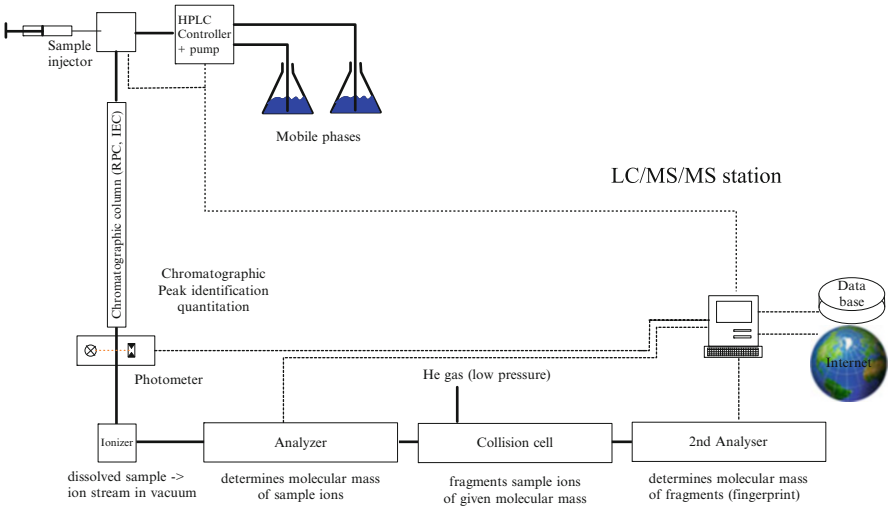


Fig. 3.13 Liquid chromatography coupled to tandem mass spectrometry (LC/MS/MS) can be used to identify and quantify unusual compounds in biological fluids. The sample is first separated on a chromatographic column (gas chromatography may be used for volatile compounds), and then the mass of compounds is determined by MS. Finally, the fingerprint spectrum of each compound of interest is determined. Together, these data identify a compound "beyond reasonable doubt" even in court

spectrum will be generated both by the N- and C-terminal fragments, the sequence is determined twice and the results can be checked for agreement.

The tedious task of converting a spectrum into a sequence is done automatically by computer. The computer will also take care of the complications such as chain breakage in places other than peptide bonds or uneven distribution of charges between the fragments.

If a protein is too large for this approach, it can be digested with endoproteases and the resulting peptides analysed. This digestion can be performed, for example, directly in a protein spot isolated from a 2d-electrophoresis gel. The peptides are then eluted and analysed by MS, an approach frequently used in **proteomics**.

3.3.2.3 Special Uses of MS

Electrospray ionisation is gentle enough to leave bound substrates on the protein. Thus ESI-MS can be used to monitor substrate binding and turnover. The same is true for protein–protein associations, for example, homo-oligomer formation.

Mass spectrometry may also be used to monitor changes in protein association, for example, after hormone stimulation of a cell. For this purpose cells are grown on [^{13}C]-Arg, stimulated with the hormone, and lysed. The lysate is mixed with a lysate from unstimulated cells grown with [^{12}C]-Arg. The receptor with any bound proteins is purified from the mixed lysate by affinity chromatography and subjected to mass spectrometry. Proteins from stimulated and unstimulated cells can

be distinguished by their mass difference. If the interaction of a protein with the bait is unaffected by hormone stimulation, the height of the corresponding protein peaks will be equal; if interaction increases after hormone stimulation, the ratio $^{13}C/^{12}C$ will be > 1; if interaction is reduced, it will be < 1 [2].

3.3.3 Phylogenetic Trees

Our current model of biological evolution states that all life on earth ultimately descends from a single **primordial cell**. Different organisms evolved from this cell by a series of mutations, which changed their shape, behaviour, and ecology. One aim of biologists is to understand retrospectively the sequence of events, that is, to draw evolutionary trees that show how the organisms that currently live, and those that we know as fossils, are related to each other. Classically, this has been done by looking at the anatomical structures of organisms and how they are related, for example, how mammalian inner ear bones developed from fish jaws.

Mutations in the sequence of the DNA of an organism will change the sequence and expression patterns of its proteins. Thus an alternative method to construct evolutionary trees is to compare the sequence of the same protein in a variety of organisms.

3.3.3.1 Alignment and Distance Matrix

Have a look at Fig. 3.14. You see an oversimplified example of such an evolutionary tree. Now assume that you know only the sequences, but not the tree itself, which you wish to reconstruct.

In order to do so, you have to compare the sequences and calculate a matrix of the number of changes between all sequences (note: an insertion or deletion counts as one change, even if several amino acids are involved).

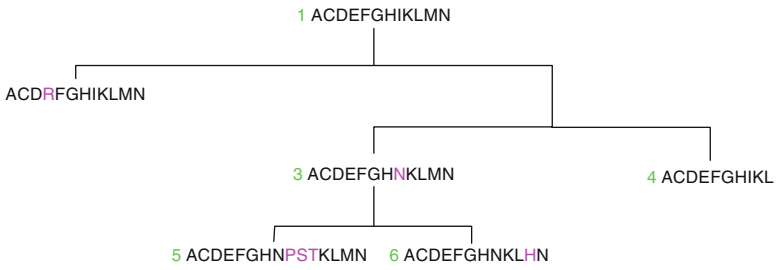


Fig. 3.14 Phylogenetic tree of a hypothetical protein sequence. Substitutions, insertions, and deletions create a variety of sequences from a single ancestor

	1	2	3	4	5	6
1	0	1	1	1	2	2
2		0	2	2	3	3
3			0	2	1	1
4				0	2	2
5					0	2
6						0

3.3.3.2 Problems and Limitations

With this information you can try to reconstruct the evolutionary events [33]. If you do so, you will notice a number of problems:

- There are several ways to draw the tree; in particular you cannot tell which of the sequences is the original one, i.e., the root of the tree.
- In practice, sequences may be directly linked even though there is more than one mutation between them. That just tells you that either there were indeed two mutations that occurred at the same time, or, more likely, that there is a missing link between them.
- The same site may have mutated twice. Although this is an unlikely event, over 4×10^9 a of evolution even rare events do happen.
- The rate of mutation may not be constant over time. Thus it is not necessarily possible to calculate the age of species from such trees.

A partial solution to these problems is to sequence more proteins, so that more information is available.

3.3.3.3 Proteins Have Different Rates of Mutation

Cytochrome c, a protein needed to burn food, for example, will mutate only very slowly, because most mutations will result in a nonfunctional protein. On the other hand, cytochrome c is present in all aerobic organisms, from bacteria to man. Sequence alignment of this protein will tell you something about long-term evolution (on the phylum or class level), but organisms within an order or family may all share the same sequence.

Serum albumin is subject to much lower evolutionary pressure, as long as it is water soluble and has some hydrophobic pockets to bind hormones it will do. Thus the serum albumins can be used to reconstruct short-term evolution, for example, how mammals evolved from each other. Long-term evolution cannot be read from such alignment, if only because many organisms do not have serum albumin.

3.3.3.4 Other Approaches to Distance Matrix Calculation

Modern approaches can calculate a distance matrix without alignment, which is computationally expensive.

Working on a computer, you have certainly used compression programs such as 7zip, bzip2, stuffit and the like. One of the algorithms used by these programs is to look for stretches of identical information. In this case only a reference to the first occurrence needs to be given, rather than repeating that information.

If two protein sequences are compressed together, the resulting file is smaller than the sum of the separately compressed sequences. The difference is the larger, the more similar the sequences are [21]. This is equivalent to calculating the amount of information in the difference between the two sequences. For another useful approach to calculate a homology matrix see [6]. Further information on the use of protein sequences for phylogeny may be found, for example, in [20, 25].

3.4 Synthesis of Peptides

In the last chapter we have seen how to determine the sequence of a protein. For many purposes it would be useful to synthesise proteins of a given sequence. The solid-phase approach of MERRIFIELD (Fig. 3.15) serves this purpose.

The process can be automated, and several companies offer peptide synthesisers that use MERRIFIELD's chemistry. Peptides of up to about 100 amino acids can be produced this way; addition of each amino acid takes about 30 min. Compare this with cellular biosynthesis, which produces a 1000-amino acid protein in about 5 min!

Other companies now offer the synthesis of peptides as a service for a very reasonable price. Thus it is usually not cost effective to perform this synthesis in one's own laboratory, unless one has to do it routinely.

Modern approaches of **combinatorial chemistry** use special paper as solid support instead of polystyrene beads. The reaction is performed on small spots of this paper: on each spot a slightly different peptide is synthesised. Thus several hundred or even several thousand peptides can be tested for a particular biological activity.

3.5 How Do We Determine Secondary Structure?

Three experimental methods are available for the determination of protein secondary structure: X-ray diffraction and electron microscopy of protein crystals and Nuclear magnetic resonance (NMR) of proteins in solution. Structures are accessible via the Internet on the PDB, which has grown dramatically over the last 40 years (see Fig. 3.16). For more details on these methods see [5].

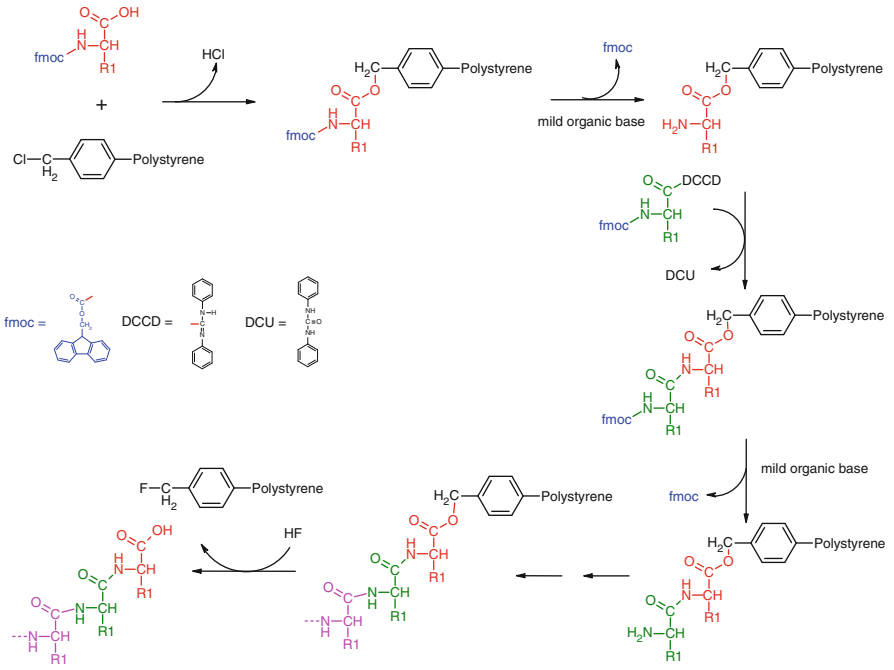


Fig. 3.15 Solid phase peptide synthesis by the method of MERRIFIELD. The amino-groups of the amino acids are protected with a FMOC- group. Then the carboxy-groups are activated with DCCD, so it may be coupled to a free amino group. After each coupling step the FMOC-protection group is removed with a mild organic base. Note that the synthesis goes from the C- to the N-terminus, the reverse order of biological protein synthesis. This should be kept in mind when talking to a peptide chemist! Finally the crude peptide is removed from the solid support with HF; it then needs to be purified, usually by reversed phase chromatography on a C_{18}-column

3.5.1 X-Ray Crystallography

In a crystal the atoms and molecules have a very regular pattern. If an X-ray beam is passed through the crystal, the pattern acts as a grid and diffraction occurs. The diffraction pattern is recorded on a photographic film or (nowadays) a solid-state detector. From this diffraction pattern it is possible in principle to calculate the position of the atoms in the protein molecule, [28] describes how this is done.

Protein crystals need to be grown from a solution [23], but even for soluble proteins it is often difficult to get good crystals that give diffraction patterns with the required resolution to trace at least the peptide backbone (≈ 3 Å). Nevertheless, considerable progress has been made in recent years.

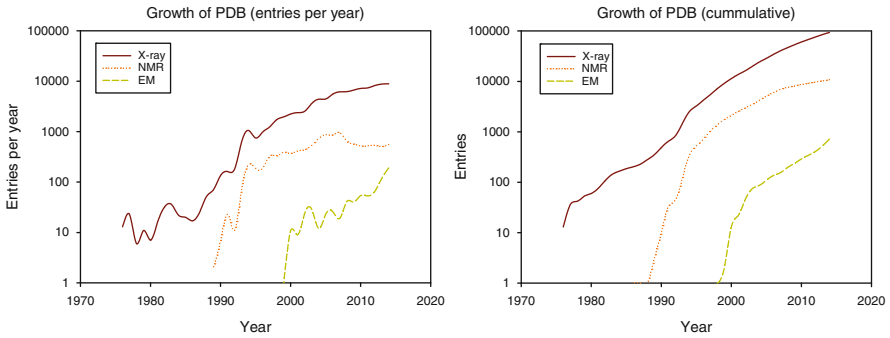


Fig. 3.16 From humble beginnings. PDB was started in 1976 with 13 resolved structures, all determined by X-ray crystallography. As of 2015-03-17, it contains 107 436 entries, determined by either X-ray crystallography (89.5 %), NMR (10.3 %), or electron microscopy (0.2 %). Data were obtained from http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

3.5.2 Electron Microscopy

For the proteins that occur in a biological membrane it is even more difficult to get crystals (the membrane has to be dissolved with detergents and the protein-detergent complex crystallised). However, such proteins can form two-dimensional arrays ("2-D crystals") in the membrane, which can be investigated using the electron microscope. Because the molecules in such a 2-D crystal all have the same orientation, it is possible to calculate an "average" image of a protein molecule. Unfortunately, the resolution of such images is in most cases limited to 10–20 Å, which is not enough to trace the peptide backbone of the protein (this is now improving; see Fig. 3.8 for an example). Some larger structures such as the proteasome can be investigated directly in the cell [24].

3.5.3 Nuclear Magnetic Resonance

Atomic nuclei with an odd number of nucleons (^{1}H, ^{13}C, ^{31}P), or an odd number of both protons and neutrons (^{14}N) have an angular momentum called spin. This spin of a positively charged body results in the generation of a tiny magnetic field, which is measured by an NMR spectrometer [27]. Because part of the field is shielded by the electrons surrounding the atoms, NMR can detect the kind of bonds that an atom is involved in, and hence the secondary and tertiary structure of proteins (see Fig. 3.17).

Fig. 3.17 Structure of ubiquitin as determined by NMR (PDB-code 1d3z). Because the structure is determined from measured interatomic distances often several different models (here 10) fit the data. Although in α-helices and β-strands the solutions often lay close together, coils are less well determined. Ubiquitin is the archetype for a whole group of small proteins, the ubiquitin-like modifiers, which are posttranslationally attached to proteins. The fold is called the ubiquitin-cradle, as the β-sheet cradles the α-helix

NMR used to be limited to small proteins up to about 20 kDa, but specialised methods can now obtain useful spectra of proteins approaching 1 MDa [10]. Unfortunately, NMR instruments are very expensive and require a considerable sample size (several mg of pure protein).

However, it is possible to investigate the structure of proteins in solution; no crystallisation (which might result in artifacts) is required. One would therefore expect results of greater biological significance than those obtained by X-ray diffraction. A comparison of structures determined by the two methods shows fortunately that they are fairly similar, giving us faith in both.

3.5.4 Computer Predictions

The biggest problem for the determination of protein structure is the vast amount of highly purified protein required ($\approx$ 100 mg) for any of the above methods. Many proteins of particular biological interest occur in very small amounts in our body. In the last 30 years the sequences of many proteins have been determined by DNA sequencing (an almost complete sequence of the human genome was published in 2001 [7, 8, 36]). This gives us the chance to produce interesting proteins in larger quantities by genetic engineering, alleviating the sample size problem somewhat. However, experimental structure elucidation is still a difficult, tedious, and expensive business.

This could be circumvented if it were possible to predict the secondary structure of a protein from its known primary structure. After all, a protein can fold unaided

(at least under ideal circumstances). Thus all information necessary for folding must be contained in its primary structure (this is known as the ANFINSEN-hypothesis [1]).

This has in part to do with bond angles, for example, Pro rarely occurs in β-sheets, because the nitrogen is part of a ring structure and ϕ cannot assume the value required. On the other hand, bulky R-groups can prevent the tight packing of amino acids required for the formation of α-helices. Secondary structures are stabilised, if R-groups with opposite charge are brought close together; they are destabilised by R-groups with the same charge.

Attempts to use these principles for computer-aided structure prediction have been made since the 1970s (e.g., [13, 30]). The idea is to look at proteins with known secondary structure and to determine which amino acids (or amino acid combinations) occur most frequently in a particular secondary structure, and to use this statistical information to predict the secondary structure of newly sequenced proteins. Some progress has been made in this field, and for soluble proteins the secondary structure at any amino acid can be determined with about 75 % accuracy. However, this means that a quarter of the amino acids in a protein will be assigned wrong secondary structures. Because the secondary structure of only a few membrane proteins has been solved and the tables derived for soluble proteins may not apply, structure prediction for them is even more uncertain. Such predictions should therefore be viewed with due caution. They serve better to generate hypotheses to guide experiments than theories to be published!

If the structure of a related protein (with similar sequence) has been solved, one can also try to "thread" the unknown protein onto this template. This method is somewhat more reliable than prediction by statistical methods and has become more and more important as the number of known protein structures available for comparison increased.

Internet Resources
http://www.predictprotein.org/ and http://bioinf.cs.ucl.ac.uk/psipred/ are servers that offer protein structure prediction for submitted sequences, using different algorithms.

3.6 Exercises

3.6.1 Problems

3.1. Sickle cell anæmia, an inherited disease, is caused by the mutation Glu6Val in the β-subunit of hemoglobin. Compared to the normal protein you would expect the mutated protein to move in a native electrophoresis experiment (i.e., without SDS):

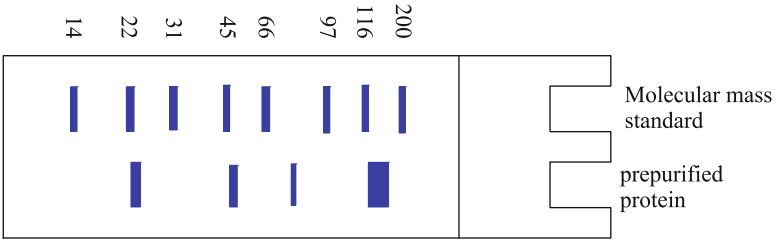
A) less to (+) at pH 8, same distance at pH 1
B) same at pH 8, more to the positive at pH 1
C) more to (+) pH 8, more to the (−) at pH 1
D) less to the (−) at pH 8, less to the (−) pH 1
E) same distance under all conditions

3.2. Blood serum contains many different proteins at a fairly constant concentration (6.0–7.8 g/dl). However, in several serious diseases protein concentration is lowered (e.g. liver cirrhosis) or elevated (e.g. multiple myeloma). A quick way to determine the serum protein concentration is to measure the UV-absorbance.

A 1:100 dilution of serum gives an absorbance of 0.4 at 280 nm in a standard cuvette of 1 cm path length. Assume a molar extinction coefficient of 4×10^4 L mol^{-1} cm^{-1} and an average molecular mass of 65 kDa for blood proteins. The protein concentration is approximately

A 6.0 g/dL
B 6.5 g/dL
C 7.0 g/dL
D 7.5 g/dL
E 8.0 g/dL

3.3. You are helping to treat a patient with a mutation in protein X, which renders this protein nonfunctional. The molecular mass of X is 150 kDa. You have expressed the wild-type protein in large quantities in *E. coli* with a suitable tag which allowed you to purify it by affinity chromatography. The product gives the depicted SDS-PAGE gel.



Which of the following methods should be used to completely purify the product, before you inject it into the patient?

A gel filtration
B anion exchange chromatography
C cation exchange chromatography
D hydroxyapatite chromatography
E hydrophobic interaction chromatography

3.6.2 Solutions

3.1 At *p*H 8.0 Glu is negatively charged, whereas Val is not. A Glu→Val mutation will reduce movement towards the positive pole.

At *p*H 1.0 Glu is uncharged, as is Val. Hence the mutation has no effect on electrophoretic mobility at this *p*H.

3.2 Light absorption A as a function of concentration c and path length l is described by LAMBERT-BEER's law $A = -\log\left(\frac{I}{I_0}\right) = \epsilon_\lambda c\, l$, where ϵ_λ is the extinction coefficient at wavelength λ. In our case the sample was diluted 100-fold, as LAMBERT-BEER's law is valid for most substances only if $A \leq 1.0$.

Hence the molar protein concentration $c = 100 \times \frac{A}{\epsilon_\lambda l} = 100 \times \frac{0.4\,\text{mol cm}}{4 \times 10^4 \times 1\,\text{L cm}} = \frac{40}{4} \times 10^{-4}\,\text{mol/L} = 1.0 \times 10^{-3}\,\text{mol/L}$.

Inasmuch as the average molecular mass is 65 kDa, and hence 65×10^3 g is equivalent to 1 mol, we can convert the molar concentration into a weight concentration: $1.0 \times 10^{-3}\,\text{mol/L} \times 65 \times 10^3\,\text{g/mol} = 65\,\text{g/L} = 6.5\,\text{g/dL}$, which is in the normal range.

3.3 The proteins differ in molecular mass. Hence one would choose a method separating by this parameter, that is, gel filtration.

References

1. C. Anfinsen, Principles that govern the folding of protein chains. Science **181**, 223–230 (1973). doi: 10.1126/science.181.4096.223
2. B. Blagoev, I. Kratchmarova, S.-E. Ong, M. Nielsen, L.J. Foster, M. Mann, A proteomics strategy to elucidate functional protein-protein interactions. Nature Biotech. **21**, 315–318 (2003). doi: 10.1038/nbt790
3. M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein-dye binding. Anal. Biochem. **72**, 248–254 (1976). doi: 10.1016/0003-2697(76)90527-3
4. E. Buxbaum, Cationic electrophoresis and electrotransfer of membrane glycoproteins. Anal. Biochem. **314**, 70–76 (2003). doi: 10.1016/S0003-2697(02)00639-5
5. E. Buxbaum, *Biophysical Chemistry of Proteins: An Introduction to Laboratory Methods* (Springer, New York, 2011). ISBN 978-1-4419-7250-7
6. K.-C. Chou, Y.-D. Cai, Predicting enzyme family class in hybridisation space. Protein Sci. **13**, 2857–2863 (2004). doi: 10.1110/ps.04981104
7. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. Nature **409**, 860–921 (2001). doi: 10.1038/35057062
8. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. Nature **431**, 931–945 (2004). doi: 10.1038/nature03001
9. M.I. de Michelis, R.U. Spanswick, H^+-pumping driven by the vanadate-sensitive ATP-ase in membrane vesicles from corn root. Plant Physiol. **81**, 542–547 (1986). doi: 10.1104/pp.81.2.542
10. J. Fiaux, E.B. Bertelsen, A.L. Horwich, K. Wüthrich, NMR analysis of a 900K GroEL–GroES complex. Nature **418**(6894), 207–211 (2002). doi: 10.1038/nature00860

11. S. Fleischer, M. Kervina, Subcellular fractionation of rat liver. Meth. Enzymol. **31**, 6–41 (1974). doi: 10.1016/0076-6879(74)31005-1
12. Y. Fujiki, A.L. Hubbard, S. Fowler, P.B. Lazarow, Isolation of intracellular membranes by means of sodium carbonate treatment: Application to endoplasmic reticulum. J. Cell Biol. **93**, 97–102 (1982). doi: 10.1083/jcb.93.1.97
13. J. Garnier, D.J. Osguthorpe, B. Robson, Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**(1), 97–120 (1978). doi: 10.1016/0022-2836(78)90297-8
14. K.E. Geckeler, H.Eckstein, *Bioanalytische und Biochemische Labormethoden* (Vieweg, Braunschweig - Wiesbaden, 1998). ISBN 978-3-6429-3589-3
15. K. Graham, J.E. Shively, Improved initial yields in C-terminal sequence analysis by thiohydantoin chemistry using purified diphenylphosphoryl isothiocyanate: NMR evidence for a reaction intermediate in the coupling reaction. Anal. Biochem. **307**(2), 202–211 (2002). doi: 10.1016/S0003-2697(02)00025-8
16. T. Hisabori, K. Inoue, Y. Akabane, S. Iwakami, K. Manabe, Two-dimensional gel electrophoresis of membrane-bound protein complexes, including photosystem I, of tylacoid membranes in the presence of sodium oligooxyethylene alkyl ether sulfate / dimethyl dodecylamine oxide and sodium dodecyl sulfate. J. Biochem. Biophys. Meth. **22**(3), 253–260 (1991). doi: 10.1016/0165-022X(91)90073-6
17. P. James, Protein identification in the post-genome era: the rapid rise of proteomics. Quart. Rev. Biophys. **30**, 279–331 (1997). URL http://journals.cambridge.org/article_S0033583597003399
18. J. Klose, U. Kobalz, Two-dimensional electrophoresis of proteins: An updated protocoll and implications for functional analysis of genomes. Electrophoresis **16**, 1034–1059 (1995)
19. U.K. Laemli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680–685 (1970). doi: 10.1038/227680a0
20. P. Lemey, M. Salemi, A.M. Vandamme, *The Phylogenetics Handbook: A Practical Approach to DNA and Protein Phylogeny*, 2nd edn. (Cambridge University Press, Cambridge, 2009). ISBN 978-0-5217-3071-6
21. M. Li, J.H. Badger, X. Chen, S. Kwong, P. Kearney, H. Zhang, An information-based sequence distance and its application to whole mitochondrial genome phylogeny. Bioinformatics **17**, 149–154 (2001). doi: 10.1093/bioinformatics/17.2.149
22. O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**, 265–275 (1951). URL http://www.jbc.org/content/193/1/265.full.pdf+html
23. A. McPherson, *Crystallization of Biological Macromolecules* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999). ISBN 978-0-8796-9527-9
24. O. Medalia, I. Weber, A.S. Frangakis, D. Nicastro, G. Gerisch, W. Baumeister, Macromolecular architecture in eukariotic cells visualized by cryoelectron tomography. Science **298**, 1209–1213 (2002). doi: 10.1126/science.1076184
25. D. Mount, *Bioinformatics. Sequence and Genome Analysis*, 2 edn. (Cold Spring Harbor Laboratory Press, Woodbury, NY, 2004). ISBN 978-0-8796-9712-9
26. L. Ornstein, Disk electrophoresis: I. Background and theory. Ann. N. Y. Acad. Sci. **121**, 321–351 (1962). doi: 10.1111/j.1749-6632.1964.tb14207.x
27. H. Rattle, *An NMR Primer for Life Scientists* (Partnership Press, Fareham (GB), 1995). ISBN 978-0-9516-4363-1
28. G. Rhodes, *Crystallography Made Crystal Clear*, 3rd edn. (Academic Press, Amsterdam, 2006). ISBN 978-0-1258-7073-3
29. J.L. Rigaud, Membrane proteins: functional and structural studies using reconstituted proteoliposomes and 2-D crystals. Braz. J. Med. Biol. Res. **35**, 753–766 (2002). doi: 10.1590/S0100-879X2002000700001. URL http://www.scielo.br/pdf/bjmbr/v35n7/4512.pdf
30. B. Rost, C. Sander, Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. **232**(2), 584–599 (1993). doi: 10.1006/jmbi.1993.1413

31. H. Schägger, W.A. Cramer, G. von Jagow, Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. Anal. Biochem. **217**, 220–230 (1994). doi: 10.1006/abio.1994.1112
32. P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Malla, F.H. Garter, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchonic acid. Anal. Biochem. **150**, 76–85 (1985). doi: 10.1016/0003-2697(85)90442-7
33. P.H.A. Sneath, R.R. Sokal, *Numerical Taxonomy: The Principles and Practice of Numerical Classification* (Freeman, San Francisco, 1973). doi: 978-0-7167-0697-7
34. M. Tswett, Physikalisch-chemische Studien über das Chlorophyll. Die Adsorptionen. Ber. Dtsch. Botan. Ges. **24**(6), 316–323 (1906a). doi: 10.1111/j.1438-8677.1906.tb06524.x
35. M. Tswett, Adsorptionsanalyse und chromatographische Methode. Anwendung auf die Chemie des Chlorophylls. Ber. Dtsch. Botan. Ges. **24**(7), 384–393 (1906b). doi: 10.1111/j.1438-8677.1906.tb06534.x
36. J.C. Venter et al., The sequence of the human genome. Science **291**, 1304–1351 (2001). doi: 10.1126/science.1058040
37. O. Warburg, W. Christian, Isolierung und Kristallisation des Gärungsferments Enolase. Biochem. Z. **310**, 384–421 (1942)

Part II
Enzymes

Chapter 4
Enzymes Are Biocatalysts

Life is an ordered sequence of enzymatic reactions. (R. WILLSTÄTTER)

Abstract Enzymes are biocatalysts that can increase the velocity of a reaction by several orders of magnitude. They have no influence on the equilibrium, because they accelerate both the forward and reverse reaction. Most enzymes are proteins, but some RNA-molecules also have enzymatic properties (ribozymes).

4.1 The Nature of Catalysis

A cube of household sugar (sucrose, saccharose) is indefinitely stable under common conditions, even though its oxydation would liberate a considerable amount of energy (see an introduction to general chemistry (e.g., http://www.rossu.edu/medical-school/students/documents/Basic_Chemistry.pdf) for a definition of the terms used):

$$C_{12}H_{22}O_{11} + 12\,O_2 \rightarrow 12\,CO_2 + 11\,H_2O,\ \Delta G^0 = -5700\,\text{kJ/mol}$$

Even with a match it is not possible to ignite the sugar. A little cigarette ash, however, makes igniting possible. The ash acts as a catalyst. Catalysts have the following properties:

- A catalyst does not change the equilibrium of a chemical reaction, but it increases the speed with which the equilibrium is attained.
- Catalysts act by lowering the activation energy of a chemical reaction. This allows a reaction to take place under milder conditions (lower temperature or pressure, *p*H closer to neutral) than would be possible without the catalyst.
- A catalyst is not used up in the reaction; it can be recovered quantitatively when the reaction has stopped. Even in those instances where catalysts form intermediate compounds with the reactants they are regenerated later (e.g., $A + B + C \rightarrow ABC \rightarrow AB + C$).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_4

4.1.1 A Brief History of Enzymology

Most of the topics in this brief overview are covered in more detail in the following chapters. For further reading into enzymology [2, 4, 13] are suggested. If you have to deal with rate equations in special cases [17] is still useful.

4.1.1.1 Chemical Catalysts

The properties of catalysts were described by the German-Russian chemist GOTTLIEB SIGISMUND CONSTANTIN KIRCHHOFF, who could show in 1812 that the hydrolysis of starch to glucose is accelerated by **acid**, and that the acid is not used up in the process.[1] In 1814 he showed that wheat extract has the same properties. The British chemist HUMPHRY DAVY discovered the catalytic properties of platinum in 1816; the chemist and physicist MICHAEL FARADAY speculated in 1825 that catalysts work by temporarily binding the reactants, thereby increasing their local concentration and orienting them properly. The famous Swedish chemist JÖNS JACOB BERZELIUS then coined the name **catalysis** in 1835 (from the Greek word for degradation).

4.1.1.2 Living Organisms Produce Very Powerful Catalysts

The use of yeasts for the production of wine is a very old technique, first mentioned in the *Codex Hamurabi* around 2100 BC. But also the use of crude enzyme preparations rather than complete living organisms is old: In the fifth song of the *Iliad* by HOMER (around 600 BC) the use of ficin (fig tree extract) for curdling of milk in cheese production is mentioned. The French chemist REAUMUR noted in 1752 that the stomach juice of buzzards would digest meat and soften bone, but leave plant material unchanged, the first demonstration of enzyme **specificity**. SPALLANZANI continued these experiments; he found in 1783 that the active ingredient of stomach juice lost its activity upon storage, showing that it was **unstable**.

The French chemists PAYEN and PERSOZ in 1833 extracted a substance from sprouting barley which was even more effective than acid in hydrolysing starch. This substance they could precipitate from the extract with ethanol; this was the

[1] CORNISH-BOWDEN cites on his website (bip.cnrs-mrs.fr/bip10/fulhame.htm, visited 2014-04-30) an earlier report by the Scottish chemist ELIZABETH FULHAME, who published in 1794 a book, *An Assay on Combustion*, in which she put forward the hypothesis that oxydations use water as a catalyst, the oxygen in water would react with the reducing agent, and the resulting hydrogen with air oxygen so that water was re-formed. She also described the photosensitivity of silver salts, which later fored the basis for photography in the 19th and 20th century. A facsimile of her book was published in 2009 and makes interesting reading for anybody interested in the history of science. No biographic data on ELIZABETH FULHAME seem to be available.

first attempt to **purify** an enzyme. They called it **diastase** (from the Greek word for separation, as this substance separated soluble sugar from insoluble starch). The German scientist THEODOR SCHWANN isolated **pepsin** (from Greek "digestion") from stomach juice in 1834, which could hydrolyse meat much more efficiently than hydrochloric acid, the other known component of stomach juice.

Thus it became clear that catalysts in living organisms allow chemical reactions to proceed faster and under milder conditions than in the test tube. These substances were collectively called **ferments** (because the conversion of starch into sugar is required for fermentation).

At that time the prevailing theory in science was **vitalism**, the assumption that living organisms are distinguished from their surrounding by a "force of life" (Lat. *vis vitalis*). Thus free ferments such as diastase and pepsin could degrade compounds, but synthesis was possible only to organised catalysts inside the cell. WILHELM KÜHNE suggested in 1876 to limit the term **ferment** to those organised catalysts of living cells, and to use **enzyme** (from the Greek word for sour dough) for both the unorganised catalysts in, for example, digestive juices, and the organised enzymes of living organisms.

The German scientists HANS and EDUARD BUCHNER tried in 1897 to preserve a cell-free juice pressed from yeast with sugar, so that they could market it for medical purposes. Much to their surprise they found that yeast juice could turn the sugar into alcohol just as whole yeast cells do; this observation effectively put an end to vitalism. It is now held that metabolism inside cells follows the same rules of chemistry and physics that govern reactions in a test tube (**reductionism**). This assumption has proven fruitful, however, it should always be remembered that these laws are applied in an environment of high order and organisation, which is unlikely to be created from unordered surroundings. This is reflected in the rule "every cell comes from a cell" originally put forward by SPALLANZANI and PASTEUR. In this sense living organisms are indeed special.

Because the distinction between ferments and enzymes has lost its significance these terms are now used synonymously.

4.1.1.3 Mathematical Description of Enzyme Kinetics

In 1902 the French chemist VICTOR HENRI quantitatively described the relationship between substrate concentration and reaction velocity for single substrate reactions in mathematical terms [8]. After PETER LAURITZ SÖRENSEN in 1909 pointed out the importance of hydrogen ion concentration for biological reactions and introduced the logarithmic *p*H-scale and the concept of buffering [18] (the theoretical basis was worked out later by HASSELBALCH [7]), the German chemist LEONOR MICHAELIS and his Canadian postdoc MAUDE LEONORA MENTEN were able to confirm HENRI's result with much higher precision (for a detailed discussion on the HENRI–MICHAELIS–MENTEN law see page 111ff); they also introduced the concept of initial velocity. G.E. BRIGGS & J.B.S. HALDANE generalised the HMM-law in 1925. W.W. CLELAND suggested a systematic approach to multisubstrate reactions and enzyme inhibition in 1963.

Fairly early it was realised that not all enzymes follow the hyperbolic HMM-law, in 1910 HILL [9] suggested a way to extend the HMM-equation to such cases. A mechanistic understanding of this "**cooperativity**" came with the papers of J. MONOD, J. WYMAN & J.P. CHANGEUX in 1965 [14] and D.E. KOSHLAND, G. NEMETHY & D. FILMER in 1966 [12].

4.1.1.4 The Chemical Nature of Enzymes

GERARDUS JOHANNES MULDER coined the term "protein" in 1838, the term is derived from Greek and means "of prime importance". M. TRAUBE suggested in 1877 that enzymes were proteins, a notion that was not generally accepted until J.B. SUMNER crystallised urease in 1926 [19] and J.H. NORTHROP pepsin, trypsin, and chymotrypsin in the 1930s. (The first protein crystallisation (hæmoglobin) was performed by ERNST FELIX IMMANUEL HOPPE-SEYLER in 1864.) Shortly afterwards in 1934 BERNAL & CROWFORD noticed that protein crystals produce well-resolved X-ray diffraction patterns.

In 1955 F. SANGER reported the complete amino acid sequence of insulin and in 1957 J.C. KENDREW the crystal structure of myoglobin from X-ray diffraction, proving that proteins were chemically well-defined molecules. This was confirmed when in 1969 GUTTE & MERRIFIELD chemically synthesised ribonuclease, molecular mass 13.7 kDa [6]. The protein had enzymatic activity, finally proving TRAUBE's hypothesis.

In 1980 THOMAS R. CECH discovered an RNA molecule in *Tetrahymena thermophila* which has enzymatic properties (**ribozyme**). Today we know that although the number of ribozymes is small, they are necessary for life. For example, the actual peptide synthase in ribosomes is an RNA [5].

Indeed, RNA is probably a molecule with a lot of surprises in store. Thus it was recently discovered that the mRNA for D-glucosamine-6-phosphate synthase in certain bacteria contains a GlcN-6-P binding site; in the presence of this metabolite the mRNA digests itself, preventing unnecessary synthesis of the enzyme [20]. Noncoding RNA has made many headlines in the last 10 years or so; its role in metabolic regulation is so exciting that people speak of *Noncodarnia*, feeling that they walked through a wardrobe into a strange new world [15]. JACOB & MONOD have speculated that the earliest life on earth was based exclusively on RNA (**RNA-world hypothesis**), and that DNA and proteins are later additions [10].

It was noted by BETRAND in 1897 that some enzymes required dialysable, nonproteinaceous cofactors; their nomenclature is given in Fig. 4.1.

4.1.1.5 Enzyme Mechanism

The German chemist EMIL FISCHER had suggested in 1894 that the specificity of enzymes was the result of a binding site in the enzyme, into which the substrate fitted like a key into a lock (**lock-and-key**-hypothesis). LINUS PAULING proposed in 1948

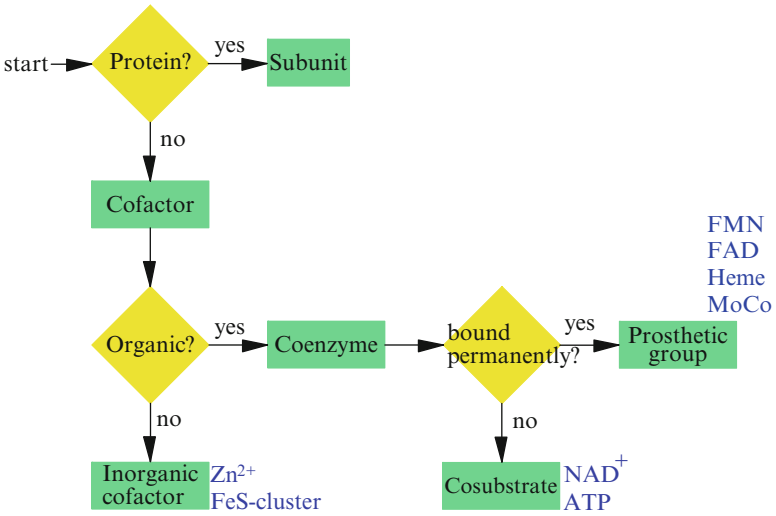


Fig. 4.1 Nomenclature of cofactors

that enzymes act by stabilising the transition state of the substrate. D.E. KOSHLAND realised in 1959 that the substrate binding site was not preformed, but that during substrate binding the enzyme changed its conformation to fit the substrate molecule (**induced-fit**-hypothesis, [11]). Thus he could explain, for example, that in some enzymes with two substrates the second substrate is bound only after the first one.

The actual existence of the enzyme-substrate complex was demonstrated by B. CHANCE in 1943 when he measured the changes in the absorption spectrum of the hæme group in peroxydase following the binding of hydrogen peroxyde [3].

4.2 Enzyme Classification and EC Code

Each cell contains several thousand different enzymes, each a specialist that performs one reaction on one substrate (or at least a small group of similar substrates). Originally each enzyme was given a name by its discoverer, and many such names are still in use today, such as trypsin, pepsin, or invertase. However, as the number of known enzymes increased, this "system" became untenable.

Today enzymes are named in a systematic way. The name of the substrate is followed by the name of the reaction performed on it and the ending -ase (following a suggestion of E. DUCLAUX in 1898). For example, the enzyme that breaks down acetylcholine into acetic acid and choline in our synaptic gaps is called acetylcholine esterase. If several substrates are used, they are listed, separated by a colon, such as NAD:alcohol dehydrogenase. Water as substrate is, however, left out (thus it is acetylcholine esterase rather than acetylcholine:water esterase).

Note that according to International Union for Biochemistry and Molecular Biology (IUBMB) rules enzyme names should not be abbreviated. However, because of the frequently inconveniently long names of enzymes, this rule is most often honoured in the breach. Do make sure, however, that you know what an abbreviation means before you use it!

Enzyme Commission (EC) codes are also used for unambiguous designation of enzymes. The Enzyme Commission is a body of experts affiliated with the IUBMB. EC codes consist of four numbers, separated by points, such as 1.2.3.4.

The first number designates the class of reaction that is performed by the enzyme. There are six known classes of enzymes (see also Fig. 4.2):

Oxydoreductases catalyse the transfer of electrons, hydride ions, or hydrogen atoms between molecules, for example, NAD^+:alcohol dehydrogenase.
Transferases catalyse the transfer of functional groups between molecules, for example, ATP:glucose phosphotransferase (glucokinase).
Hydrolases catalyse the transfer of functional groups to water, for example, glucose-6-phosphatase.
Lyases form double bonds by removing functional groups from a molecule, or open them by adding functional groups. An example would be the aconitate hydratase, which produces iso-citrate from aconitate in the KREBS-cycle.
Isomerases transfer functional groups within a molecule, for example, the conversion of glucose-6-phosphate to fructose-6-phosphate by glucose-6-phosphate isomerase.
Ligases use the energy of ATP hydrolysis to form $\mathsf{C{-}C}, \mathsf{C{-}O}, \mathsf{C{-}S}$ or $\mathsf{C{-}N}$ bonds, e.g., the formation of oxaloacetate from pyruvate and bicarbonate by pyruvate carboxylase.

The second number in the EC code denotes the subclass, and the third number further specifies the reaction. Thus ATP:glucose phosphotransferase has the EC-code 2.7.1.1: it is a transferase (2), a phosphotransferase (7), and transfer is to a hydroxy-group (1). The last number has no systematic meaning; it simply allows the unique identification of an enzyme.

Internet Resources
`http://www.chem.qmw.ac.uk/iubmb/` is the website of the IUBMB with all relevant standards. Amongst other things, you can determine the EC-code for an enzyme or obtain information on an enzyme with given EC-code.

Valuable information on kinetic constants, reaction conditions, substrates, inhibitors, and physiological roles of enzymes can be found on BRENDA (`www.brenda-enzymes.info/`). The Kyoto Encyclopedia of Genes and Genomes (KEGG, `http://www.genome.jp/kegg/`) or UniProt (`http://www.uniprot.org`). provide additional information on biochemical pathways, diseases, sequences, and the like.

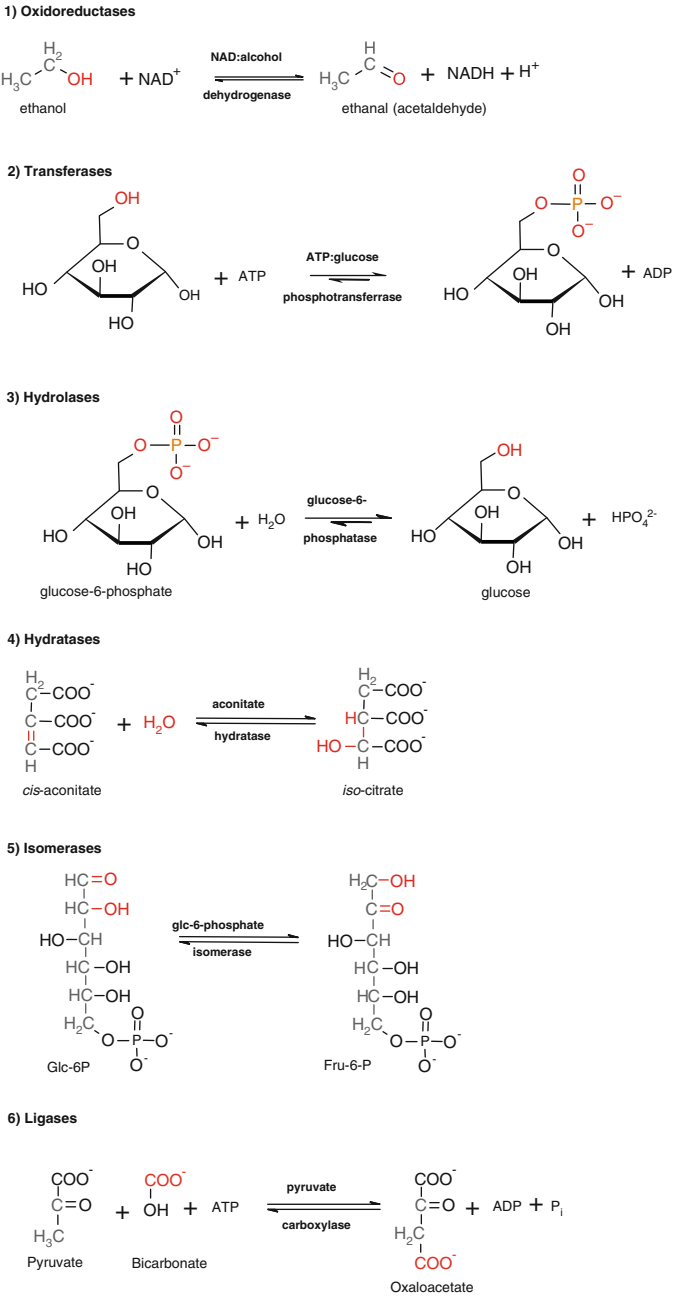


Fig. 4.2 Examples of the reactions performed by the six known classes of enzymes. For details see text

4.2.1 Chemical and Biological Direction of a Reaction

All reactions are, at least in principle, reversible. Catalysts only accelerate the establishment of the equilibrium; they do not influence the equilibrium constant. Chemists by convention write a reaction so that when it proceeds from left to right ΔG^0 is negative, for example:

$$\mathrm{ATP} + \mathrm{H_2O} \rightleftharpoons \mathrm{ADP} + \mathrm{P_i} \qquad \Delta G'^0 = -30.5\,\mathrm{kJ/mol}$$

Within the mitochondria of a cell, however, the reaction runs in the reverse direction, using the energy liberated by oxydation of food to drive it. Thus the biological direction of the reaction is (in this case) opposite to the chemical. Nevertheless, enzymes are named by chemical direction, hence the enzyme responsible is called systematically ATPase rather than ATP synthase (even though the latter is frequently used as a trivial name).

4.3 Inherited Diseases of Metabolism

Partial or complete absence of enzymatic activity results in inborn errors of metabolism (Fig. 4.3). These diseases are individually rare (10^{-3}–10^{-6} per life birth), but because there are 7028 known, together they affect 1 in 2500 live births and make a significant contribution to infant morbidity and mortality (10 % of infant hospitalisations, 20 % of deaths [1]).

Defects in enzymes can lead to

- lack of product.
- accumulation of substrate.
- conversion of substrate to toxic substances by other enzymes.
- gain of function—protein does something it shouldn't. Such diseases are autosomal dominant; that is, persons with one defective copy will be affected.

Such defects are the result of either

inappropriate transcription leading to too few/many copies of mRNA

splicing defect so that an incorrect mRNA is produced. About 15 % of single nucleotide mutations interfere with splicing.

translation e.g. nonsense-mutations resulting in an inappropriate Stop-signal

folding protein folds too slowly and is destroyed by ERAD/proteasome (e.g. cystic fibrosis, see page 450).

amyloid formation improperly folded protein accumulates and interferes with cellular function. Amyloidoses are discussed in Sect. 10.2 on page 206.

defective protein sequence or structure does not allow function. In oligomers this may lead to **dominant negative** inheritance (see, e.g., the discussion of collagen in the chapter on cell matrix on page 324)

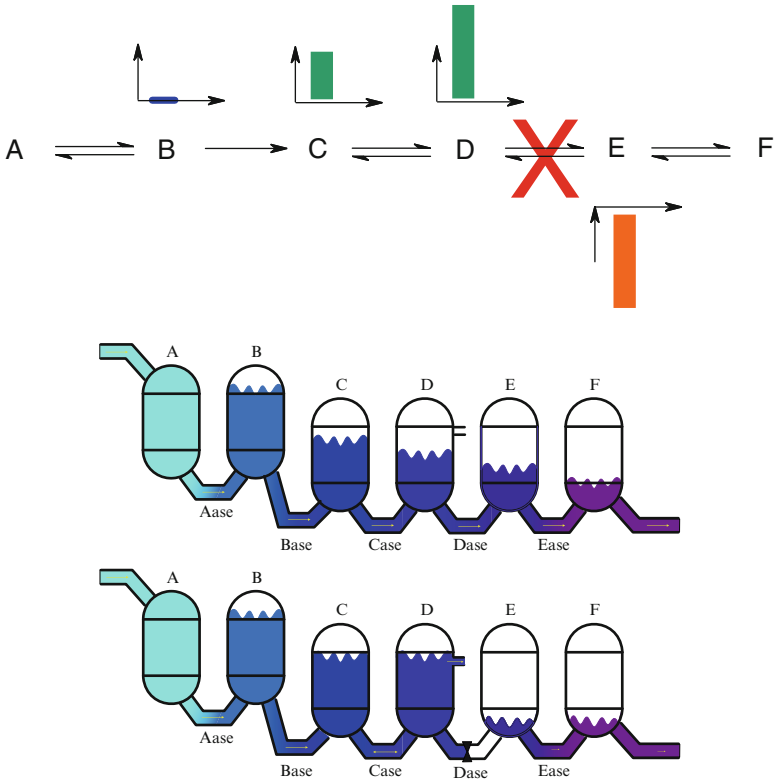


Fig. 4.3 *Top*: Effect of a defective enzyme in an idealised metabolic pathway. Aase, Case, Dase, and Ease catalyse reversible steps, but the reaction of Base is irreversible. Defective Dase leads to a drop in [E] and an increase in [D], which in turn will change [C] because of LE ECHATELIER's principle. However, [B] will not be changed as the conversion of B to C is irreversible. The increase in substrate and decrease in product concentrations can be used to identify the defective enzyme. *Bottom*: Hydraulic model of the pathway. Metabolite concentrations are represented by the fluid level in tanks, and the enzymes by the pipes connecting these tanks. In reversible reactions, tanks are at similar height (representing ΔG). In irreversible reaction the height difference is sufficient that changes in product level do not affect the substrate. Note also the overflow at tank D, which could represent urinary or fecal excretion or conversion by a different pathway

regulation of activity site for posttranslational modification destroyed or protein independent of modification (e.g. oncogenes).

regulation of survival time too low/high concentration of protein.

Most metabolic diseases are **autosomal recessive** (see Fig. 4.4); heterozygotes have 50 % of normal enzymatic activity and appear healthy as metabolite concentrations in our body are usually not determined by kinetics, but by thermodynamics. Thus, carriers have essentially the same substrate and product concentration as homozygote normals. However, they can pass the problem on to their offspring.

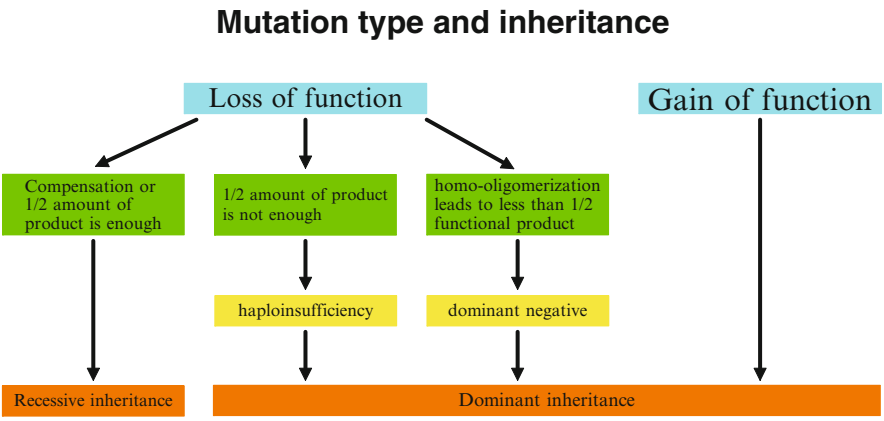


Fig. 4.4 Mechanism for different modes of inheritance

Thus one sees increased frequency in **inbred** populations. **Founder effects** occur if a population descends from few individuals some of whom carried the mutation.

In 104 unrelated, healthy test persons, carrier status for 448 of the 1139 recessive disorders known at the time was measured [1]; there were 0–7 (average 2.8) such defects present per individual. It is, of course, these carriers who pass on inherited diseases to future generations, affected individuals (homozygotes) are usually too sick to do so. In the first half of the twentieth century several countries instituted **eugenics** programs to prevent persons with genetic diseases from reproducing by incarceration, forced sterilisation or, in Nazi Germany, even the murder of "people unworthy of living". You should now see why these programs were not only morally criminal, but also bound to fail on scientific grounds. To eliminate metabolic diseases, we would have to stop carriers from reproducing, and in effect that means all of us. This would eliminate disease at the expense of eliminating the human species[2].

The genetic makeup can predispose to some diseases; those can be relatively frequent. For example, diabetes type I occurs after viral infections by autoimmune destruction of the β-cells in the pancreas of persons of specific HLA-types. For the much more common diabetes type II genetic cofactors, that predispose carriers to disease under conditions of high food availability, are also discussed [16].

Depending on the severity of the enzyme defect **age of onset**, that is, the time it takes for toxicity to build up, can vary. The first symptoms may appear within 24 h of birth or after many years, suddenly or insidiously. We distinguish three main forms:

infantile Results from essentially complete absence of enzymatic function. Mostly neurological symptoms (seizures, dementia, brain stem dysfunction), death frequently occurs within the first year of life.

[2]Please note that humans are a species, not – as one frequently reads – a race.

juvenile Onset in toddlers, intermediate severity.

adult Some remaining enzyme activity, signs are mostly systemic: hepatosplenomegaly, heart and kidney injury, abnormal bone formation, muscle atrophy, and ocular disease. They appear during adolescence.

Pointers to IEoM

The variability in presentation of metabolic disease makes their diagnosis very difficult. The following signs, though unspecific, should raise suspicion if not otherwise explainable:

- lethargy, convulsions, hypotonia
- hepatomegaly, also kardio- or splenomegaly
- acidæmia, increased anion gap
- hyperammonæmia
- unusual colour, smell, or structure of hair, eyes, skin, stool, or urine
- hypoglycæmia
- vomiting
- elevated liver enzymes

Some signs may increase after stress.

Some metabolic diseases are associated with characteristic odours of urine, sweat, or breath. These can be valuable for an astute clinician:

Disease	Pathway	Description
Diabetic ketoacidosis	Glc	Acetone, fruit
Hawkinsinuria	Tyr	Chlorine-like
Isovaleric acidæmia	Leu	Sweaty feet
Maple syrup urine disease	Leu, Ile, Val	Caramel
Oast house syndrome	Met	Drying hops
Phenylketonuria	Phe	Mouse urine
Trimethylaminuria	Choline, carnitine	Fish
Tyrosinæmia I	Tyr	Cabbage
Methylmalonic aciduria	Odd-chain fa, aa	Ammonia
Propionic acidæmia	Ile, Val, Thr, Met	Ammonia
Urea cycle defect	Amino acids	Ammonia

Metabolic diseases have a highly **variable presentation** even for patients with the same mutation (i.e., within kinships), simply because patients are individuals with different genetic makeup. These diseases may also be **phenocopies** of other, nongenetic, diseases. For example, urea cycle defects present like REYE syndrome. As a result, metabolic diseases are underdiagnosed. Many developed countries have **screening** programs where body fluids of all newborns are tested for unusual metabolite concentrations. Some urine screening tests are very old:

$FeCl_3$ blue colour in phenylketonuria, tyrosinæmia, maple syrup urine disease, alcaptonuria, and ketonuria. Interference by salicylates and phenothiazines.

Tests for reducing substances (e.g., FEHLING's or BENEDICT's reagent) detect Glc, Gal, Fru (spontaneous endiol-conversion to Glc + Man under alkaline conditions!) or sialic acid. Interference by cephalothin and ampicillin.

Nitroprusside sodium $Na_2[Fe(CN)_5NO]$ gives a red complex with ketones, cystine[3] and homocysteine.

Azure A metachromatic colour change with mucopolysaccharides

These tests are superseded in developed countries, however, in developing countries they are still useful as they are cheap and can be applied by semi-trained personnel such as village nurses; the reagents are stable and do not require cooling. Patients identified by these tests need to be transferred for further diagnosis and management.

Many hospitals use heel-prick blood for testing, which is taken 24–72 h after birth. Blood drops are placed on GUTHRIE-cards, allowed to air-dry, and mailed to the testing laboratory. The most modern approach to testing involves LC/MS/MS (see Fig. 3.13 on page 82). This method requires considerable investment in the equipment, but can handle high sample loads with minimal staff involvement. As sequencing becomes cheaper, whole genome sequencing of each newborn will likely become normal. Note, however, that this can identify only known mutations; patients with unknown mutations will still require diligent physicians for early diagnosis.

The Online Mendelian Inheritance in Man Database

OMIM is the online version of a list of genetic diseases and their components, first published as book in 1960. It is regularly updated by the *National Center for Biotechnology Information (NCBI)* and an authoritative resource for genetic diseases found at `omim.org`. Each gene or disease is assigned a six-number code. The preceding char means:

- ***** Gene description
- **#** Phenotype description, molecular basis known
- **+** Gene and phenotype, combined
- **%** Phenotype description or locus, molecular basis unknown
- **^** Entry has been deleted or moved to another entry

The first number means:

1. autosomal dominant
2. autosomal recessive

(continued)

[3] Cystine is a compound made from two molecules of cysteine by -S-S- bridge formation. It occurs in urine in the disease cystinuria.

3. X-linked
4. Y-linked (46 entries!)
5. mitochondrial
6. autosomal, defined after 1994

4.4 Exercises

4.4.1 Problems

4.1. Connect the following reactions with enzyme classes:

1)	Ethanol + NAD^+ + $\rightleftharpoons$ ethanal + NADH + H^+	A)	Ligase
2)	Glucose + ATP $\rightarrow$ glucose-6-phosphate + ADP	B)	Oxydoreductase
3)	Glucose-6-phosphate $\rightleftharpoons$ fructose-6-phosphate	C)	Isomerase
		D)	Transferase
		E)	Hydrolase

4.2. A 25 a old man has difficulties regulating his blood sugar level (Maturity onset diabetes of the young (MODY)). Molecular analysis shows that in the enzyme glucokinase—which catalyses the first step of glucose to glycogen conversion in the liver—a leucine residue is replaced by proline. This mutation occurred in a hinge region far away from the active site. What is the most likely functional effect of this mutation?

A None, because it is not in the active site.
B None, because both Pro and Leu are hydrophobic amino acids.
C It will cause the reaction to proceed by a different mechanism.
D It will prevent the induced fit of the substrate to the enzyme.
E It will lower the activation energy of the reaction.

4.4.2 Solutions

4.1

1) The removal of hydrogen from ethanol is an oxydation.
2) In this reaction a phosphate group is transferred from the donor ATP to the acceptor glucose.
3) In this reaction functional groups within a molecule change place, creating an isomer.

4.2 Pro is distinguished from all other amino acids by the fact that the $N-C^{\alpha}$-bond is part of a ring. This limits the dihedral angle ϕ and reduces the flexibility of the protein backbone. This makes conformational changes upon substrate binding more difficult.

References

1. C.J. Bell, D.L. Dinwiddie, et al., Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci. Transl. Med. **3**(65), 65ra4–65ra4 (2011). doi: 10.1126/scitranslmed.3001756
2. H. Bisswanger, *Enzyme Kinetics, Theories and Methods*, 3rd edn. (Wiley-VCH, Weinheim, 2008). ISBN 978-3527319572
3. B. Chance, The kinetics of the enzyme-substrate compound of peroxidase. J. Biol. Chem. **151**, 553–577 (1943). URL http://www.jbc.org/content/151/2/553.full.pdf
4. R.A. Copeland, *Enzymes. A Practical Introduction to Structure, Mechanism and Data Analysis*, 2nd edn. (Wiley-VCH, New York, 2000). ISBN 978-0-4713-5929-6
5. W. Darge, *The Ribosome* (Ribosomen-Verlag, Krefeld, 2004). ISBN 978-3-9808-7530-1
6. B. Gutte, R.B. Merrifield, The total synthesis of an enzyme with ribonuclease A activity. J. Am. Chem. Soc. **91**(2), 501–502 (1969). doi: 10.1021/ja01030a050
7. K.A. Hasselbalch, Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. Biochem. Z. **78**, 112–144 (1916)
8. V. Henri, Theorie generale de l'action de quelques diastases. Compt. Rend. Hebd. Acad. Sci. Paris **135**, 916–919 (1902). URL http://gallica.bnf.fr/ark:/12148/bpt6k64607148/f1257.image
9. A.V. Hill, The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. J. Physiol. **40**, 190 (1910)
10. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3**(3), 318–356 (1961). doi: 10.1016/S0022-2836(61)80072-7. URL http://www.pasteur.fr/ip/resource/filecenter/document/01s-000046-03t/genetic-regulatory.pdf
11. D.E. Koshland, Enzyme flexibility and enzyme action. J. Cell. Comp. Physiol. **54**(Suppl. 1), 245 (1959). doi: 10.1002/jcp.1030540420
12. D.E. Koshland, G. Nemethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry **4**, 365–385 (1966). doi: 10.1021/bi00865a047. URL http://www.chem.umass.edu/~rmweis/Chem728/papers/P15_Koshland_et_al_Biochemistry_1966.pdf
13. S.A. Kuby, *A Study on Enzymes. Volume I: Enzyme Catalysis, Kinetics, and Substrate Binding* (CRC Press, Boca Raton, 1990). ISBN 978-0-8493-6987-2
14. J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model. J. Mol. Biol. **12**, 88–118 (1965). doi: 10.1016/S0022-2836(65)80285-6
15. J.M. Perkel, Visiting "noncodarnia". BioTechniques **54**(6), 301–304 (2013). doi: 10.2144/000114037
16. W. Rathmann, C. Scheidt-Nave, M. Roden, C. Herder, Type 2 diabetes: Prevalence and relevance of genetic and acquired factors for its prediction. Dtsch. Arztebl. Int. **110**(19), 331–337 (2013). doi: 10.3238/arztebl.2013.0331
17. I.H. Segel, *Enzyme Kinetics* (Wiley, New York, 1975), reprinted 1993. ISBN 978-0-4713-0309-1
18. S.P.L. SÃ¸rensen, Enzymstudien II. Über die Messung und Bedeutung der Wasserstoffionenkonzentration bei enzymatischen Prozessen. Biochem. Z. **21**, 131–304 (1909)
19. J. Sumner, The isolation and crystallization of the enzyme urease. J. Biol. Chem. **69**, 435–441 (1926). URL http://www.jbc.org/content/69/2/435.full.pdf+html
20. W.C. Winkler, A. Nahvi, A. Roth, J.A. Collins, R.H. Breaker, Control of gene expression by a natural metabolite-responsive ribozyme. Nature **428**, 281–286 (2004). doi: 10.1038/nature02362

Chapter 5
Enzyme Kinetics and Mechanism

Abstract Enzyme and substrate form an ES-complex, which reacts further to enzyme and product. This process requires time, so each enzyme molecule can only handle a certain number of substrate molecules per unit time, called the turnover-number k_{cat}. This number multiplied with the number of enzyme molecules is the limiting reaction velocity, V_{max}, reached only at infinite substrate concentration. At $[S] < \infty$ enzyme molecules need time to find a new substrate after the release of product, hence $v < V_{max}$.

Measuring the enzyme concentration in biological fluids is of great clinical significance.

Some enzymes are produced as inactive precursors and activated only when needed.

For some enzymes, the molecular mechanism of their action has been determined. If there are several substrates and/or products, substrates can bind to, and products be released from enzymes in specific order or randomly. Reaction mechanism is determined by keeping the concentration of one substrate constant, while varying the second. Different enzyme mechanisms result in characteristic LINEWEAVER-BURK-plots.

5.1 The HENRI–MICHAELIS–MENTEN Equation

To derive the relationship between the substrate concentration $[S]$ and the enzymatic velocity v, we start with the assumption of E. FISCHER [6] that enzymes act by binding their substrate in a special pocket, called the **binding site**, into which the substrate fits like a **key into a lock**. Within the enzyme the substrate is turned over into product and finally released:

$$\mathsf{E} + \mathsf{S} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathsf{ES} \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} \mathsf{EP} \underset{k_{-3}}{\overset{k_{+3}}{\rightleftharpoons}} \mathsf{E} + \mathsf{P}$$

The first to develop a mathematical model for this reaction was VICTOR HENRI [9]; his work was later extended by LEONOR MICHAELIS and his postdoc MAUDE LEONORA MENTEN [13]. The model is therefore called the HENRI–MICHAELIS–MENTEN equation (from now on we abbreviate this to HMM-equation).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_5

We start with the assumption that there is no product present, so that the back-reaction cannot occur. Under these conditions we get:

$$\mathsf{E} + \mathsf{S} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathsf{ES} \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} \mathsf{EP} \xrightarrow{k_{+3}} \mathsf{E} + \mathsf{P}$$

Because the conversion of ES to EP does not involve binding or release steps, its speed is independent of the concentration of reactants. Thus we can further simplify:

$$\mathsf{E} + \mathsf{S} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathsf{ES} \xrightarrow{k_{\text{cat}}} \mathsf{E} + \mathsf{P}$$

If we mix enzyme and substrate, the ES-complex will begin to form. Initially, the rate of formation of P will be zero, because there is no ES. As $[ES]$ increases, the rate of its breakdown into $\mathsf{E} + \mathsf{P}$ will also increase, until an equilibrium is established, where the rate of formation of ES from $\mathsf{E} + \mathsf{S}$ is equal to its rate of breakdown into $\mathsf{E} + \mathsf{P}$ (see Fig. 5.1). The HMM-law of enzyme kinetics, which we are about to derive, applies only in this **steady-state** phase of the reaction.

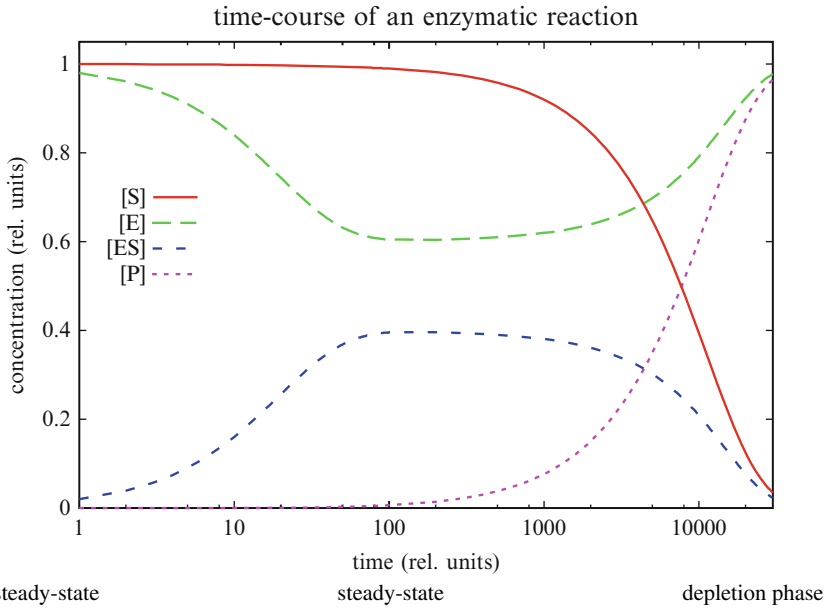


Fig. 5.1 Phases of an enzymatic reaction. In the first phase (pre-steady state) the rate of product formation is low because the ES-complex must be formed. In the second, steady-state, phase the rate of ES formation from $\mathsf{E} + \mathsf{S}$ is equal to the rate of its breakdown into $\mathsf{E} + \mathsf{P}$, $[ES]$ is constant, and $[P]$ increases linearly over time. In the last phase the substrate concentration becomes so low that ES-formation slows down, leading in turn to a decrease in the rate of product formation. Note the logarithmic time scale

The rate at which product is formed is, according to the law of mass action, given by

$$v = k_{\text{cat}} \times [ES] \tag{5.1}$$

In the absence of the second reaction step, $[ES]$ would depend on $[E]$, $[S]$ and $K_d = k_{-1}/k_{+1} = 1/K_a$:

$$[ES] = \frac{[E] \times [S]}{K_d} \tag{5.2}$$

where the **dissociation-constant** (K_d) is the concentration of substrate where $[ES]$ is half the total enzyme concentration $[E]_t$; in other words $[E] = [ES]$. Its unit is M. If $[S] = K_d$, then the fraction of enzyme with substrate bound $\Theta = K_d/(K_d + K_d) = 1/2$. K_d is the smaller, the higher the affinity of an enzyme is for its substrate. It is the reciprocal of the **association constant** K_a.

Strictly speaking, we would have to use the concentration of free substrate, $[S]$, for these calculations. This concentration we do not know, however; all we know is the total substrate concentration $[S]_t = [S] + [ES]$. However, if $[S]_t \gg [E]_t$, formation of the ES complex does not appreciably change the concentration of free substrate. In experiments, we can therefore use $[S]_t$ as an approximation for $[S]$. This assumption cannot be used for the enzyme, however, $[E]_t > [E]$ under usual conditions.

Because, however, in an enzymatic reaction ES can break down not only into E + S, but also into E + P, $[ES]$ must be lower than predicted by Eq. (5.2). G.E. BRIGGS and J.B.S. HALDANE [2] have shown that Eq. (5.2) can still be used, but we have to replace the dissociation constant $K_d = k_{-1}/k_{+1}$ with the MICHAELIS-constant $K_{\text{m}} = \frac{k_{-1}+k_{\text{cat}}}{k_{+1}}$. Although K_d is the concentration of substrate where half the enzyme molecules have substrate bound, K_{m} is the substrate concentration where the rate of the enzymatic reaction is half maximal. If we replace $[ES]$ in Eq. (5.1) with the modified Eq. (5.2) and rearrange, we get the HMM-equation:

$$v = \frac{k_{\text{cat}} \times [E]_t \times [S]}{K_{\text{m}} + [S]} = \frac{V_{\text{max}} \times [S]}{K_{\text{m}} + [S]} \tag{5.3}$$

This equation is graphed in Fig. 5.2. We note that

- if $[S] = \infty$, all enzymes will be present as ES ($[ES] = [E]_t$), and the rate of product formation will approach the limit $k_{\text{cat}} \times [E]_t = V_{\text{max}}$. This is commonly called "limiting velocity" (in the older literature "maximal velocity", but this term is now discouraged by IUBMB as it is not a maximum in the mathematical sense).
- if $[S] \gg K_{\text{m}}$, relative large changes in $[S]$ will have only a small effect on v; in other words, the reaction will be of zeroth order with respect to S. Example: If $[S] = 10 \times K_{\text{m}}$, then $v = V_{\text{max}} \times \frac{10K_d}{11K_d} = V_{\text{max}} \times 0.91$. If on the other hand

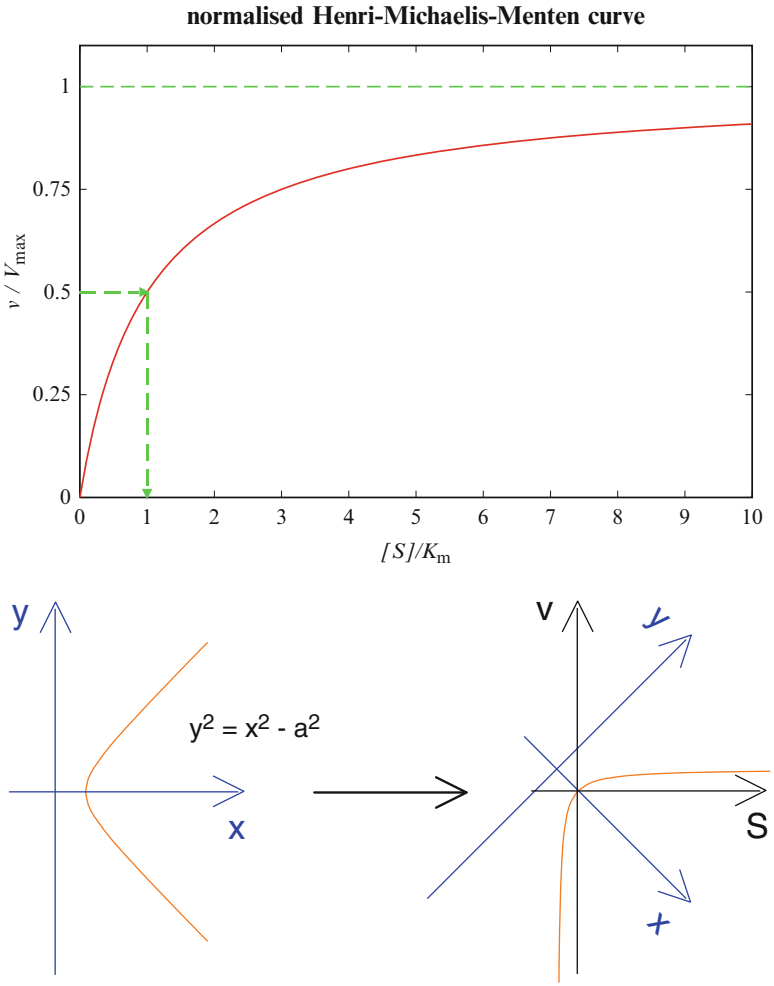


Fig. 5.2 *Top*: Velocity of an enzyme reaction as function of the substrate concentration $[S]$. The curve is a hyperbola. K_m can be obtained as the substrate concentration at which $v = 1/2\ V_{max}$. *Bottom*: The graph of the HMM-equation is obtained from a rectangular hyperbola by a 45° rotation of the coordinate system, followed by a translation to bring the vertex of the hyperbola to the origin

$[S] = 20 \times K_d$, then $v = V_{max} \times 0.95$. In other words, a doubling of the substrate concentration resulted in a 4.4 % increase in v.

- if $[S] = K_m$, then $v = \frac{V_{max} \times K_m}{2K_m} = V_{max}/2$
- if $[S] \ll K_m$, then the relationship between $[S]$ and v can be approximated by a straight line $v \approx [E] \times [S]$.

The unit of the velocity of an enzyme reaction is the **katal**, 1 kat = 1 mol/s. In the older literature you find the **enzyme unit**, 1 U = 1 µmol/min = 16.7 nkat.

We can also calculate Θ, the fraction of enzyme molecules that has substrate bound to it:

$$\Theta = \frac{[ES]}{([ES] + [E])} = \frac{K_a \times [E] \times [S]}{(K_a \times [E] \times [S] + [E])} \tag{5.4}$$

$$= \frac{K_a \times [S]}{(K_a \times [S] + 1)} = \frac{[S]}{([S] + \frac{1}{K_a})} = \frac{[S]}{([S] + K_d)} \tag{5.5}$$

$$v = V_{\max} \times \Theta \tag{5.6}$$

For our derivation of the HMM-equation we have made some simplifying assumptions:

- We have assumed that the reaction rate is determined by a single rate-limiting step, and got $V_{\max} = k_{\text{cat}} \times [E]_t$. This condition is not necessarily true, but even if it is false, the HMM-equation is still applicable. However, we can no longer interpret k_{cat} as a single first-order rate constant. Instead, turnover number of an enzyme may be a complicated function of several partially rate-limiting reaction constants.
- We have assumed that $k_{+1} \gg k_{+2}$. This too need not be true; the HMM-equation still holds even when $k_{+1} \ll k_{+2}$, but the interpretation becomes different. In the general case, we have to replace K_d with the MICHAELIS-constant $K_{\text{m}} = \frac{k_{+2}+k_{-1}}{k_{+1}}$ which is not inversely proportional to the affinity of the enzyme for its substrate. K_{m} is the substrate concentration where $v = V_{\max}/2$, K_d is the substrate concentration where $\Theta = 1/2$.

The turnover number k_{cat} can assume very different values for different enzymes. Some enzymes, such as Hsc70 or RecA, turn over less than 1 substrate molecule per second, whereas a molecule of catalase (which detoxifies H_2O_2 in our cells) can turn over 40×10^6 molecules per second.

Derivation of the HMM-Equation in Its General Form

In the previous section we have derived the HMM-equation for the reaction

$$E + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} ES \xrightarrow{k_{\text{cat}}} E + P$$

from equilibrium considerations by assuming that the establishment of the binding equilibrium is much faster than the conversion of the ES complex to E and P. This was how HENRI derived the HMM-equation [9], but for some enzymes the assumption of $k_{+1} \gg k_{\text{cat}}$ does not hold. BRIGGS & HALDANE [2] derived the HMM equation for the general case in 1925.

The enzymatic reaction can be described by a system of differential equations:

(continued)

$$\begin{cases} \dfrac{d[S]}{dt} = k_{-1}[ES] - k_{+1}[S][E] \\ \dfrac{d[E]}{dt} = (k_{-1} + k_{\text{cat}})[ES] - k_{+1}[S][E] \\ \dfrac{d[ES]}{dt} = k_{+1}[E][S] - (k_{-1} + k_{\text{cat}})[ES] \\ v = \dfrac{d[P]}{dt} = k_{\text{cat}}[ES] \end{cases} \tag{5.7}$$

Such differential equations describe the rate of change of a parameter over time. The integrated form would describe the parameter over time. In many processes it is easy to write the differential, but very difficult to get the integrated form.

A considerable simplification of this system of differential equations may be achieved if we limit ourselves to the steady-state phase of the reaction, where ES is formed as rapidly as it is broken down; that is, $\frac{d[ES]}{dt} = \frac{d[E]}{dt} = 0$ (see Fig. 5.1). This is known as the **steady-state-assumption**. Note that such an equilibrium phase exists independent of the ratio of k_{+1}/k_{cat}. Then the above equations simplify to

$$k_{+1}[S][E] = (k_{-1} + k_{\text{cat}}) \times [ES] \tag{5.8}$$

Because of mass conservation $[E]_t = [E] + [ES] = \text{const}$ (the total enzyme concentration is the sum of free enzyme and ES-complex concentrations and is constant over time) we can also write Eq. (5.2) as:

$$k_{+1}[S]([E]_t - [ES]) = (k_{-1} + k_{\text{cat}})[ES] \tag{5.9}$$

$$k_{+1}[S][E]_t - k_{+1}[S][ES] = (k_{-1} + k_{\text{cat}})[ES] \tag{5.10}$$

$$k_{+1}[S][E]_t = (k_{+1} + k_{-1} + k_{\text{cat}})[ES] \tag{5.11}$$

$$[ES] = \frac{k_{+1}[E]_t[S]}{k_{+1}[S] + k_{-1} + k_{\text{cat}}} \tag{5.12}$$

$$= \frac{[E]_t[S]}{\frac{k_{-1}+k_{\text{cat}}}{k_{+1}} + [S]} \tag{5.13}$$

$$= \frac{[E]_t[S]}{K_{\text{m}} + [S]} \tag{5.14}$$

(continued)

If we replace $[ES]$ in the system of differential equations with Eq. (5.13), and consider that the maximal velocity is achieved when all the enzyme is converted to the enzyme-substrate complex ($[ES] = [E]_t$), we get

$$v = \frac{k_{\text{cat}}[E]_t[S]}{K_{\text{m}} + [S]} = \frac{V_{\max}[S]}{K_{\text{m}} + [S]} \tag{5.15}$$

Thus the rapid equilibrium approach of HENRI and the steady-state approach of BRIGGS & HALDANE yield the same result, except that K_d is replaced by K_{m}.

5.1.1 *Efficiency Constant and Catalytic Perfection*

The HMM-equation may be written as

$$v = \frac{V_{\max}[S]}{K_{\text{m}} + [S]} = \frac{k_{\text{cat}}}{K_{\text{m}}}[E][S] \tag{5.16}$$

The term $k_{\text{cat}}/K_{\text{m}}$ is called the **efficiency constant** of the enzyme; its unit is that of a second-order reaction constant ($\text{M}^{-1}\text{s}^{-1}$).

Enzymes, like all catalysts, do not change the equilibrium of a reaction; they merely increase the speed at which it is obtained. Under equilibrium conditions the forward and reverse reactions proceed with equal speed, hence

$$\left(\frac{k_{\text{cat}}}{K_{\text{m}}}\right)_f [E][S] = \left(\frac{k_{\text{cat}}}{K_{\text{m}}}\right)_r [E][P] \tag{5.17}$$

$$K_{eq} = \frac{(k_{\text{cat}}/K_{\text{m}})_f}{(k_{\text{cat}}/K_{\text{m}})_r} \tag{5.18}$$

which is known as the HALDANE-**relationship**.

Note that at very low substrate concentration ($[S] \ll K_{\text{m}}$) the HMM-hyperbola can be approximated by a straight line, with the equation

$$v = \frac{k_{\text{cat}}}{K_{\text{m}}} \times E_t \times [S] \tag{5.19}$$

The rate-limiting step in an enzyme's reaction can be either the association of enzyme with its substrate, or the conversion to and release of product. The association velocity is described by the association rate constant k_{+1}, which is on

the order of $10^9\,M^{-1}s^{-1}$. This is the rate at which substrate can diffuse and bind to the enzyme in a dilute aqueous medium at physiological temperatures (Table 5.1).

Table 5.1 Efficiency constants of some almost perfect enzymes. Data were obtained from BRENDA

Enzyme	Organism	Substrate	k_{cat} s^{-1}	K_m M	k_{cat}/K_m $M^{-1}\,s^{-1}$
β-lactamase	*E. coli*	Ampicillin	1090	8.0×10^{-5}	8.7×10^8
Carbonic anhydrase	*Mus musculus*	CO_2	940 000	1.6×10^{-3}	5.9×10^8
Catalase	*N. crassa*	H_2O_2	125 000	2.5×10^{-4}	5.0×10^8
AcChE	*H. sapiens*	Ac-S-choline	6500	4.6×10^{-5}	1.4×10^8
Peroxydase	*Strep. faecalis*	NADH	83.3	2.0×10^{-6}	4.2×10^7
Fumarase	*S. scrofa*	Fumarate	364	5.0×10^{-6}	7.3×10^7
Triose-P isomerase	*S. cerevisiae*	d-GA3P	16 700	1.1×10^{-3}	5.6×10^7

If the efficiency constant is of the same order of magnitude (note that it has the same unit!) the enzyme is called **catalytically perfect**, because the total reaction velocity is limited only by diffusion of the substrate to the enzyme (**diffusion controlled reaction**). The enzyme can turn over the substrate as fast as it is delivered [1].

The limitation of the efficiency constant by diffusion is overcome by **multienzyme complexes**, where the product of one enzyme is fed directly to the next enzyme as substrate, without time-consuming diffusion steps in between. This not only increases the overall speed of a biochemical pathway, it also may prevent the destruction of an unstable intermediate by some side reaction.

Elucidating the **interactom** of cells has become an important research question. This association of enzymes into complexes can be observed already in bacteria. In eukaryotes sometimes the bacterial operons fuse into a single gene which encodes a multifunctional protein.

What Determines Association Velocity?

According to EINSTEIN:

$$D = \mu_p kT \qquad \mu_p = \frac{V_d}{F} \tag{5.20}$$

with D the diffusion coefficient, k the BOLTZMANN constant, T the absolute temperature, V_d the terminal drift velocity, and F the force applied. According to STOKES

$$\mu = \frac{1}{6\pi\nu r} \tag{5.21}$$

(continued)

with ν the viscosity of the medium and r the radius of the protein. This gives the EINSTEIN-STOKES-relation

$$D = \frac{kT}{6\pi \nu r} \tag{5.22}$$

The association constant for small molecules $\mathsf{A} + \mathsf{B} \rightleftharpoons \mathsf{AB}$ can be described in terms of the SMOLUCHOWSKI-equation

$$k_{+1} = \frac{4\pi N_a}{1000}(D_a + D_b) \tag{5.23}$$

which is on the order of $7 \times 10^9\ \mathrm{M^{-1}\ s^{-1}}$. However, this assumes the molecules to be small billiard balls. Binding does not require specific orientation and the molecules are uncharged, so that there are no electrostatic steering forces.

For the interaction of a protein with a small ligand a modified SMOLUCHOWSKI-equation may be used, which takes these effects into account:

$$k_{+1} = \frac{4\pi \kappa a f N_a}{1000}(D_a + D_b) = Zpfe^{-\frac{E_a}{RT}} \tag{5.24}$$

where κ is the fraction of the surface involved in binding and f the electrostatic steering factor. In the second part of the equation Z is the collision factor, p the fraction of collision with correct orientation, R the universal gas constant, and E_a the activation energy. Then the association rate constant k_{+1} is on the order of 10^9–$10^{11}\ \mathrm{M^{-1}\ s^{-1}}$.

5.1.2 *Application: Forensic Determination of Blood Alcohol Concentration*

In 7 % of all road accidents, 10 % of the accidents that cause injuries and 31 % of all fatal accidents in the US at least one person involved (not necessarily, but often the guilty one) was legally drunk, according to NHTSA statistics. Because alcohol has a very detrimental effect on the ability of motorists to control their vehicle, most countries have legal limits on [alcohol] in the blood of motorists. Depending on jurisdiction, these tend to vary between 0.2–0.8 ‰, but some countries enforce a strict 0 ‰ rule. In an "untrained" drinker 0.3 ‰ leads to marked changes in behaviour, 1.5 ‰ to unconsciousness, and 4 ‰ to death by respiratory suppression (Fig. 5.3).

Ethanol is oxydised in our livers by alcohol dehydrogenase, which has a K_m of 1 mM ≡ 0.046 ‰. In other words, at legally relevant blood alcohol concentrations the enzyme works near substrate saturation; the reaction is of zeroth order with

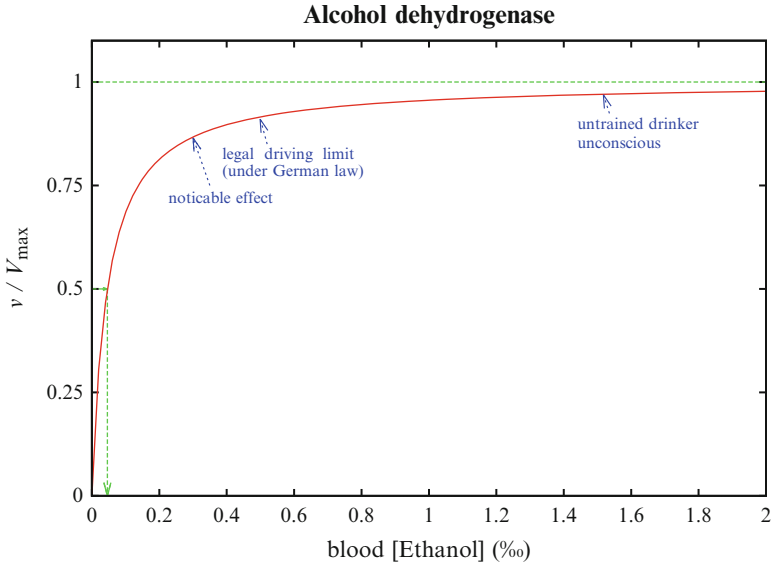


Fig. 5.3 Alcohol dehydrogenase's K_m for ethanol is so low that the enzyme works zeroth order at physiologically and legally relevant blood alcohol concentrations. This makes evolutionary sense: any ingested ethanol (*e.g.*, with ripe fruit) is rapidly detoxified. Consumption of alcoholic beverages (especially distilled spirits) overwhelms this mechanism

$v \approx 0.15..0.2\,‰/\,\mathrm{h}$. There is usually a considerable delay between an arrest for drunk driving and the taking of a blood sample, but the above relationship allows the back-calculation of the alcohol concentration at the time of arrest or the time of an accident (which is the legally relevant one).

5.1.3 Linearisation of the HMM-Equation

If you have experimentally determined the reaction rate v of an enzyme as a function of the substrate concentration $[S]$, it is not always easy to determine V_{max}, because the curve at high substrate concentrations is so flat. This is especially true if your data points contain experimental error (which they always do!).

5.1.3.1 The LINEWEAVER-BURK-Transformation

BARNETT WOOLF had the idea to linearise the HMM-equation by using either one of three mathematical transformations. However, due to injuries received in a road accident he was unable to publish his results [7]. He had shown them to his friend HALDANE, who referred to them in his seminal book on enzyme kinetics [8].

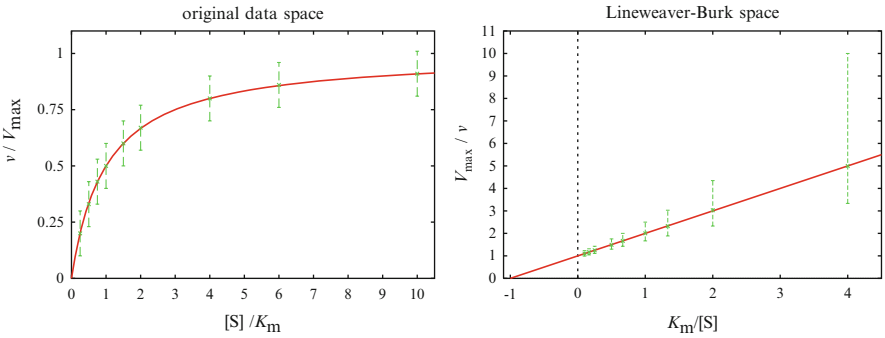


Fig. 5.4 Transformation of data into LINEWEAVER-BURK-coordinates has a significant effect on the distribution of errors. Linear regression assumes that the error is independent of the controlled variable, and hence is not applicable to LINEWEAVER-BURK-plots

However, at that time those methods had little impact; they were later independently developed by the authors whose name they bear today. The most important one is the LINEWEAVER-BURK-transformation, which consists of taking reciprocals:

$$v = \frac{V_{\max} \times [S]}{K_m + [S]} \qquad \Rightarrow \qquad \frac{1}{v} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \times \frac{1}{[S]} \tag{5.25}$$

This is the equation of a straight line ($y = a + bx$), the y-intercept is $1/V_{\max}$, the slope is $K_m/V_{\max}$ and, as can be shown easily by setting $1/v$ to 0, the x-intercept is $-1/K_m$.

Using the LINEWEAVER-BURK-transformation is not without problems: the taking of reciprocals transforms not only the data points, but also their standard deviation. This becomes not only unsymmetrical, but extremely large for low values of $[S]$ (see Fig. 5.4).

The LINEWEAVER-BURK-transformation should therefore not be used to estimate K_m and $V_{\max}$ from experimental data by linear regression, which assumes the error of data points to be independent of the x-value. We have better procedures to do that by curve-fitting directly to the original data on a computer (for experts: NELDER/MEAD-simplex and MARQUARDT/ LEVENBERG algorithms [12, 14]). However, as we will see later, the LINEWEAVER-BURK-transformation is still very useful for presenting the results of kinetic experiments.

Not all enzymes follow HMM-kinetics (see the discussion of cooperativity on page 165ff), and in such cases the data do not form a straight line in a LINEWEAVER-BURK-plot. Thus linear transforms are an easy way to judge whether the correct model has been chosen.

Apart from the LINEWEAVER-BURK-transformation there are two other transformations which achieve the same purpose. These are:

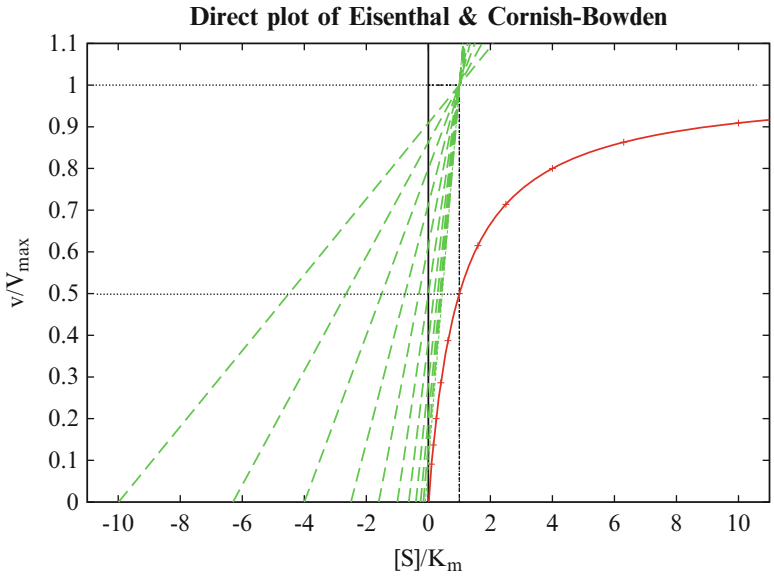


Fig. 5.5 The direct plot (introduced by EISENTHAL & CORNISH-BOWDEN) is an alternative way to plot enzyme kinetics data

$$\frac{[S]}{v} = \frac{[S]}{V_{\max}} + \frac{K_{\mathrm{m}}}{V_{\max}} \tag{5.26}$$

$$v = V_{\max} - K_{\mathrm{m}} * \frac{v}{[S]} \tag{5.27}$$

These linearisations, known as HANES- and as EADIE-HOFSTEE-plots respectively, lead not only to a change in error distribution (although somewhat less than the LINEWEAVER-BURK-transform), but also do not separate dependent and independent variables onto different axes. For this reason they are rarely used; medical students may safely ignore them.

Because of the close relationship between ligand binding and enzyme activity all three linearisation methods may also be used in binding experiments for plots of $[ES]$ versus $[S]$, however, in this case they have different names. The plot of $[ES]$ vs $[ES]/[S]$ is most commonly used and known as SCATCHARD-plot (related to the HANES-plot). The y-intercept gives the maximum binding capacity; the slope is K_d.

5.1.3.2 The "direct plot" of EISENTHAL & CORNISH-BOWDEN

If the negative substrate concentrations on the x-axis are connected by straight lines to the corresponding measured velocities on the y-axis, all these lines should meet in a single point, with the coordinates $K_{\mathrm{m}}/V_{\max}$ (see Fig. 5.5). The data are not

transformed, therefore the associated statistical problems do not occur. With K_m and V_{max} known, the HMM-curve through the data can be plotted [5].

With actual experimental data, however, the lines do not intersect in a single point, but in a number of closely-spaced points. The median of their coordinates can be used to determine K_m and V_{max}; the spread (minimal and maximal values) gives a good indication of the error margins.

5.1.4 Experimental Pitfalls

5.1.4.1 Initial Velocity

We have deduced the HMM equation under the condition that no product is present and that therefore the speed of the backreaction is zero. Of course this will be the case only the moment we start the reaction. If we let the reaction proceed, the substrate concentration will decrease, and the product concentration increase.[1] Both effects will lower the velocity of the forward reaction, and the product will not linearly increase with time. Rather, a hyperbolic curve will be seen if $[P]$ is measured as a function of time (see Fig. 5.6). This curve can be approximated by a straight line at the beginning. The slope of this line is called the **initial velocity** of the reaction; kinetic results are valid only if measurements are taken under these conditions.

The Use of Average Concentrations

Alternatively, if the reaction is irreversible and product inhibition does not occur (conditions that can be forced to be true by removing the product in a second reaction), one can use the **average substrate concentration** $[\overline{S}] = ([S]_0 + [S]_t)/2$ over the measurement time t as an approximation for the "true" substrate concentration $[S]$ [11]. This is especially useful if the decrease in $[S]$, rather than the increase in $[P]$ is followed, because of the statistical problems involved in measuring small differences.

[1] HENRI's version of the HMM equation tries to take care of the increasing product concentration by treating the product as a competitive inhibitor of the reaction:

$$v = \frac{V_{max}([S]_0 - [P])}{1 + \frac{[S]_0 - [P]}{K_s} + \frac{[P]}{K_p}} \tag{5.28}$$

with K_s and K_p the dissociation constants for substrate and product, respectively. $[S]_0$ is the substrate concentration at the start of the experiment.

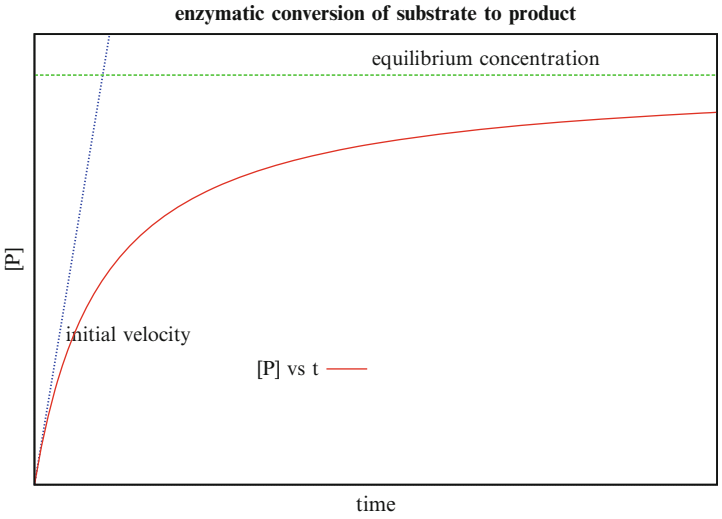


Fig. 5.6 Increase of $[P]$ over time in an enzymatic reaction. The reaction velocity ($\frac{d[P]}{dt}$) is highest at the $t = 0$; it approaches zero as the ratio of $[P]/[S]$ approaches the equilibrium. The reaction velocity at $t = 0$ is called the "initial velocity"

Substrate-Regenerating Systems

If enzyme reactions are to be followed over longer time periods, the substrate can sometimes be kept constant by the use of a **regenerating system**. For example, if the reaction velocity of an ATPase is to be measured, [ATP] can be kept constant by adding phospho*enol*pyruvate (PEP) and pyruvate kinase (PK) to the system (ADP + PEP $\xrightarrow{\text{PK}}$ ATP + pyruvate). Thus any ADP produced would be converted back to ATP. Of course in this case the speed of Pi production would have to be measured, as overall no ADP is produced nor ATP destroyed. A second option would be to measure the produced pyruvate.

5.1.4.2 Measurement Error

As mentioned, not all enzymes follow HMM-kinetics. It is thus important to experimentally distinguish those that do from those that don't. This is made more difficult by the fact that all experimental data are subject to measurement error. For this reason enough data points must be collected over a large enough concentration range to decide whether the kinetics of an enzyme follows the HMM-equation and, if so, reliably calculate V_{max} and K_m. As a rule of thumb at least 10–12 data points should be collected, and they should evenly cover the range of 0.1–5 times K_m, resulting in reaction velocities between 9 and 83 % of V_{max}. It is also important to

calculate error estimates for V_{max} and K_m. (Make this a general rule: results without error estimates are almost useless.)

The Integral of the HMM-Equation

The HMM-equation, which relates the substrate concentration with the resulting reaction velocity ($d[P]/dt$ or $-d[S]/dt$) can be integrated, yielding the HENRI-equation:

$$v = -\frac{d[S]}{dt} = \frac{V_{max} * [S]}{K_m + [S]} \tag{5.29}$$

$$V_{max}dt = -\frac{K_m + [S]}{[S]}d[S] \tag{5.30}$$

$$V_{max}\int_{t_0}^{t} dt = -\int_{[S]_0}^{[S]} \frac{K_m + [S]}{[S]}d[S] \tag{5.31}$$

$$= -K_m \int_{[S]_0}^{[S]} \frac{d[S]}{[S]} - \int_{[S]_0}^{[S]} d[S] \tag{5.32}$$

$$V_{max}t = -K_m \ln\left(\frac{[S]}{[S]_0}\right) - ([S] - [S]_0) \tag{5.33}$$

$$= 2.303K_m \log_{10}\left(\frac{[S]}{[S]_0}\right) + ([S] - [S]_0) \tag{5.34}$$

$$= 2.303K_m \log_{10}\left(\frac{[S]}{[S]_0}\right) + [P] \tag{5.35}$$

$$-\frac{1}{K_m}\frac{[P]}{t} + \frac{K_m}{V_{max}} = \frac{2.303}{t}\log_{10}\left(\frac{[S]_0}{[S]}\right) \tag{5.36}$$

$$= y \tag{5.37}$$

Thus a plot of y as a function of $[P]/t$ yields a straight line with y-intercept V_{max}/K_m, slope $-1/K_m$ and x-intercept V_{max} [4, 16]. Several data points should be obtained, starting at a concentration significantly higher than K_m and letting the reaction proceed until $[S] \ll K_m$. In actual practice the HENRI-equation is rarely used, as results tend to be imprecise.

5.1.4.3 Stability of Enzyme

Enzymes may be unstable, especially in dilute solutions. Thus if an enzyme stock solution is prepared in the morning, it may give different results from those obtained with the same solution in the evening after several hours of storage.

To minimise such problems, enzyme stock solutions are prepared fresh each day, they are stored in an ice bath during the experiment, and it is good practice to establish enzyme stability in control experiments. If very dilute solutions are required, inert carrier proteins may be added. Gelatin is very effective; it also does not contain any aromatic amino acids, which could absorb UV-light in photometric assays.

5.1.4.4 Control Experiments Required

Enzyme kinetic experiments should *always* include controls for the turnover of substrate without enzyme (all reagents except enzyme solution, which is replaced by buffer) and for turnover of other compounds that might be present (all reagents except substrate). Otherwise serious misinterpretation of experimental results is possible.

For example, if the breakdown of *para*-nitrophenylphosphate (pNPP) by alkaline phosphatase is studied as a function of *p*H, it might appear that the phosphatase becomes more active at acidic *p*H. This increased breakdown of pNPP, however, is also seen without the enzyme; it is caused by H^+-ions acting as catalysts.

5.1.5 Environmental Influences on Enzyme Activity

You have already seen that protein structure depends on environmental factors such as temperature, *p*H, ionic strength, or the presence of denaturing compounds. Of course, enzyme function depends on proper folding, so any of these factors will also affect enzymatic activity. In addition, the following points are relevant:

temperature Higher temperatures mean more molecular motion and hence higher reaction rates. However, beyond a certain temperature the enzyme gets inactivated, hence enzymatic activity versus temperature is an optimum curve.

Osmolarity Enzymes need water as a "grease" for conformational changes. Osmolytes reduce the available water concentration and hence the enzymatic activity.

***p*H** The protonation state of both the substrate and of amino acid R-groups in the catalytic center of the enzyme are influenced by *p*H.

5.2 Enzymes with Several Substrates

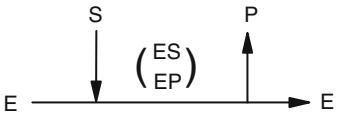
We have looked at the kinetics of enzymes with a single substrate and a single product. However, many enzymes in our bodies have several substrates and/or products. Although a detailed kinetic treatment of these enzymes is beyond the scope of this course, we should at least look briefly at the mechanism of such enzymes. Again we use the nomenclature and formalism suggested by W.W. CLELAND [3].

5.2.1 Nomenclature

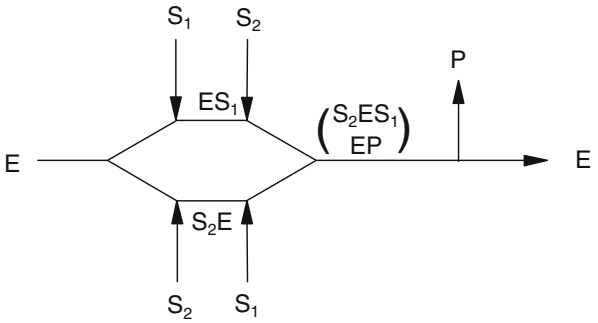
To graphically represent the reaction mechanism of enzymes, CLELAND suggested a special type of diagram, where the reaction is represented by a long horizontal arrow. Vertical arrows denote entering substrates or leaving products. A conversion inside the enzyme, without entering or leaving substances is denoted by a pair of brackets and called an **inner complex**. Thus our single substrate, single product enzyme reaction:

$$E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P \tag{5.38}$$

would look like this in a CLELAND-diagram:

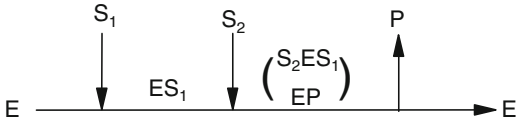


For enzymes with two substrates it may be irrelevant whether S_1 binds before or after S_2. Such a mechanism is called **random-bi**, and the CLELAND-diagram would look like this:



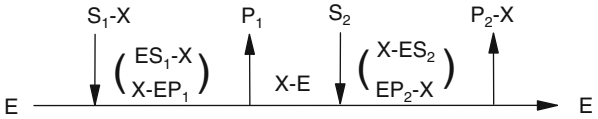
If the enzyme has three substrates rather than two, it would be called **random-ter**.

On the other hand, both substrates may have to bind in a predetermined order; S_2 cannot bind to the enzyme unless S_1 has already bound. Such a mechanism would be called **ordered bi**:

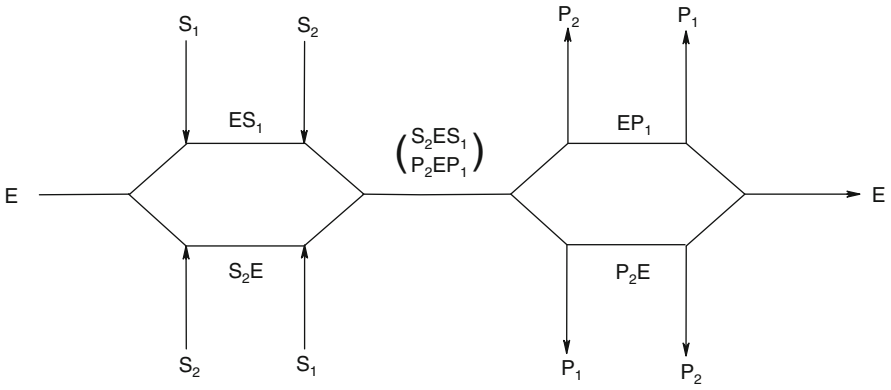


A third possibility, realised with many transferases, is that the first substrate binds to the enzyme, transfers a functional group to the enzyme, and is released as the first

product. Only then can the second substrate bind and accept this functional group, to be released as the second product. This mechanism is called **ping-pong**:



Of course, enzymes may have more than one product, and those may be released in random or sequential order. CLELAND-diagrams are constructed the same way as for substrates, and the nomenclature is identical too. For example, a random-bi-bi mechanism would refer to an enzyme with two substrates and two products, which can be bound and released in any order.



5.2.2 *How Do We Determine the Mechanism of Multisubstrate Enzymes?*

The experimental procedure is to measure the reaction velocity as the function of the concentration of one substrate, while the concentration of the second is held constant. This experiment is repeated with different concentrations of the second substrate, thus a number of HMM-curves are obtained with different parameters depending on $[S_2]$. In the LINEWEAVER-BURK-transformation the different reaction mechanisms result in characteristic patterns; secondary plots are then used to determine the K_m-values for the substrates.

Enzymes with ping-pong mechanism give LINEWEAVER-BURK-plots with parallel lines. If the lines intersect, a ternary complex is formed (random-bi or ordered-bi).

Similar approaches can be taken with multisubstrate enzymes, but the work involved increases geometrically with the number of substrates.

5.3 Enzyme Precursors and Their Activation

Some cells produce and secrete enzymes which would damage the cell if those enzymes were already active inside the cell. A typical example are the digestive proteases such as pepsin and trypsin. These are produced in an inactive form, secreted, and then activated.

Such enzymes are produced as pro-enzymes consisting of two domains, the enzymatic domain and a regulatory one, which also may carry a marker that is recognised by the protein export machine in the cell, and which may act as a chaperone during protein folding. This domain is cleaved off during or after secretion.

In the case of the digestive proteases the cleavage process is autocatalytic; that is, pepsin molecules in the stomach juice (or trypsin molecules in pancreas juice) cleave off the regulatory sequences of newly secreted pro-enzymes. Trypsin then activates all other digestive hydrolases secreted by the pancreas.

In chymotrypsin the structural changes during activation have been elucidated by X-ray crystallography. Cleavage of the N-terminal 15 amino acid pro-sequence allows the newly freed amino group of Ile-16 to form a salt bridge with Asp-194 (see Fig. 5.7), the resulting conformational changes activate the enzyme. Activation can be reversed by increasing the *p*H and thus abolishing salt bridge formation;

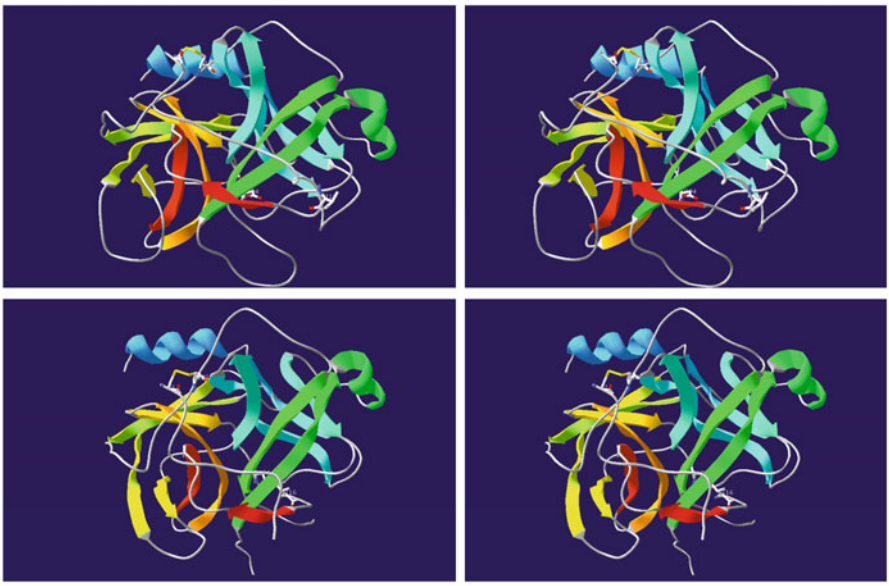


Fig. 5.7 Activation of α-chymotrypsinogen (top, PDB-code 1CGI) to α-chymotrypsin (bottom, PDB-code 1OXG) by proteolytic cleavage at Ile-16. The free amino-group of Ile-16 can form a salt bridge to Asp-194, leading to substantial conformational changes. The pro-peptide stays bound to chymotrypsin because of a disulphide bond between Cys-1 and -122

the enzyme then returns to its inactive conformation. Somewhat unusually the propeptide and the enzyme stay together after activation, because of a disulphide bridge between Cys-1 and -122.

Use of enzymes for diagnostics
Many enzymes occur only in certain specialised cells in our body. If those cells are damaged, enzyme molecules are released into the circulation. Thus determination of enzyme activities is an important diagnostic tool in the clinical laboratory. For example, CK is prevalent in muscle cells. Destruction of heart muscle cells in an acute myocardial infarct or of skeletal muscle cells in DUCHENNE muscular dystrophy leads to the appearance of CK in the blood. Interestingly, different organs contain different isoforms of CK, so that it is possible to tell from which organ an increased blood CK activity originated.

5.4 The Coupled Spectrophotometric Assay of WARBURG

The concentration of substrates and enzymes in clinical samples is often determined using the coupled spectrophotometric assay which was introduced by O. WARBURG [17].

This test is based on the observation that NADH and NADPH absorb UV-light at 345 nm, whereas NAD^+ and NADP do not (see Fig. 5.8). These coenzymes are used by many dehydrogenases. Thus the activity of dehydrogenases can be determined by the rate at which the absorbtion of UV-light of a sample changes. On the other hand, the concentration of substrates of such dehydrogenases can be determined by the total change of absorbtion.

There are many enzymes that do not themselves use NAD^+ or NADP, but which can be coupled to an enzyme that does. The creatine kinase discussed above is an example:

$$\text{Creatine} + \mathit{ATP} \xrightleftharpoons[\text{kinase}]{\text{Creatine}} \text{Creatinephosphate} + \mathit{ADP} \tag{5.39}$$

$$\mathit{ADP} + \text{Phospho}\mathit{enol}\text{pyruvate} \xrightleftharpoons[\text{kinase}]{\text{Pyruvate}} \text{Pyruvate} + \mathit{ATP} \tag{5.40}$$

$$\text{Pyruvate} + \mathit{NADH} \xrightleftharpoons[\text{dehydrogenase}]{\text{Lactate}} \text{Lactate} + \text{NAD}^+ \tag{5.41}$$

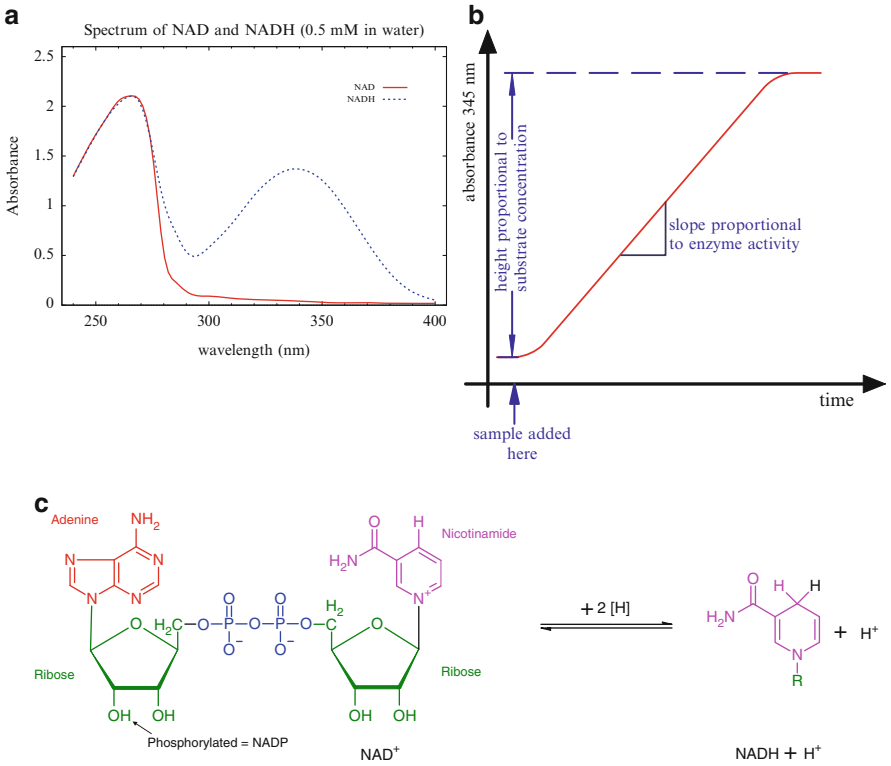


Fig. 5.8 **(a)** UV-spectra of NAD and NADH in water. Both have an absorbance maximum at 260 nm, attributable to the adenine ring. NAD(P)H has an additional absorbance maximum at 340 nm, which can be used to distinguish it from $NAD(P)^+$. **(b)** Change of UV-absorbance over time during a coupled spectrophotometric test. Substrate and enzyme concentrations can be determined. **(c)** Reduction of NAD^+ to NADH does not affect the structure of the adenine ring, thus its absorbance at 260 nm is unchanged. However, the nicotinamide ring does change, resulting in an additional peak at 340 nm

Clinical laboratories do this type of assay fully automated and with reagent kits bought premixed from suitable manufacturers.

5.5 How Do Enzymes Work?

We have extensively discussed *what* enzymes do; now we take a look at *how* they do it (Fig. 5.9).

An enzyme will orient the reactants properly. For this reason if the reaction of a nonchiral substrate leads to a chiral product only one of the possible isomers will be produced (e.g., aconitase in the citric acid cycle).

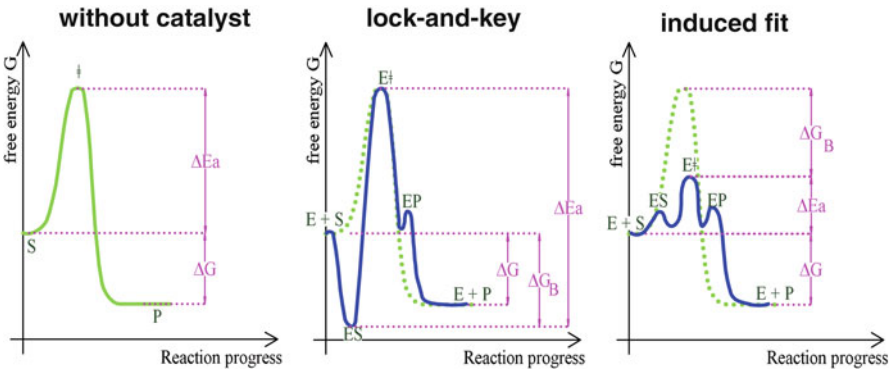


Fig. 5.9 *Left*: A chemical reaction without a catalyst. The product has a lower free energy (G) than the substrate, therefore the reaction should proceed. However, because the substrate S first needs to be activated to a transition state ‡ with high energy, the reaction rate is slow. Catalysts accelerate the reaction rate by lowering the activation energy (ΔE_a). Because they do not influence ΔG, they also have no influence on the equilibrium position. *Middle*: If the substrate binding site in the enzyme were complementary to the substrate (**lock-and-key model**), enzymes would stabilise the substrate and increase the activation energy for the reaction. The reaction rate would be lowered. *Right*: If the binding site is complementary to the transition state (**induced fit model**) a small amount of activation energy is required to form the initial ES complex, but the binding energy (ΔE_b) between enzyme and transition state mostly compensates for the activation energy required for the formation of the transition state. The change in free energy resulting from the conversion of the transition state into product then drives the formation of the EP complex, breaking the E−‡ bonds in the process. A further small amount of activation energy is then required to release the product from the enzyme

Substrates are held in the optimal position for much longer than they would be during random collisions. Effectively, the concentration of substrates in the catalytic centre of an enzyme is considerably higher than that in solution, and second-order reactions between molecules are converted to first-order reactions inside a molecule.

Acidic and/or basic side-chains of the enzyme may participate in **acid/base catalysis**; their $p\mathrm{K_a}$-values are adjusted by the environment to the needs of the reaction. For example, in the xylanase from *Bacillus circulans* Glu-172 acts as a proton donor and has a $p\mathrm{K_a}$ of 6.7, whereas Glu-78 with $p\mathrm{K_a}$ = 4.6 is ionised and stabilises the positively charged intermediate. Calculation of $p\mathrm{K_a}$-values in proteins (rather than in water) is possible, if the 3D-structure has been solved [15].

Redox reactions may be catalysed by metal centres in proteins; their redox potential again can be adjusted to the needs of the reaction by the protein environment. For example, free Fe^{3+}/Fe^{2+} has a standard potential of 771 mV; for protein-bound iron the potential can reach from 365 mV (cytochrome f) down to −432 mV (ferredoxin).

Electrostatic fields between a charged group in the transition state and a group of opposite charge in the enzyme may stabilise the complex.

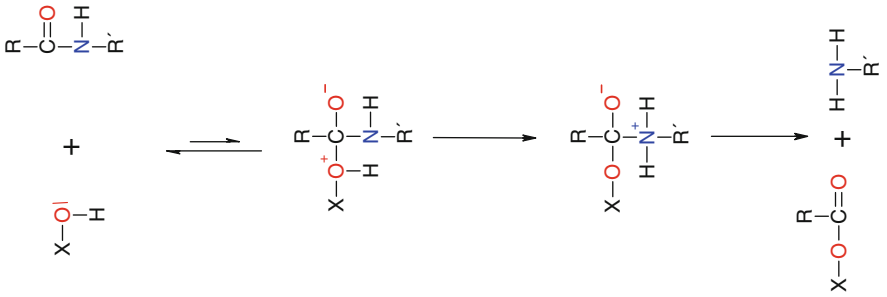
The binding site may be hydrophobic or hydrophilic, depending on the needs of the reaction.

In order to form the strong bond between the transition state of the substrate and the enzyme, bonds in the substrate may be strained and become susceptible to break (**rack mechanism**).

The induced-fit hypothesis of enzyme mechanism [10] has a practical application: if an antibody is raised against a stable transition state analogue, it may have enzymatic properties. Such **catalytic antibodies** can then be engineered with molecular biology methods for improved catalytic properties. That way it is possible to generate enzymes for reactions that do not occur in nature, a process of considerable commercial interest. Although some catalytic antibodies have indeed been obtained, results with this approach have been largely disappointing. Although enzymes can accelerate a reaction by a factor of 10^8 to 10^{11}, the best catalytic antibodies reach about 10^6. Thus other possible mechanisms of enzyme catalytic action are currently investigated, based in particular on quantum mechanics. It is postulated that electrons or protons can tunnel through the energy hill, rather than crossing it. This tunnelling may be aided by vibrational movement in the protein (**vibration assisted tunnelling**). This approach, however, is still in its infancy.

5.5.1 *Molecular Mechanism of Serine-Proteases and -Esterases*

Proteins and esters are relatively stable molecules, because their hydrolysis involves the formation of an unstable intermediate that easily reverts back to the original compound.



Acids and bases can increase the speed of hydrolysis by catalysing the conversion between the first and second intermediate.

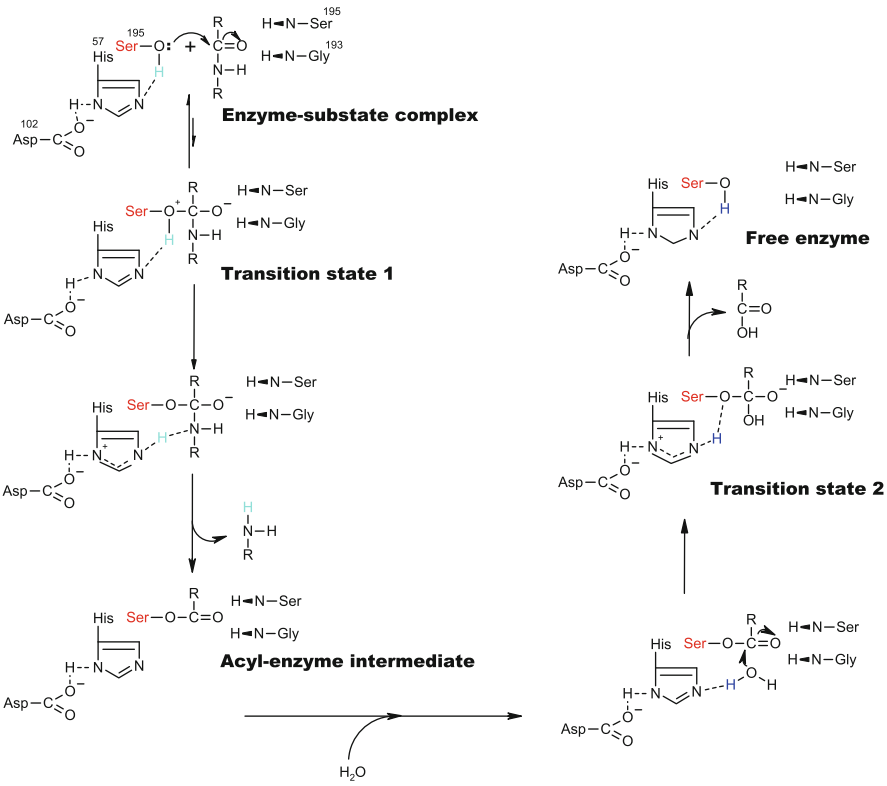


Fig. 5.10 Reaction mechanism of Ser-hydrolases. Many proteases (such as chymotrypsin) and esterases (such as acetylcholine esterase) contain Ser in their active centre. This serine forms a so-called "catalytic triad" with two other residues, His and Asp. The His acts as a general base and removes the proton from the Ser-substrate complex; the resulting positive charge on the His is stabilised by the negative charge on the Asp. The His then transfers the proton to the nitrogen on the leaving group. Thus no positive charge develops on the Ser-oxygen, and its nucleophilicity is increased. The negative charge on the carbonyl-oxygen is stabilised by hydrogen bonding to additional Ser- and Gly-nitrogens (called the "oxanion hole")

This is the reason that such compounds are caustic to our skin.

In Ser-dependent hydrolases the H^+ on the nucleophile (Ser) is removed quickly by other amino acids; this prevents the transition state from reverting to substrate (see Fig. 5.10).

5.6 Exercises

5.6.1 Problems

5.1. The plasma membrane of a cell contains 10 000 receptors for the hormone X. Of these, 1000 need to have hormone bound to stimulate the maximal response in the cell. The dissociation constant $K_d = 1$ nM. How high does the concentration of the hormone need to be to get maximal response of the cell?

A 0.05 nM
B 0.11 nM
C 0.25 nM
D 0.50 nM
E 0.73 nM

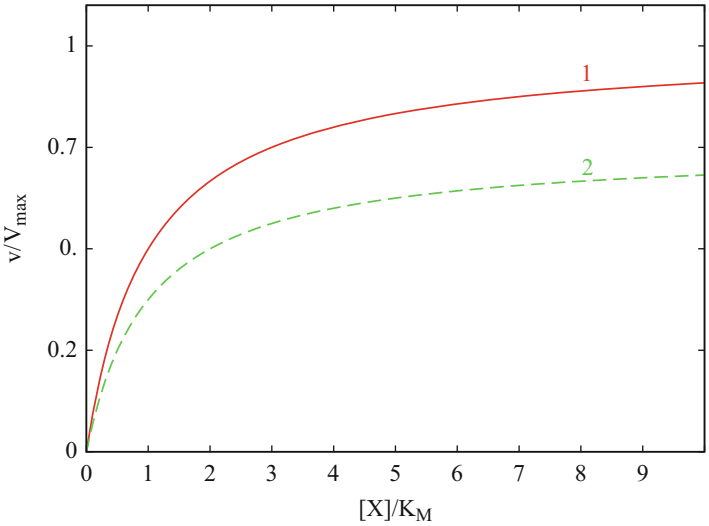
5.2. If a cell is constantly stimulated by the hormone X it internalises the hormone receptor by endocytosis. This reduces the number of receptor molecules on the cell surface. Assume that the above cell has only 5000 receptor molecules left. How high would the $[X]$ have to be to elicit a maximal response?

A 0.05 nM
B 0.11 nM
C 0.25 nM
D 0.50 nM
E 0.73 nM

5.3. An enzyme is turning over at a substrate concentration of $[S] = 3 \times K_m$. The reaction velocity will be

A 0.25 * $V_{\max}$.
B 0.33 * $V_{\max}$.
C 0.50 * $V_{\max}$.
D 0.75 * $V_{\max}$.
E $V_{\max}$.

5.4. In the figure the enzymatic reaction rate (v) of a substrate X is plotted as a function of $[X]$. Which change(s) is/are required to get from curve 1 to curve 2?



a) Increase the total concentration of X.
b) Decrease the number of enzyme molecules.
c) Increase the affinity of the enzyme for X.
d) All of the above.
e) None of the above.

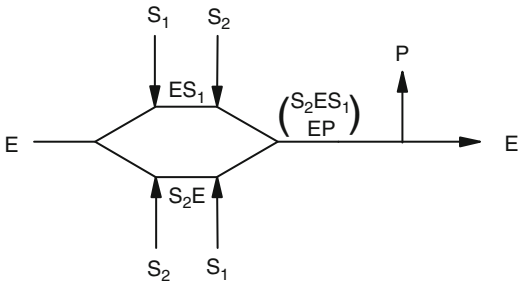
5.5. 10 µg of a pure enzyme with a molecular mass of 50 kDa turn over 10 µmol/min of substrate. The turnover number is _____

5.6. Two enzymes compete for the same substrate. Most of the substrate will be turned over by the enzyme with:

a) The higher molecular mass.
b) The higher activity and lower K_M-value.
c) The lower activity and higher K_M-value.
d) A *p*I-value nearer to the *p*H of the solution.
e) The higher specificity for the substrate.

5.7. You are supposed to measure the kinetics of an enzyme reaction. Critically discuss possible experimental pitfalls and the methods to control for them.

5.8. The mechanism of the following reaction is called: _____



5.9. Under conditions of substrate saturation you double [E]. The velocity v

A) remains constant.
B) is halved.
C) is doubled.
D) is increased 10-fold.
E) is reduced 100-fold.

5.10. The enzyme glucose phosphate isomerase catalyses the reaction Fru-6P $\rightleftharpoons$ Glu-6P, $\Delta G'^0 = -1.8$ kJ/mol. For this reaction to proceed in the physiological direction (toward Fru-6P) one has to

A) Remove [Glu-6P] in a second reaction.
B) Remove [Fru-6P] in a second reaction.
C) Do nothing, as reactions proceed toward positive ΔG.
D) Increase the [Fru-6P].
E) Increase the enzyme concentration.

5.11. At a substrate concentration of 1 mM an enzyme is turning over at 75 % of its maximal velocity. The K_m is

A 0.25 mM
B 0.33 mM
C 0.50 mM
D 0.75 mM
E 1.00 mM

5.6.2 Solutions

5.1

$$\mathrm{X} + \mathrm{R} \rightleftharpoons \mathrm{RX}$$

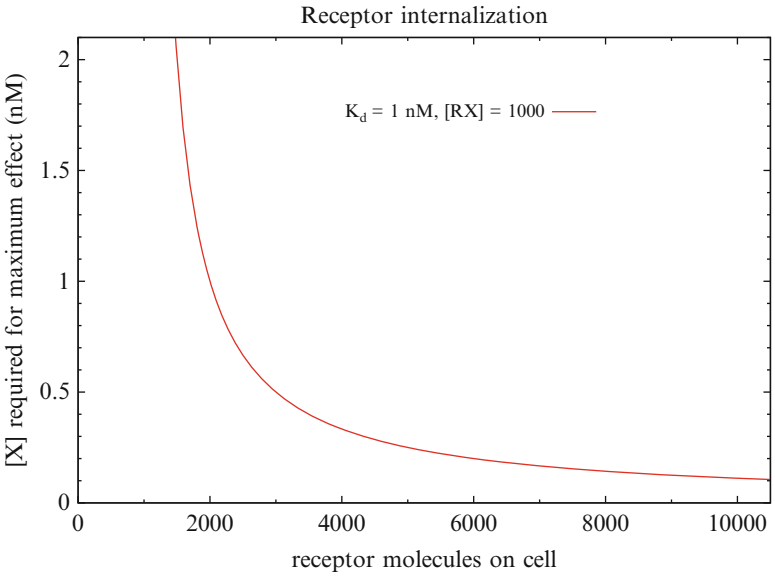
$$K_d = \frac{[R][X]}{[RX]} = \frac{([R]_t - [RX])[X]}{[RX]} \qquad | \times [RX] \tag{5.42}$$

$$K_d[RX] = ([R]_t - [RX])[X] \qquad | : ([R]_t - [RX]) \tag{5.43}$$

$$[X] = \frac{K_d[RX]}{([R]_t - [RX])} \tag{5.44}$$

$$= \frac{1\text{nM} \times 1000}{10000 - 1000} = \frac{1}{10-1}\text{nM} = \frac{1}{9}\text{nM} \tag{5.45}$$

$$\approx 0.11\text{nM} \tag{5.46}$$



5.2

- Receptor internalization reduces the sensitivity of a cell toward a hormone X.
- Note: If $[R]_t = 2000$ and $[RX] = 1000$, then $[X] = K_d$: physical meaning of K_d.

5.3 Because the substrate concentration is higher than K_m the reaction velocity needs to be larger than 0.5 * V_{max}, so answers A–C have to be incorrect. On the other hand, V_{max} is reached only at [S] = ∞, thus answer E is incorrect too. That leaves D as the only possible answer, which can be confirmed by solving the HMM-equation: $v = \frac{3K_m}{K_m + 3 * K_m} * V_{max} = 3/4 * V_{max}$.

5.4 **A**: In curve 2 the maximal rate is reduced, but the K_m is unchanged. Such behaviour is obtained by reducing the number of enzyme molecules.

5.5 If 50 000 g are equivalent to 1 mol of enzyme, then 10 µg are equivalent to 0.2 nmol. Thus 1 molecule of enzyme will turn over 10.000 nmol min^{-1}/0.2 nmol = 50.000 molecules of substrate per min. Note that a correct answer must also contain the unit!

5.6 Higher activity (faster turnover) and lower K_m-value (higher affinity) both are important. Molecular mass and *p*I have little to do with it. Specificity would influence only the competition of two substrates for the same enzyme.

5.7 For a (nonexclusive) list see Section 5.1.4

5.8 This is a reaction with two substrates and one product; the substrates may bind in any order. Hence the reaction mechanism is called random-bi.

5.9 When the substrate concentration is (nearly) saturating, $v \approx V_{max}$. However, $V_{max} = k_{cat} \times [E]$, thus doubling [E] doubles the velocity.

5.10 LE CHATELIER's principle: removal of the product stimulates its formation.

5.11 We have to rearrange the HMM-equation: $v = V_{max} \times [S] / (K_m + [S]) \Rightarrow K_m = V_{max} \times [S] / (v - [S]) = 0.33\,mM$

References

1. W.J. Albery, J.R. Knowles, Evolution of enzyme function and the development of catalytic efficiency. Biochemistry **15**, 5631–5640 (1976). doi: 10.1021/bi00670a032
2. G.E. Briggs, J.B.S. Haldane, A note on the kinetics of enzyme action. Biochem. J. **19**, 338–339 (1925)
3. W.W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products 1. Nomenclature and rate equations. Biochim. Biophys. Acta **67**, 104–137 (1963). doi: 10.1016/0926-6569(63)90211-6
4. R.G. Duggleby, Quantitative analysis of the time courses of enzyme-catalyzed reactions. Methods **24**(2), 168–174 (2001). doi: 10.1006/meth.2001.1177
5. R. Eisenthal, A. Cornish-Bowden, The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters. Biochem. J. **139**(3), 715–720 (1974). URL http://www.biochemj.org/bj/139/0715/1390715.pdf
6. E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme. Ber. Dtsch. Chem. Ges. **27**(3), 2985–2993 (1894). doi: 10.1002/cber.18940270364
7. J.B.S. Haldane, Graphical methods in enzyme chemistry. Nature **179**, 832 (1957). doi: 10.1038/179832b0
8. J.B.S. Haldane, K.G. Stern, *Allgemeine Chemie der Enzyme* (Steinkopf, Dresden, 1932)
9. V. Henri, Theorie generale de l'action de quelques diastases. Compt. rend. hebd. Acad. Sci. Paris **135**, 916–919 (1902). URL http://gallica.bnf.fr/ark:/12148/bpt6k64607148/f1257.image
10. D.E. Koshland, Enzyme flexibility and enzyme action. J. Cell. Comp. Physiol. **54**(Suppl. 1), 245 (1959). doi: 10.1002/jcp.1030540420
11. H.J. Lee, I.B. Wilson, Enzymic parameters: Measurement of V and K_m. Biochim. Biophys. Acta **242**(3), 519–522 (1971)
12. D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**(2), 431–441 (1963). URL http://137.204.42.130/~bittelli/materiale_lettura_fisica_terreno/marquardt_63.pdf
13. L. Michaelis, M.L. Menten, Die Kinetik der Invertin-Wirkung. Biochem. Z. **49**, 333–369 (1913). URL http://path.upmc.edu/divisions/chp/PDF/Michaelis-Menten_Kinetik.pdf
14. J.A. Nelder, R. Mead, A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965). doi: 10.1093/comjnl/7.4.308

15. J.E. Nielsen, J.A. McCammon, Calculating pKa values in enzyme active sites. Protein Sci. **12**, 1894–1901 (2003). doi: 10.1110/ps.03114903
16. B.A. Orsi, K.F. Tipton, Kinetic analysis of progress curves. Meth. Enzymol. **63**, 159–183 (1979). doi: 10.1016/0076-6879(79)63010-0
17. O. Warburg, W. Christian, Pyridin, der wasserstoffübertragende Bestandteil von Gährungsfermenten. Biochem. Z. **287**, 291–328 (1936). doi: 10.1002/hlca.193601901199

Chapter 6
Inhibition and Inactivation of Enzymes

Abstract In addition to their substrates, enzymes also bind substances that reduce their activity. Such binding can be reversible resulting in inhibition, or irreversible resulting in inactivation of the enzyme. We distinguish four types of inhibition:

competitive The inhibitor binds only to the free enzyme, not to the enzyme substrate complex. The EI-complex cannot bind substrate and therefore has no catalytic activity.

uncompetitive The inhibitor binds only to the enzyme substrate complex, not to the free enzyme. The EIS-complex has no catalytic activity.

noncompetitive The inhibitor binds to both the free enzyme and to the enzyme substrate complex. The EI-complex can still bind to the substrate, but the EIS-complex has no catalytic activity.

partial The EIS-complex still has some catalytic activity. It can be either partially noncompetitive or partially uncompetitive.

These inhibition types result in characteristic patterns in LINEWEAVER-BURK-plots.

Inactivators reduce enzymatic activity in a time-dependent manner. Of particular medical use are suicide-inactivators, which have to be converted into the inactivating species by the enzyme itself. These are very specific and often have few side effects.

The usual view of enzymes (and receptors) is that they bind their substrates very specifically, but nothing else. If this were the case it would be impossible to influence disease with pharmaceuticals, because most of them work by binding to enzymes, receptors, and transporters (see Fig. 6.1), blocking their action (Fig. 6.2). Binding can be **reversible** or **irreversible**; reversible binding results in **inhibition**, and irreversible binding in **inactivation** of the enzyme or receptor (the latter is dealt with in the next section).

6.1 Enzyme Inhibition

Depending on the mode of interaction among inhibitor, substrate, and enzyme we distinguish several forms of inhibition. **Note**: The nomenclature was worked out by W. W. CLELAND [1–3], but is often misrepresented in textbooks. You may therefore have learned things differently in previous courses. The nomenclature represented

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_6

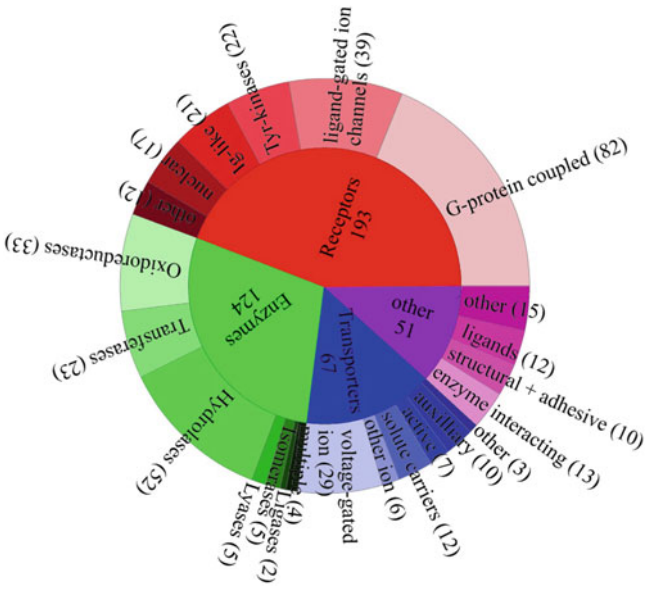


Fig. 6.1 The 435 targets that our medicine chest is directed against [4]

here is, however, used by professional enzymologists and largely agrees with a proposed IUBMB-standard (www.chem.qmul.ac.uk/iubmb/kinetics/).

competitive The inhibitor binds only to the free enzyme, not to the enzyme substrate complex. The EI-complex cannot bind substrate and therefore has no catalytic activity.

uncompetitive The inhibitor binds only to the enzyme substrate complex, not to the free enzyme. The EIS-complex has no catalytic activity.

noncompetitive The inhibitor binds to both the free enzyme and to the enzyme substrate complex. The EI-complex can still bind substrate, but the EIS-complex has no catalytic activity.

partial The EIS-complex still has some catalytic activity. It can be either partially noncompetitive or partially uncompetitive.

We now look at these four mechanisms in more detail (see also Table 6.1).

6.2 Competitive Inhibition

In competitive inhibition substrate and inhibitor compete for the enzyme *i.e.*; they cannot bind at the same time:

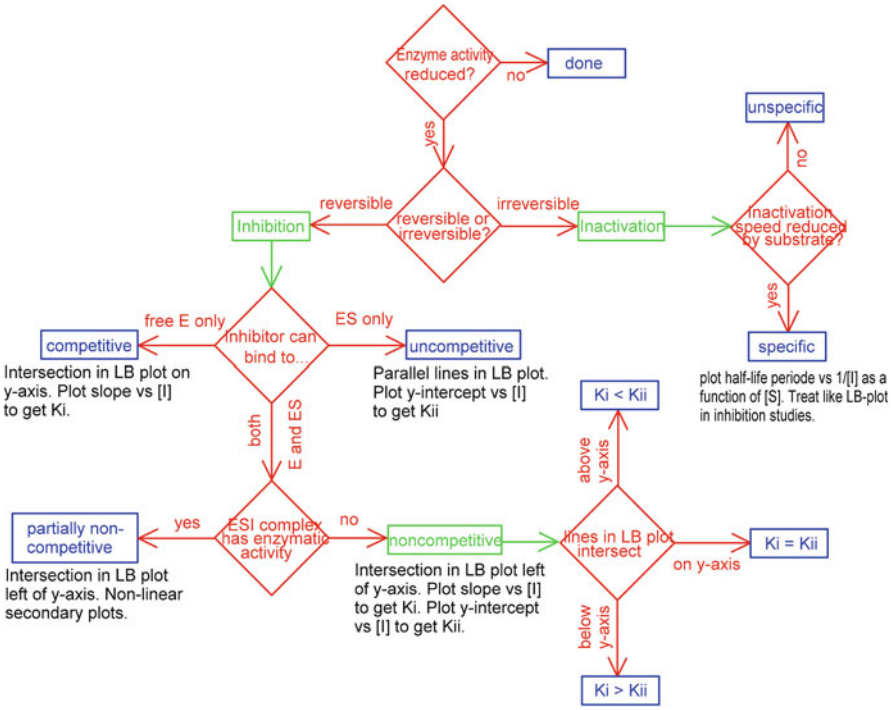
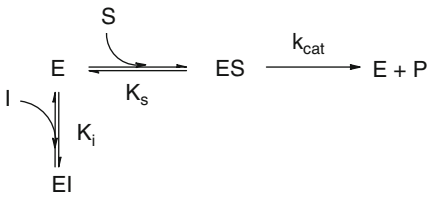


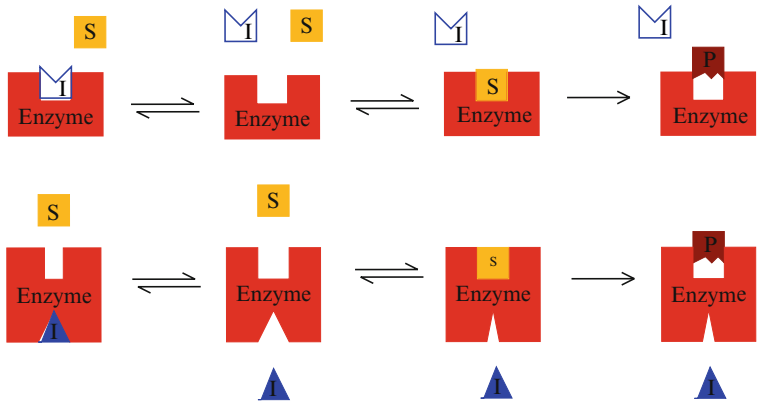
Fig. 6.2 Decision tree for enzyme kinetics: Summary of the various forms of inhibition and inactivation



From a competitive inhibition mechanism it has often been concluded that substrate and inhibitor share the same binding site on the enzyme. This, however, is not necessarily correct; if substrate binding changes the conformation of the enzyme in such a way that the inhibitor can no longer bind and *vice versa*, then the inhibition mechanism will also be competitive:

Table 6.1 If the x- or y-intercepts or the slopes of the lines in a LINEWEAVER-BURK are plotted as a function of the inhibitor concentration (secondary plot), the resulting straight lines can be used to estimate K_i and K_{ii}. With the advent of personal computers, this method should no longer be used, but the diagrams are still useful for data presentation. Partially noncompetitive inhibition does not result in straight lines in secondary plots, so this inhibition type is not included in the table

	Slope vs. [I]			y-intercept vs. [I]			x-intercept vs. [I]		
Inhibition type	x-intercept	y-intercept	Slope	x-intercept	y-intercept	Slope	x-intercept	y-intercept	Slope
Competitive	$-K_\mathrm{i}$	$\frac{K_\mathrm{m}}{V_\mathrm{max}}$	$\frac{K_\mathrm{m}}{K_\mathrm{i}V_\mathrm{max}}$	–	$\frac{1}{V_\mathrm{max}}$	0	$-K_\mathrm{i}$	K_m	$\frac{K_\mathrm{m}}{K_\mathrm{i}}$
Uncompetitive	–	$\frac{K_\mathrm{m}}{V_\mathrm{max}}$	0	$-K_\mathrm{ii}$	$\frac{1}{V_\mathrm{max}}$	$\frac{1}{K_\mathrm{ii}V_\mathrm{max}}$	–	–	–
Noncompetitive	$-K_\mathrm{i}$	$\frac{K_\mathrm{m}}{V_\mathrm{max}}$	$\frac{K_\mathrm{m}}{K_\mathrm{i}V_\mathrm{max}}$	$-K_\mathrm{ii}$	$\frac{1}{V_\mathrm{max}}$	$\frac{1}{K_\mathrm{ii}V_\mathrm{max}}$	–	–	–



Of course, if substrate and inhibitor do share the same binding site, then the inhibition mechanism must be competitive as no two objects can occupy the same space at the same time.

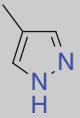
In the presence of a competitive inhibitor the HMM-equation and its LINEWEAVER-BURK-transform need to be modified as follows:

$$v = \frac{V_{\max} * [S]}{(1 + \frac{[I]}{K_i}) * K_m + [S]} \tag{6.1}$$

$$\frac{1}{v} = \frac{1}{V_{\max}} + \left(1 + \frac{[I]}{K_i}\right) * \frac{K_m}{V_{\max}} * \frac{1}{[S]} \tag{6.2}$$

Because high $[S]$ can exclude the inhibitor from binding to the enzyme, the effect of any given $[I]$ can be counteracted by increasing $[S]$. However, as the $[I]$ increases, higher and higher $[S]$ are required to achieve a given v. Thus V_{max} is not changed in competitive inhibition (see Fig. 6.3). In the LINEWEAVER-BURK-plot all lines intersect at a common point on the y-axis, that is, at a common $1/V_{max}$. If one plots the slopes of the lines in the LINEWEAVER-BURK-PLOT as a function of $[I]$ (*not* its reciprocal!) the data points are on a straight line, K_i can be determined from its intersection with the x-axis (**secondary plot**).

Clinical Use of Competitive Inhibition



4-methyl pyrazole

(Fomepizole) Ingested **methanol** is oxydised by alcohol dehydrogenase to methanal (formaldehyde), which is further metabolised to formic acid by

(continued)

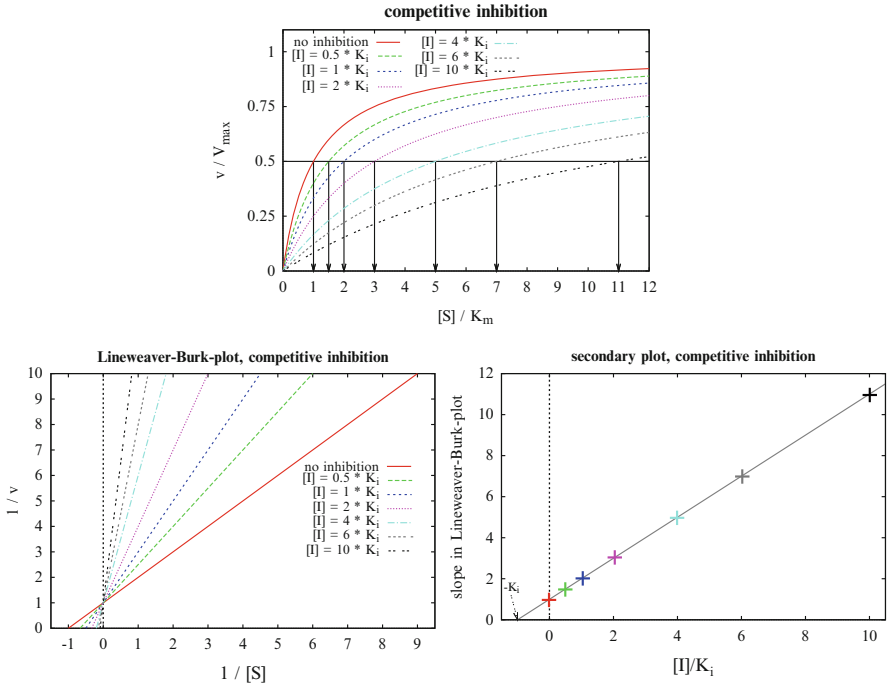


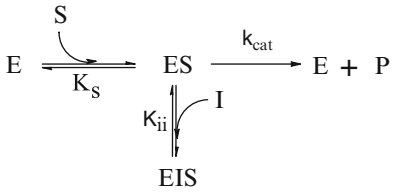
Fig. 6.3 Competitive inhibition. *Top*: plot of v vs $[S]$ at several $[I]$. *Bottom*: LINEWEAVER-BURK-transformation and secondary plot

aldehyde dehydrogenase. The resulting acidosis can be lethal and damages the optic nerve, leading to blindness. **Ethanol** inhibits methanol oxydation competitively and hence protects the optic nerve from damage. Ethanol concentration has to be kept very high (close to lethal) for several days, until the methanol has been excreted by the kidneys. Instead of ethanol, one can also use **Fomepizole**, which has a much higher affinity for alcohol dehydrogenase than ethanol. It can therefore be used in concentrations that do not cause unconsciousness in the patient. Unfortunately, it is extremely expensive.

Other important pharmaceuticals including methotrexate or sulphonamides also act as competitive inhibitors.

6.3 Uncompetitive Inhibition

In uncompetitive inhibition the inhibitor reacts exclusively with the enzyme-substrate complex, not with the free enzyme:



If we increase [*S*], we increase [*ES*] and this will lead to higher [*ESI*]. Increasing [*I*] will pull ES into ESI; following LE CHATELIER's principle more E must be converted to ES to maintain the equilibrium constant K_s. As a result, the inhibitor increases the apparent affinity of the enzyme for the substrate and *vice versa*.

Inasmuch as we cannot combat the inhibitor's effect by increasing [*S*], V_{max} must decrease with increasing [*I*] (see Fig. 6.4). Because the intersections on the y- and x-axis in the LINEWEAVER-BURK-plot are the reciprocals of K_m and V_{max}, respectively, both go farther away from 0 with increasing [*I*]. Characteristic for

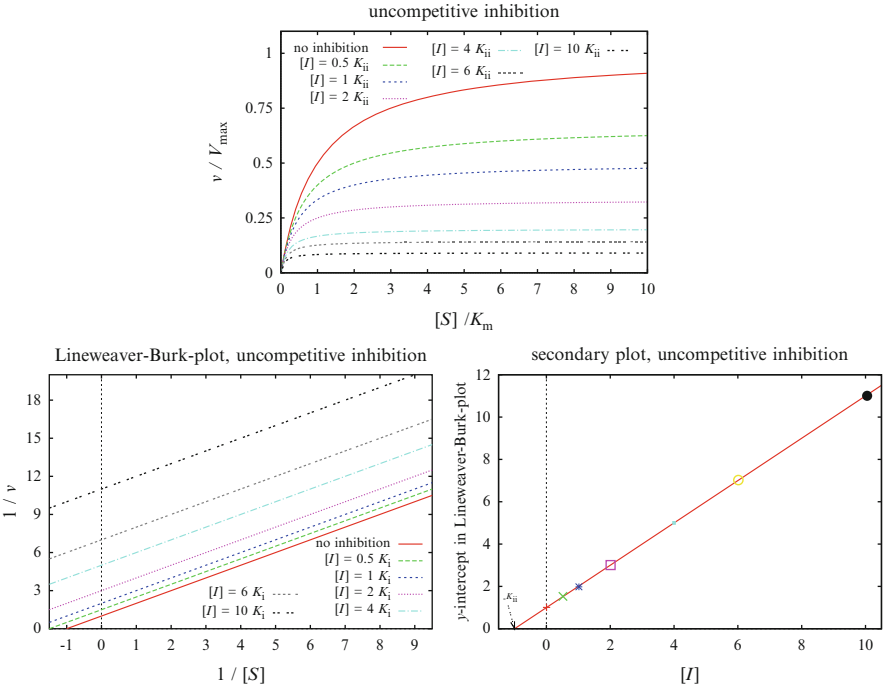


Fig. 6.4 Uncompetitive inhibition. *Top*: plot of *v* versus [*S*] at several [*I*]. *Bottom*: LINEWEAVER-BURK-transformation and secondary plot

uncompetitive inhibition are the parallel lines in the LINEWEAVER-BURK-plot. Plotting the y-intercepts of the LINEWEAVER-BURK-plot against $[I]$ in a secondary plot gives a straight line, intersecting the x-axis at $-K_{ii}$.

Uncompetitive inhibition is rare and is a sign that the enzyme acts as an oligomer.

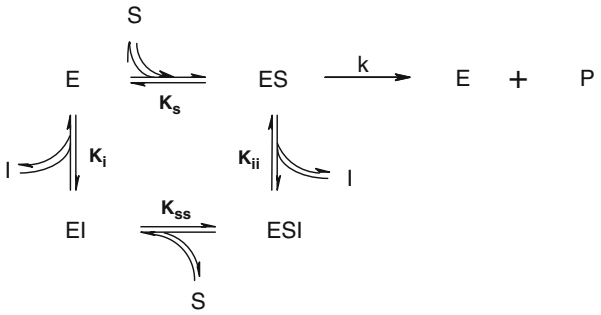
The HMM-equation and its LINEWEAVER-BURK-transform for this case are:

$$v = \frac{V_{max} * [S]}{K_m + (1 + \frac{[I]}{K_{ii}}) * [S]} \tag{6.3}$$

$$\frac{1}{v} = \left(1 + \frac{[I]}{K_{ii}}\right) * \frac{1}{V_{max}} + \frac{K_m}{V_{max}} * \frac{1}{[S]} \tag{6.4}$$

6.4 Noncompetitive Inhibition

In noncompetitive inhibition both the free enzyme and the enzyme-substrate complex can bind the inhibitor:



Going around the cycle from E via EIS back to E (i.e., without product formation) the free energy change must be zero for reasons of energy conservation, irrespective of whether S or I binds first or dissociates first. Thus

$$\frac{K_s K_{ii}}{K_i K_{ss}} = 1 \quad \Rightarrow \quad \frac{K_s}{K_{ss}} = \frac{K_i}{K_{ii}} \equiv \alpha \tag{6.5}$$

This is known as the **law of microreversibility**. In other words, whatever effect the binding of S to E has for the binding of I must be mirrored exactly by the effect the binding of I has on the binding of S. If binding of S reduces the affinity of E for I, then binding of I must also reduce the affinity of E for S. We encounter such linked reactions frequently throughout this book; an example is the regulation of the oxygen affinity of hemoglobin by 2,3-BPG (see page 173).

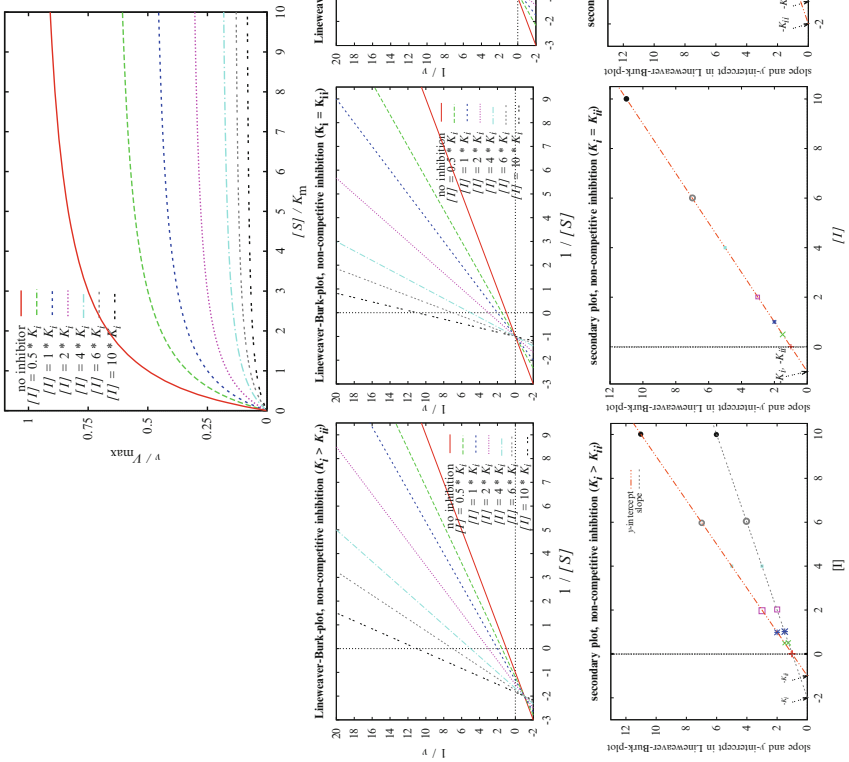


Fig. 6.5 Noncompetitive inhibition. *Top*: plot of v vs $[S]$ at several $[I]$. *Bottom*: LINEWEAVER-BURK-transformation and secondary plot for the three possible cases, $K_i > K_{ii}$, $K_i = K_{ii}$ and $K_i < K_{ii}$

What does this mean? Let's look at the three possible cases in turn (see also Fig. 6.5):

$K_i = K_{ii}$ This means that the affinity of the enzyme for the inhibitor is independent of whether it has substrate bound. In other words, substrate binding does not change the conformation of the enzyme in such a way that the affinity for the inhibitor would be changed. If this is the case, however, then the opposite must also be true: $K_s = K_{ss}$. In the LINEWEAVER-BURK-plot all lines intersect in a common point on the x-axis, because K_s and K_m are related and if K_s is not changed by inhibitor binding, K_m can't be either.

$K_i > K_{ii}$ In this case the affinity of E for the inhibitor is smaller than the affinity of ES (remember that dissociation constants are reciprocals of the affinity!). In other words, binding of substrate changes the conformation of the enzyme so that binding of the inhibitor is facilitated. This, however, must also mean that the affinity of EI for the substrate is higher than that of E; binding of I increases the affinity and lowers the dissociation constant (and hence the apparent K_m).

$K_i < K_{ii}$ This is the opposite case from above, binding of the substrate changes the conformation of the enzyme in such a way that binding of the inhibitor is made more difficult. Then in turn binding of the inhibitor must change the enzyme conformation so that binding of the substrate becomes more difficult; the dissociation constant (and hence the apparent K_m) is increased.

K_i and K_{ii} are determined from secondary plots, K_{ss} is then calculated from the law of microreversibility.

Noncompetitive inhibition looks like a mixture of competitive and uncompetitive inhibition (indeed, IUBMB suggest the term "mixed inhibition" for this case), and this is also apparent from the equations describing this case:

$$v = \frac{V_{max} * [S]}{(1 + \frac{[I]}{K_i}) * K_m + (1 + \frac{[I]}{K_{ii}}) * [S]} \tag{6.6}$$

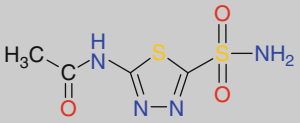
$$\frac{1}{v} = \left(1 + \frac{[I]}{K_{ii}}\right) * \frac{1}{V_{max}} + \left(1 + \frac{[I]}{K_i}\right) * \frac{K_m}{V_{max}} * \frac{1}{[S]} \tag{6.7}$$

There may be more than one binding site for the inhibitor on the enzyme; then

$$\log\left(\frac{V_0}{v} - 1\right) = n \log([I]) - \log K \tag{6.8}$$

Hence, a plot of $\log(\frac{V_0}{v} - 1)$ versus $\log([I])$ for several substrate concentrations will result in parallel lines.

Medical Use of Noncompetitive Inhibition

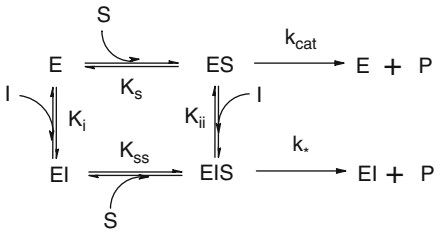


A pharmaceutical working by noncompetitive inhibition is **acetazolamide** (Diamox), which binds to an essential Zn^{2+} ion in carbonic anhydrase. The enzyme catalyses the reaction $CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$, for example, on the luminal membranes of kidney tubule cells. The bicarbonate in the primary urine is broken down to carbon dioxyde, which diffuses through the membrane into the cell and then into the blood. If carbonic anhydrase is inhibited, more bicarbonate ions stay in the urine; Na^+ and K^+ follow for electroneutrality and water osmotically. As a result urine volume is increased (diuresis). By a similar mechanism the drug is effective in treating increased intraocular pressure (**glaucoma**) and increased intracranial pressure (resulting in **absence seizures**). It is also used to prevent or treat **acute mountain sickness** which may befall people at great heights where the air pressure is reduced and hence the amount of oxygen in a given volume of air. The hypoxæmia results in faster breathing (**hyperventilation**), and this in turn to a loss of carbon dioxyde (**hypocapnia**) resulting in **alkalosis**. Inhibition of carbonic anhydrase slows the decomposition of bicarbonate into carbon dioxyde and hence helps to stabilise blood *p*H.

6.5 Partial Inhibition

If the EIS complex can form product, albeit with a lower velocity than ES, we speak of partial inhibition. Partial inhibition is rare in practice; pharmacologically speaking it would be useless: if we need to inhibit an enzyme to help a patient, we don't want it to retain activity after binding the inhibitor! In physiology you will learn about **partial (ant)agonists** for receptors.

Partially noncompetitive inhibition is similar to the noncompetitive:



The LINEWEAVER-BURK-plot (Fig. 6.6) also looks similar to the noncompetitive case (all three possibilities exist too), but the secondary diagram is curved. This allows partially noncompetitive and noncompetitive inhibition to be easily

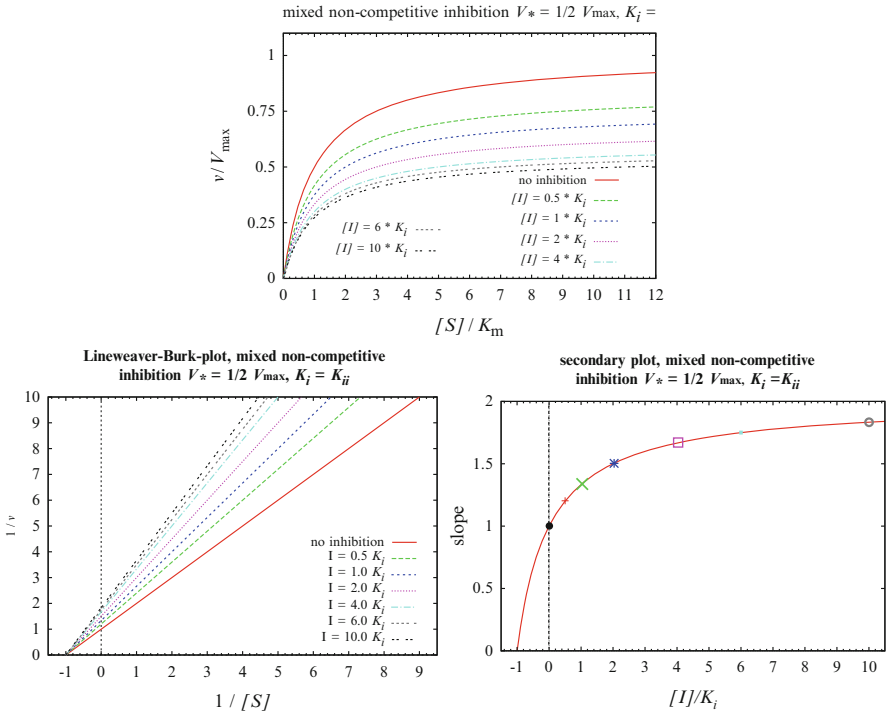


Fig. 6.6 Partially noncompetitive inhibition. *Top*: Plot of v versus $[S]$ at several $[I]$. *Bottom*: LINEWEAVER-BURK-transformation and secondary plot

distinguished. Note that partially noncompetitive inhibition is the most universal mechanism of inhibition; by simply setting some rate constants to 0 we can get any of the other mechanisms from it. For the sake of completeness, the HMM-equation and its LINEWEAVER-BURK-transform for this case are:

$$v = \frac{V_{\max} + \frac{V_* [I]}{K_{\mathrm{ii}}} [S]}{K_{\mathrm{m}} \left(1 + \frac{[I]}{K_{\mathrm{i}}}\right) + \left(1 + \frac{[I]}{K_{\mathrm{ii}}}\right) [S]} \tag{6.9}$$

$$\frac{1}{v} = \frac{1 + \frac{[I]}{K_{\mathrm{ii}}}}{V_{\max} + \frac{V_* [I]}{K_{\mathrm{ii}}}} + \frac{K_{\mathrm{m}} \left(1 + \frac{[I]}{K_{\mathrm{i}}}\right)}{V_{\max} + \frac{V_* [I]}{K_{\mathrm{ii}}}} \times \frac{1}{[S]} \tag{6.10}$$

In analogy to partially noncompetitive there is also a partially uncompetitive inhibition, but this can safely be ignored by anybody but professional enzymologists. There cannot be a partially competitive inhibition, as in competitive inhibition there is no EIS complex, which therefore cannot have enzymatic activity.[1]

[1] In the older literature one may find the expression "partially competitive" for a special case of partially noncompetitive inhibition with $k_{\mathrm{cat}} = k_*$. This results in a LINEWEAVER-BURK-plot

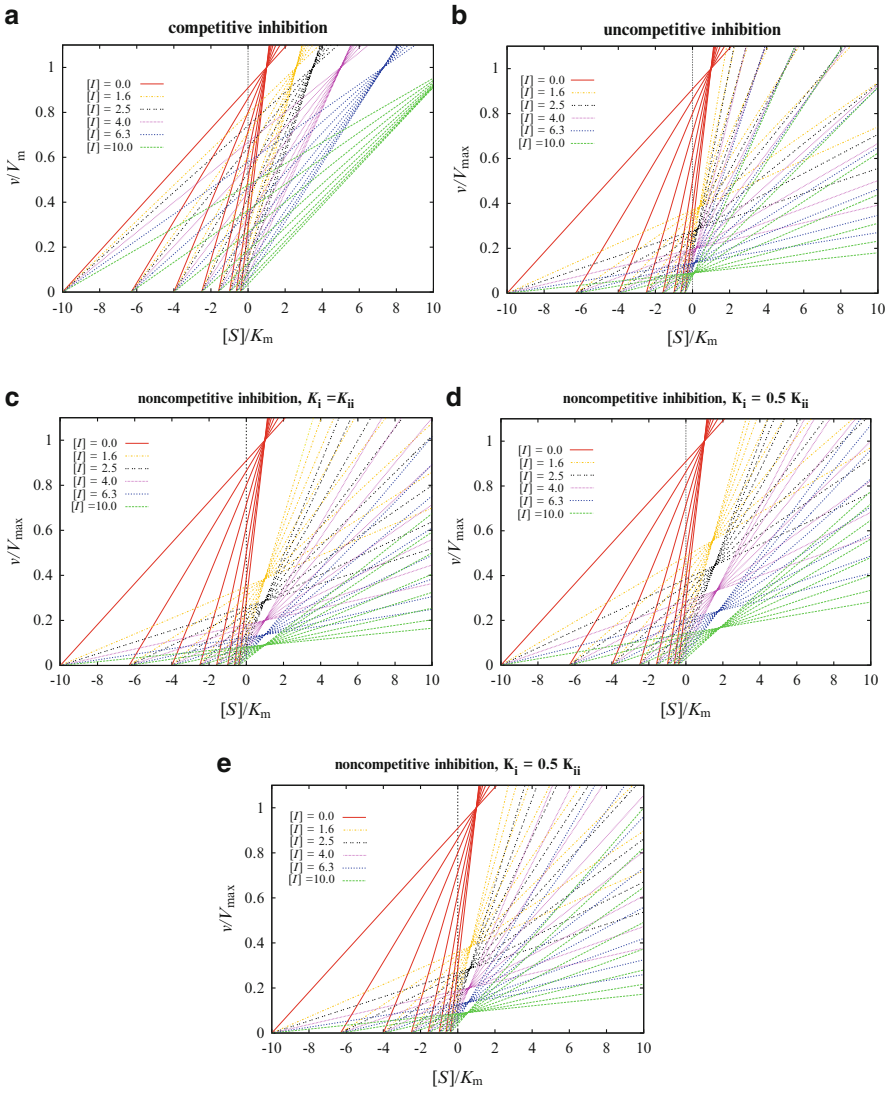


Fig. 6.7 The various types of inhibition produce characteristic patterns in direct plots, just as in LINEWEAVER-BURK-plots. **(a)** Competitive inhibition. As the apparent V_{max} doesn't change, the intersections lie on a horizontal line. **(b)** Uncompetitive inhibition. Both V_{max} and K_m change. **(c)** Noncompetitive inhibition with $K_i = K_{ii}$. Since then also $K_s = K_{ss}$ there is no change in apparent K_m, the intersections lie on a vertical line. **(d)** + **(e)** Noncompetitive inhibition with $K_s \neq K_{ss}$. In those cases the apparent K_m does change

resembling competitive inhibition, but with curved secondary plots. The term is mechanistically incorrect and outdated.

6.6 Inactivation of Enzymes

Inactivation occurs when a substance binds so tightly to an enzyme, that it can no longer dissociate.

Important clinical examples are aspirin (acetyl salicylate)—which blocks the synthesis of prostaglandins, important mediators of inflammation and pain—and β-lactam antibiotics such as penicillin (see Fig. 6.8).

Another example for enzyme inactivating compounds is organophosphates, which covalently bind to Ser-residues in proteins. Ser occurs in the catalytic centre of many hydrolases, for example, proteases and esterases. Phosphoesters are used as pesticides (Parathion) and chemical weapons (VX, Tabun), because they inactivate acetylcholine esterase, an enzyme that breaks down acetylcholine into acetic acid and choline in the synaptic gap of our motor neurons. The result is a constant contraction of all muscles in the body,

(continued)

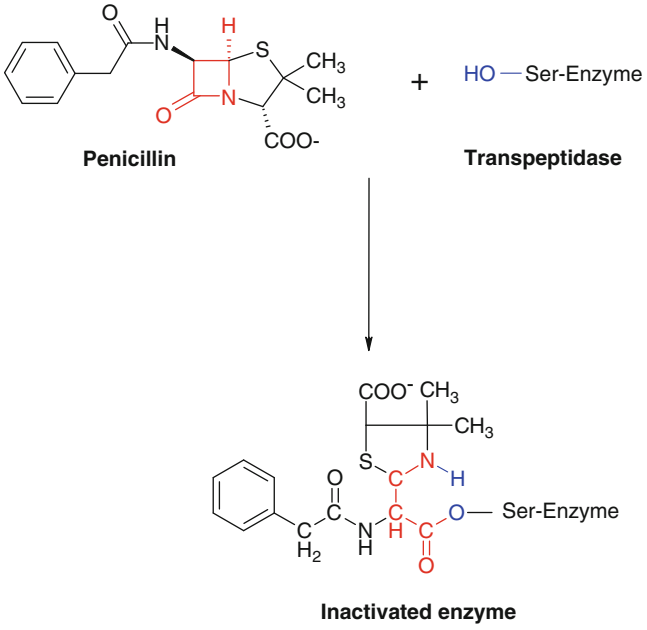


Fig. 6.8 Structure of penicillin, an important antibiotic. The β-lactam ring (*marked red*) opens in the substrate binding site of D-alanine transpeptidase to form an ester with an essential serine residue of the enzyme. Hence formation of the bacterial cell wall is prevented and bacteria lyse during cell division. Recall that 4-membered ring systems are reactive because the bond angle is forced to 90° rather than the tetrahedral angle of 109° (BAYER-tension)

resulting in painful death from suffocation and heart failure. Treatment is by competitive inhibition of the acetylcholine receptor in the postsynaptic membrane by its antagonist atropine.

A special case is the so-called **suicide inactivators**. These substances are substrate analogues which are enzymatically converted to the inactivating compound. This makes such inactivators very specific for a particular enzyme, and ideal drugs. Formally, the reaction of suicidal inactivators resembles the reaction with substrates, except that the enzyme is not regenerated at the end:

$$E + I \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} E \cdot I \xrightarrow{k_{+2}} E - I \tag{6.11}$$

If the association to the loose complex is faster than the conversion to the final stable complex the reaction rate will be determined by the second step, which as a first-order reaction is described by the following equations:

$$\frac{d[E]}{dt} = k_{+2} * [E] \tag{6.12}$$

$$[E]_t = [E]_0 * e^{-k_{+2} * t} \tag{6.13}$$

$$\ln([E]_t) = \ln([E]_0) - k_{+2} * t \tag{6.14}$$

Thus plotting the natural logarithm of the remaining enzyme activity ($[E]_t$) against time (t) will result in a straight line, the slope of which is the reaction rate constant k_{+2}. If k_{+2} is plotted as a function of the inactivator concentration, a hyperbola is obtained. Then $k_{+2\max}$ is the inactivation rate at infinite inactivator concentration, when $[E \cdot I] = [E]_0$. The concentration of inactivator that results in half-maximal reaction rate is called K_i (Fig. 6.9).

As with enzyme turnover, the LINEWEAVER-BURK-transformation can be used, by plotting $1/k_{+2}$ (the **relaxation time** τ, time required for the enzyme activity to drop to 1/e $\approx$ 36.8 % of the original) or, alternatively, the **half life period** $\mathbf{t}_{1/2} = \ln(2)/k_{+2}$ (time required for the remaining activity to drop by 50 %) against $1/[I]$.

The presence of substrate then may result in a protection of the enzyme against inactivation, which can be either competitive (substrate and inactivator exclude each other from binding to the enzyme), noncompetitive (substrate and inactivator can bind to the enzyme at the same time, but EIS cannot turn into the stable complex), or uncompetitive (substrate can bind only to EI, but EIS cannot form the final complex).

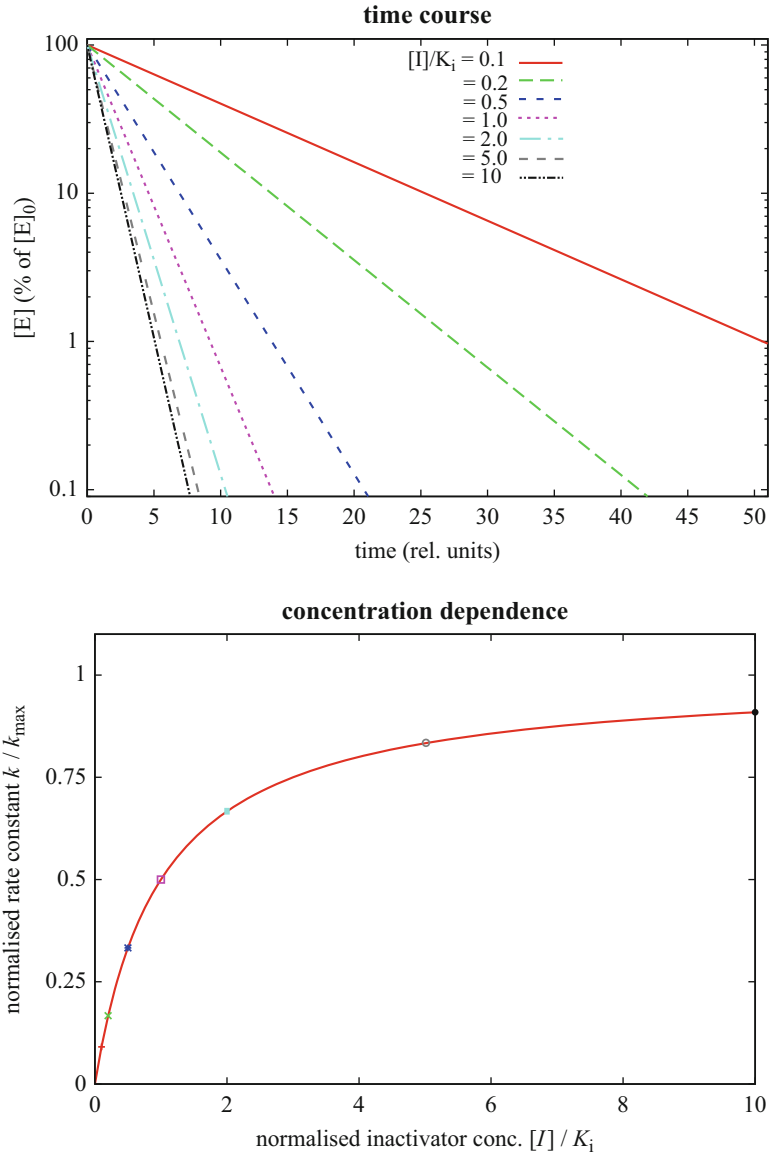
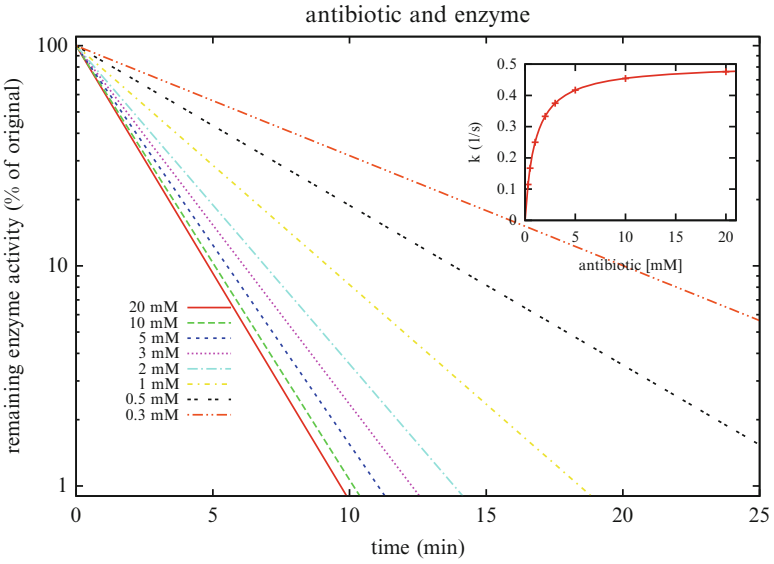


Fig. 6.9 *Top:* Time course of the inactivation of an enzyme at various concentrations of inactivator. Note that the y-axis has a logarithmic scale. *Bottom:* If the slopes are plotted against the inactivator concentration, a hyperbola is obtained

6.7 Exercises

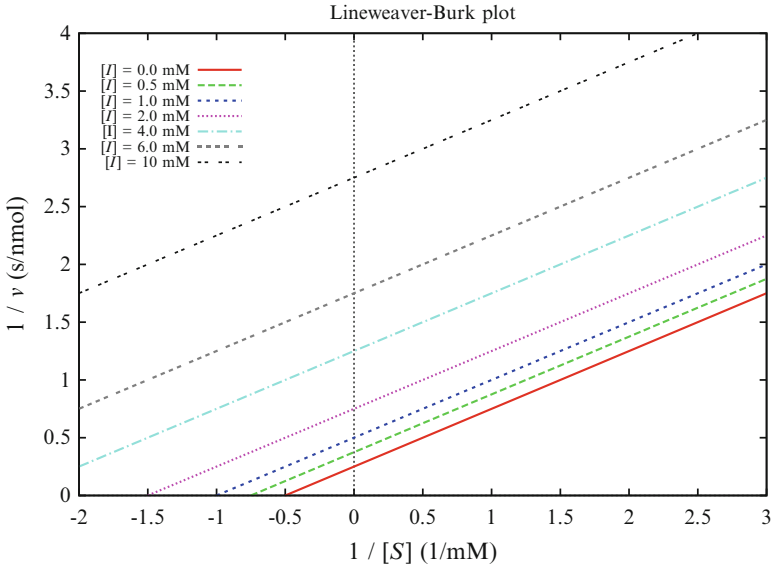
6.7.1 Problems

6.1. You are employed by a drug company developing a new antibiotic against tuberculosis. Testing the interaction of a candidate substance with an enzyme the bacterium needs for division you obtain the diagram below.



Which of the following terms best describes the interaction between the potential drug and the target enzyme?

A) Competitive inhibition.
B) Noncompetitive inhibition.
C) Uncompetitive inhibition.
D) Mixed competitive inhibition.
E) Inactivation by a suicide substrate.



6.2. The diagram shows the Lineweaver-Burk-plot of an experiment: The interaction is

A) The competitive inhibition of an enzyme with a K_m of 1 mM.
B) The competitive inhibition of an enzyme with a K_m of 2 mM.
C) The noncompetitive inhibition of an enzyme with a K_m of 1 mM.
D) The uncompetitive inhibition of an enzyme with a K_m of 1 mM.
E) The uncompetitive inhibition of an enzyme with a K_m of 2 mM.

6.3. Some phosphoesters are used as insecticides (Parathion, E605) and chemical weapons (VX, sarin, tabun). They act by chemically modifying Ser-groups in the active center of acetylcholine esterase in the synaptic gap of the neuro-muscular junction. This reaction most likely results in a

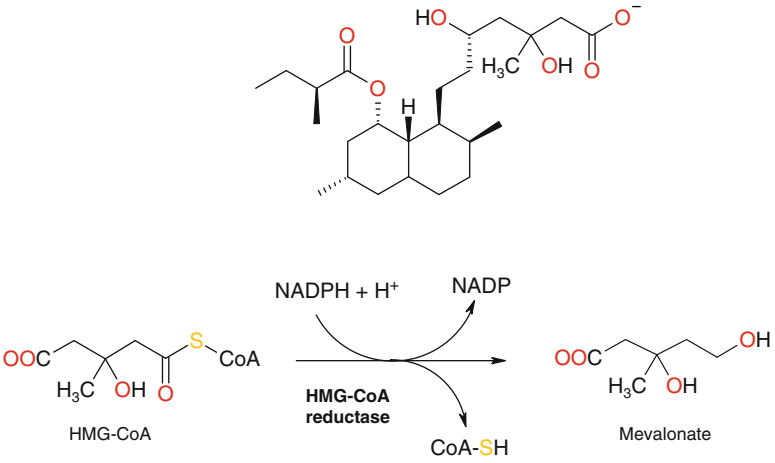
A Competitive inhibition
B Uncompetitive inhibition
C Noncompetitive inhibition
D Mixed inhibition
E Inactivation

6.4. Methotrexate is an anticancer drug which competitively inhibits dihydrofolate reductase. In the presence of this drug

A) The $V_{\max}$ of the enzyme will be reduced.
B) The apparent K_m will increase.
C) A LINEWEAVER-BURK plot will show parallel lines.
D) The inhibitor will bind only to the ES-complex.
E) The inhibitor will bind to both the free enzyme and the ES-complex.

6.5. Red yeast rice is produced by fermenting rice grains with the mold *Monascus purpureus* for 3–6 d. It is used as a spice and as food colour, for example, in rice wine. In traditional Chinese medicine (TCM) red yeast rice has been used since

the Tang Dynasty (800 A.D.) to "remove blood blockages". It turned out that red yeast rice contains the polyketide Lovastatin. After activation in our bodies Lovastatin reduces the activity of HMG-CoA reductase, the enzyme responsible for the first committed step of cholesterol biosynthesis. It is used to treat hypercholesterolæmia. The figure shows the reaction catalysed by HMG-CoA reductase and the structure of activated Lovastatin.



Based on this information, which is the most likely interaction between drug and enzyme?

A Competitive inhibition
B Uncompetitive inhibition
C Noncompetitive inhibition
D Partially competitive inhibition
E Inactivation

6.6. From a student's lab notes: the velocity of *para*-nitrophenylphosphate (pNPP) hydrolysis by acid phosphatase was measured as function of [S], both in the presence and absence of 2 mM phenylphosphate (PP). The following results were obtained:

$[S]$/mM	v/nKat [PP] = 0 mM	v/nKat [PP] = 2 mM
0.2	19.3	2.4
0.4	23.7	3.5
0.8	31.6	6.9
1.4	34.2	10.5
2.0	32.4	13.4
4.0	35.9	19.5
6.0	36.5	20.0
8.0	44.6	38.8

Discuss the result, using a direct plot.

6.7.2 Solutions

6.1 This is a time-dependent inactivation. Inhibition would be an equilibrium reaction that occurs on the ms timescale.

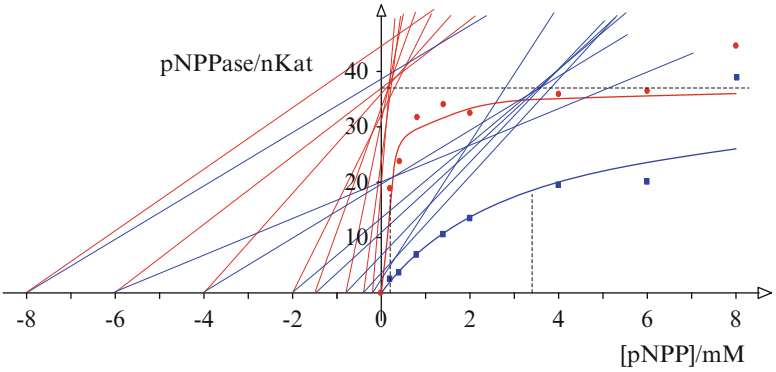
6.2 The parallel lines indicate an uncompetitive inhibition. To get K_m we locate the line in the absence of inhibitor (red). Its x-intercept is $-0.5\,\mathrm{mM}^{-1}$, thus $K_m = 2\,\mathrm{mM}$

6.3 Covalent modification results in irreversible loss of activity → inactivation

6.4 Definition of competitive inhibition: binding only to free E, $V_{\max}$ constant and apparent K_m increased. LINEWEAVER-BURK plot lines with common intersection at y-axis ($1/V_{\max}$).

6.5 The top part of Lovastatin looks very much like the substrate and interacts with the substrate binding site of the enzyme, resulting in competitive inhibition.

6.6



The intersection of two lines defined by two points each $(x_1, y_1), (x_2, y_2)$ and $(x_3, y_3), (x_4, y_4)$ is:

$$(x, y) = \frac{(x_1y_2 - y_1x_2)(x_3 - x_4) - (x_1 - x_2)(x_3y_4 - y_3x_4)}{(x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4)}, \tag{6.15}$$

$$\frac{(x_1y_2 - y_1x_2)(y_3 - y_4) - (y_1 - y_2)(x_3y_4 - y_3x_4)}{(x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4)} \tag{6.16}$$

Considering that $x2 = x4 = y1 = y3 = 0$, this simplifies to

$$(x, y) = \frac{x_1y_2x_3 - x_1x_3y4}{x_1(-y_4) + y_2x_3}, \frac{x_1y_2(-y_4) + y_2x_3y_4}{x_1(-y_4) + y_2x_3} \tag{6.17}$$

For the uninhibited reaction this produces the following solutions (in mM and nKat, respectively):

$$\begin{pmatrix} 0.12, 30.7 & 0.22, 40.1 & 0.21, 39.3 & 0.16, 35.0 & 0.19, 37.6 & 0.19, 37.7 & 0.28, 46.2 \\ 0.40, 47.4 & 0.30, 41.6 & 0.20, 35.7 & 0.24, 38.1 & 0.24, 38.0 & 0.39, 46.8 & \\ 0.17, 38.4 & 0.03, 33.0 & 0.14, 37.2 & 0.15, 37.4 & 0.38, 46.7 & & \\ 0.22, 28.9 & 0.11, 36.9 & 0.13, 37.3 & 0.55, 47.7 & & & \\ 0.48, 40.3 & 0.41, 39.0 & 1.15, 51.0 & & & & \\ 0.21, 37.8 & 2.56, 58.9 & & & & & \\ 15.9, 133.4 & & & & & & \end{pmatrix} \tag{6.18}$$

The K_m and $V_{\max}$ values need to be sorted, so that one can determine the quartiles. For K_m this gives:

0.03, 0.11, 0.12, 0.13, 0.14, 0.15, **0.16**, 0.17, 0.19, 0.19, 0.20, 0.21, 0.21, **0.22, 0.22**, 0.24, 0.24, 0.28, 0.30, 0.38, 0.39, **0.40**, 0.41, 0.48, 0.55, 1.15, 2.56, 15.9

Thus the K_m would be in the range 0.16–0.40 mM, probably near 0.22 mM. The $V_{\max}$ of the uninhibited reaction is in the range 37–47 nKat, most likely near 38 nKat.

For the inhibited reaction we get K_m = 2.1 – 8.5 mM, most likely near 3.4 mM, about 15-fold higher than that for the uninhibited reaction. For $V_{\max}$ we get a range of 27 – 64 nKat, most likely around 36 nKat, not significantly different from the uninhibited reaction. Thus the inhibition of pNPP hydrolysis by PP is competitive.

Notice how the process of determining quartiles removes the extreme outliers in the dataset. Even though the data quality is poor one can still get a reasonable estimate of the kinetic parameters of the reaction, and identify the inhibition type with confidence.

References

1. W.W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products 1. Nomenclature and rate equations. Biochim. Biophys. Acta **67**, 104–137 (1963a). doi:10.1016/0926-6569(63)90211-6
2. W.W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products 2. Inhibition: Nomenclature and theory. Biochim. Biophys. Acta **67**, 173–187 (1963b). doi:10.1016/0926-6569(63)90226-8
3. W.W. Cleland, The kinetics of enzyme-catalyzed reactions with two or more substrates or products 3. Prediction of initial velocity and inhibition patterns by inspection. Biochim. Biophys. Acta **67**, 188–196 (1963c). doi:10.1016/0926-6569(63)90227-X
4. M. Rask-Andersen, M.S. Almén, H.B. Schiöth, Trends in the exploitation of novel drug targets. Nature Rev. Drug Discov. **10**(8), 579–590 (2011). doi:10.1038/nrd347

Chapter 7
Hæmoglobin and Myoglobin: Cooperativity

Blood is a very special juice (J.W. v. GOETHE: Faust I)

Abstract Myoglobin is a protein with a single hæme group, which binds oxygen inside our cells and makes it available for metabolism. Oxygen binding to myoglobin is described by a hyperbola. Hæmoglobin consists of two α,β-protomers, each of the four proteins contains a hæme group. Binding of oxygen to one of these hæme groups increases the affinity of the remaining sites for oxygen; this cooperation results in an S-shaped (sigmoidal) binding curve. The shape of this curve is influenced by *p*H and the presence of 2,3-Bisphosphoglycerate (2,3-BPG). Mutations in the genes for either the α- or the β-subunits of hæmoglobin can cause serious disease.

Hæmoglobin is the most abundant protein in mammals; an adult male contains about 1 kg of it, equivalent to 3.5 g iron, about 80 % of the total iron in our body. The hæmoglobin concentration in erythrocyte cytosol is 34 % by weight.

Hæmoglobin and myoglobin are not only very important proteins that attract funding, they are also available in large amounts and easily purified. Even human hæmoglobin can be produced from outdated blood conserves without ethical problems, albeit with some risk of infection. For these reasons myoglobin and in particular hæmoglobin are probably the best characterised proteins on this planet (Table 7.1).

The 3D-structure has been solved by X-ray crystallography, and the structural changes that occur during binding of O_2, CO_2 and poisons such as CO are known in great detail. The structural consequences of mutations in these proteins have also been investigated.

The kinetics of binding of ligands has been minutely studied; indeed the science of enzyme kinetics owes a lot to the progress made in studying oxygen binding to hæmoglobin, which has been dubbed an "enzyme *honoris causa*" for this reason. Ironically, the latest research shows that hæmoglobins in some species indeed have enzymatic properties; they are involved in the destruction of nitric oxyde and some xenobiotics ($2NO\cdot + 2O_2 + NAD(P)H + H^+ \rightarrow 2HNO_3 + NAD(P)^+$ [6, 7]).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_7

Table 7.1 Some important milestones in hæmoglobin research

Year	Author	Discovery
1825	J.F. ENGELHART	Iron makes 0.5 % of the weight of hæmoglobin in all species. Thus hæmoglobin is a defined molecule with a molecular mass of $n * 11$ kDa. This result was reasonably close to the modern $4 * 16.7$ kDa, given the crude techniques available at the time
1853	L. TEICHMANN	Isolates hæm by treating hæmoglobin with NaCl/glacial acetic acid. The protein remains as colourless precipitate, the hæm can be crystallised from the solution
1864	F. HOPPE-SEYLER	Suggests the name hæmoglobin and crystallises it (first crystallisation of a protein)
1904	C.H.L.P.E. BOHR	Measures oxygen binding to hæmoglobin as function of oxygen and carbon dioxyde concentration
1912	W. KÜSTER	Discovers the tetrapyrrole ring structure of hæm. In the following year R. WILLSTÄTTER finds a similar structure in chlorophyll, but with Mg^{2+} instead of Fe as the central atom
1914	J.S. HALDANE	Describes oxygen binding as equilibrium reaction with monodisperse hæmoglobin; the binding constant depends on the animal species used
1920	L.J. HENDERSON	Investigates the influence of *p*H on the oxygen binding curve. The results are interpreted by J.B.S. HALDANE (son of J.S. HALDANE)
1925	G.S. ADAIR	Determines the molecular mass of hæmoglobin to be 65 kDa by measuring its osmotic pressure. He also discovers that oxygen binding to hæmoglobin is a cooperative process
1926	J.B. CONANT	Discovers the coordination of iron with amino acid side chains in hæmyoglobin
1929	H. FISCHER & K. ZEILE	Synthesise hæm
1934	T. SVEDBERG	Measures the molecular mass of hæmoglobin with his newly developed ultracentrifuge and finds hæmoglobin to consist of a single, well-defined molecule
1935	F. HAUROWITZ	Isolates HbF and compares oxygen binding curves of adult and fetal hæmyoglobin
1940	R.R. PORTER & F. SANGER	Show that all hæmoglobin molecules have the same end-groups
1959	M. PERUTZ & J.C. KENDREWS	Solve the three-dimensional structure of hæmoglobin by X-ray crystallography

This surprise finding shows that our knowledge about these proteins is far from complete. The role of this enzymatic activity for human physiology and pathology is still under discussion [8].

7.1 Structure

Both myoglobin, the oxygen storage protein in our tissues (and neuroglobin, which serves the same function in neuronal tissues), and hæmoglobin, the oxygen transport protein in erythrocytes, contain hæm as the oxygen binding site. Hæm is a porphyrin molecule with Fe^{2+} bound in its centre. Because it is buried deep in a small pocket inside the protein without direct access to the medium, hæm cannot leave the protein. Such firmly bound molecules in the active centre of a protein are called **prosthetic groups**. Binding of hæm into the protein has several functional consequences:

- Free iron can catalyse the formation of reactive oxygen species by the FENTON-reaction, which are dangerous to cells. Iron bound inside myoglobin or hæmoglobin has a significantly reduced tendency to do that.
- Free hæm has an extremely high affinity for some poisons, such as nitric oxyde and carbon monoxyde. Carbon monoxyde binds about 20 000 times better to free hæm than oxygen, whereas in myoglobin it is only 200-fold better. The reason is the different orbital geometry of oxygen and carbon monoxyde; whereas the axis of the O_2 molecule forms an angle with the Fe−O bond (see Fig. 7.1), CO binds perpendicular to the hæm ring. Steric hindrance by the distal His residue reduces the affinity of hæm for CO in myoglobin and hæmoglobin.
- In free hæm oxydation of the iron centre to Fe^{3+} destroys the oxygen binding site. The iron-histidine bond seems to protect the iron centre against oxydation.
- In a protein the affinity of hæm for oxygen can be controlled by protein conformation. As we see in a moment, this has important physiological consequences.

Myoglobin is a monomeric protein with a single hæm group hæmoglobin is a heterotetrameric protein (2 α- and 2 β-subunits in normal adult hæmoglobin); each of the subunits has its own hæm (Fig. 7.2). And this apparently small difference has important consequences.

7.2 Oxygen Binding and Cooperativity

In the enzyme section we have derived an equation for the binding of a ligand to a protein as a function of ligand concentration:

$$\Theta = \frac{[S]}{K_d + [S]} \tag{7.1}$$

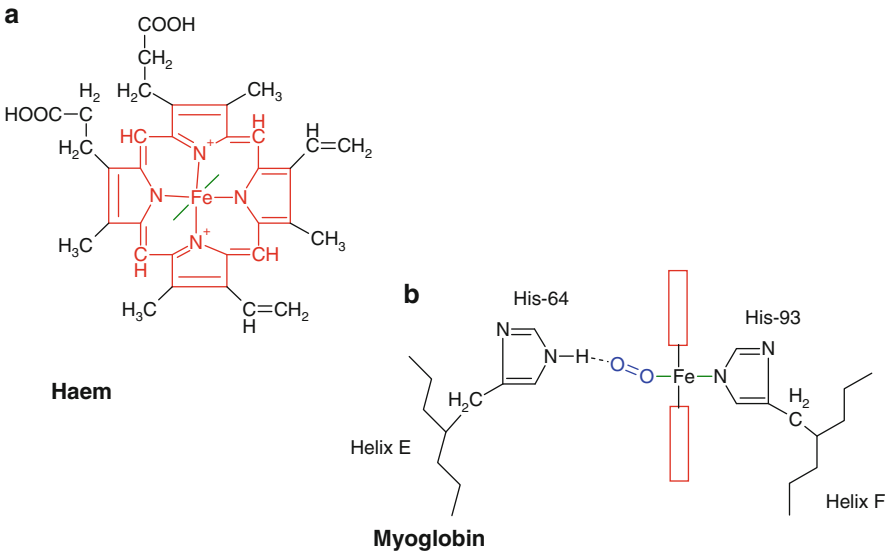


Fig. 7.1 Binding of hæm to myoglobin. (**a**) Hæm is a flat molecule with iron at its centre. The iron is held in place by complex bonds to the 4 ring nitrogen atoms. The remaining 2 coordination sites of iron are oriented perpendicular to the hæm ring system. (**b**) One of these binds to a His residue of myoglobin (called the proximal His)and the second binds the oxygen molecule. Binding of oxygen to hæm is stabilised by a hydrogen bond to a second (distal) His residue. Note that the axis of the oxygen molecule forms an angle with the Fe-O bond; in the case of the poison CO the system would be straight. The subunits of hæmoglobin have a similar structure

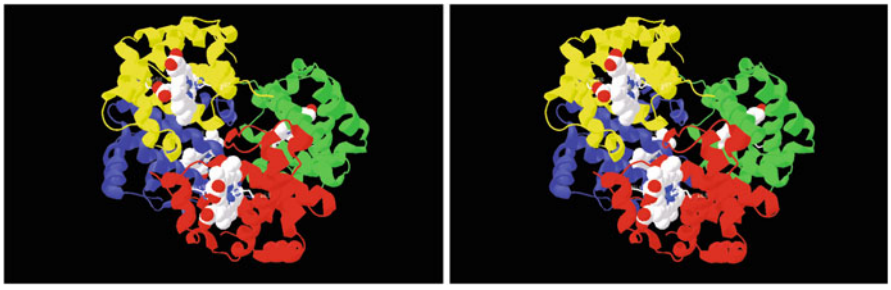


Fig. 7.2 Three-dimensional structure of hæmoglobin, in stereo (PDB-code 1HGA). The molecule consists of 2 protomers, each containing an α- and a β-subunit. The 4 hæm molecules are in protected cavities inside the subunits

Oxygen binding to myoglobin is well described by this equation. However, the binding curve of hæmoglobin looks completely different (see Fig. 7.3):

If oxygen binds to one of the 4 subunits in hæmoglobin, the conformation of the whole protein changes, and with it the affinity of the other 3 binding sites for oxygen. This interaction is called **cooperativity**. There are three key points about cooperativity that you have to remember:

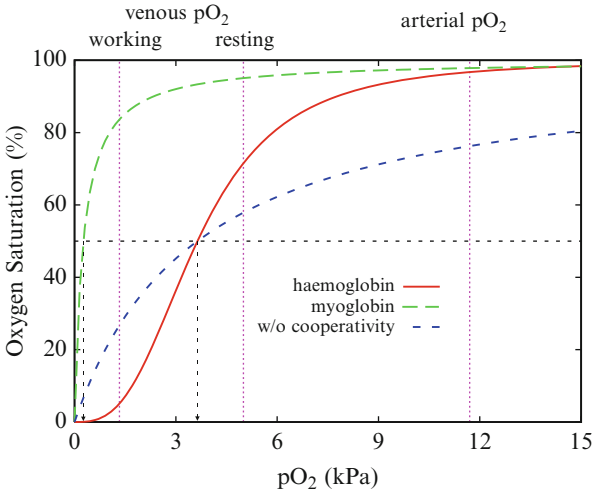


Fig. 7.3 Oxygen binding to myoglobin, hæmoglobin and a hypothetical protein with the same half-saturation point as hæmoglobin, but a hyperbolic binding curve. Oxygen partial pressure in normal air is about 20 kPa; in the lung it is ≈ 13 kPa. Venous pO_2 depends on the exercise status (1–5 kPa)

- Cooperativity requires the coexistence of several binding sites for a ligand in a protein, either on the same or (usually) on different subunits.
- Cooperativity results in sigmoidal (S-shaped) rather than hyperbolic binding curves.
- Cooperativity allows the body to regulate ligand binding and, in enzymes, substrate turnover to its needs in a given physiological situation. You will find cooperative kinetics in enzymes at key regulatory steps in metabolism (*e.g.*, phosphofructokinase in glycolysis).

7.2.1 Functional Significance of Cooperativity

Let us look at the function of myoglobin and hæmoglobin, so we can appreciate the importance of cooperativity:

Hæmoglobin picks up oxygen in the lungs, and transfers it onto myoglobin in the tissues. In the lungs, oxygen partial pressure is about 13 kPa, somewhat lower than pO_2 in outside air (20 kPa). Arterial oxygen partial pressure is about 11 kPa, in tissues, pO_2 is between 1 and 5 kPa, depending on the physiological situation. Myoglobin is half-saturated at 0.26 kPa, so even in the low pO_2 of an exercising muscle myoglobin will be 85 % saturated with oxygen, and deliver it to oxygen-utilising enzymes.

Hæmoglobin should be almost saturated with oxygen at the pO_2 of arteries (13 kPa), but should give off almost all bound oxygen in tissue, at 1 kPa. If you look at Fig. 7.3, you will find that a hypothetical protein with hyperbolic binding curve and a the same half-saturation point as hæmoglobin cannot do this. In the arteries, only 77 % of the molecules would bind oxygen, and in tissue 24–66 % of the molecules would retain their oxygen. In other words only 11–53 % of the transport capacity of such a protein would be used. Simply shifting the half-saturation point is no solution either: a lower affinity would improve oxygen release in the tissue but result in lower oxygen loading in the lung; a higher affinity would have opposite effects. Thus neither change would improve net oxygen transport.

Now look at the sigmoidal binding curve of hæmoglobin. At arterial pO_2 hæmoglobin is about 98 % saturated with oxygen. In a resting muscle, with $pO_2 = 5$ kPa hæmoglobin is about 72 % saturated, in a working muscle with $pO_2 = 1$ kPa only 2 %. This corresponds to a transport efficiency of 26–96 % and makes hæmoglobin a virtually ideal transport protein: almost completely filled at the source and, if required, virtually empty at the sink.

Note also how myoglobin buffers the oxygen supply in muscle: a change of hæmoglobin saturation from 2 to 76 % results in a 80–95 % saturation range for myoglobin.

7.2.2 Mechanism of Cooperativity

How has evolution achieved this miracle of adaptation? As already mentioned the key is cross-talk between the binding sites of hæmoglobin. Each binding site can occur in a low-affinity state (called the “tense” or T state) or in a high-affinity (“relaxed” or R) state.

Historical Note
The names “tense” and “relaxed” were originally chosen because in the T-state hæmoglobin molecules are stabilised by additional ionic bonds between the α- and β-subunits, which do not exist in the R-state. The reason is that in the absence of oxygen the hæm iron is pulled toward the proximal His, putting a strain on the hæm molecule that results in a movement of the surrounding protein helices. Oxygen binding releases this strain.

If there are no oxygen molecules bound to hæmoglobin, all 4 subunits will be in the T state, and their affinity to oxygen will be low.

If all 4 subunits have oxygen bound, then all will be in the R state and have a high affinity for oxygen.

You can already see how this works: with low oxygen saturation (in tissue) affinity will be low, and hæmoglobin will tend to release oxygen completely. With high oxygen saturation (in the lungs) hæmoglobin will have high affinity, and tend to bind as much oxygen as possible—just what is functionally required.

The interesting question is: what happens at intermediate oxygen pressures? How is the transition between all-T and all-S achieved? There are two models for this, the **concerted model** of J. MONOD *et al.* [12] and the **sequential model** suggested by D. KOSHLAND *et al.* [11] (see Fig. 7.5).

Distinguishing Between Concerted and Sequential Models

There is considerable experimental evidence for the concerted model in various proteins (see Fig. 7.4, reviewed in [4]). For example, the nicotinic acetylcholine receptor (AChR) is a homopentamer. Each subunit consists of an extracellular acetylcholine binding domain, a transmembrane domain that forms the ion channel, and a cytoplasmic domain. Patch-clamping of these receptors reveals stochastic all-or-nothing opening; the opening time, but not the conductivity, is increased by the agonist acetylcholine and decreased by the antagonist α-bungarotoxin. Thus all five subunits are either in the open or in the closed state, as otherwise a stepwise rather than an all-or-nothing opening should be seen. Agonists and antagonists simply affect the likelihood of finding the protein in either state. Some mutations in AChR affect the equilibrium between open and closed state in the absence of ligand, but not ligand binding affinity. This argues for conformational selection rather than induced fit.

The phenomenon that binding of a molecule at one site may influence binding at other sites on the same protein is called **allostery** (from the Greek words for "other shape"). When both ligands are identical, the interaction is called **homotropic**; this is equivalent to cooperativity. If the molecules are different, the interaction is called **heterotropic**. Some proteins have both homotropic and heterotropic interactions.

Thermodynamics and Mechanism of Allostery

Allostery requires consecutive binding of a ligand A (e.g., a substrate) and an effector X to a protein E (e.g., an enzyme):

(continued)

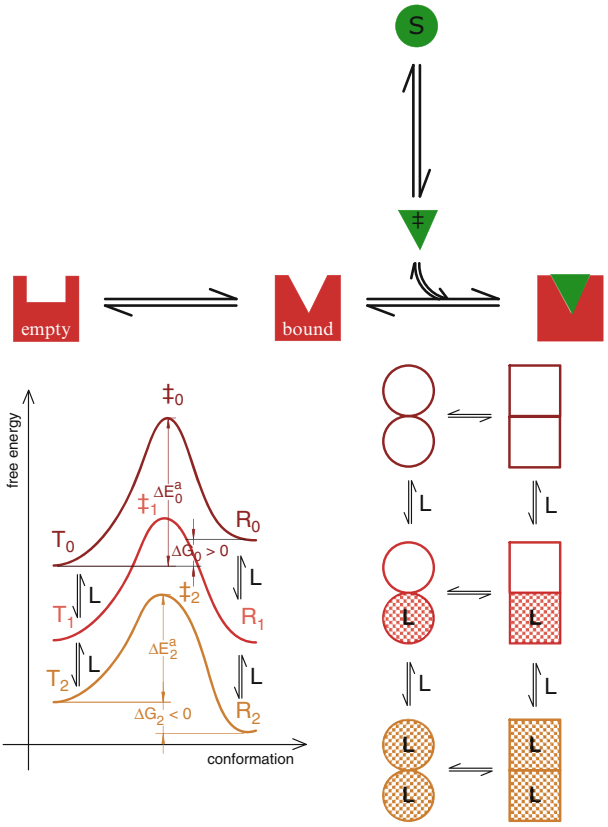


Fig. 7.4 *Top*: Conformational selection model. The protein can spontaneously switch between "empty" (T) and "bound" (R) conformation; binding of the ligand to the R conformation then stabilises it. This is in contrast to the induced fit model, where ligand binding causes the protein to assume the R conformation. The model is shown for a monomer, as might occur in enzyme/substrate interactions. Note that the ligand may also have to change conformation to one that fits the enzyme. See Fig. 14.6 on page 331 for an example of conformational selection. *Bottom*: Energetics of conformational selection (modified after [4]). This model is shown for a dimer, the minimal unit for cooperativity. The subscript figures indicate the number of bound ligand molecules. Note that a concerted change between T- and R-state is implied. At low [L] the most likely path is $T_0 \rightarrow R_0 \rightarrow R_1 \rightarrow R_2$, however, at very high [L] the sequence $T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow R_2$ may be favoured if the transition between T and R is slower than binding of the ligand to T_0 and T_1. In an induced fit model, the path would go directly from T_0 to R_2

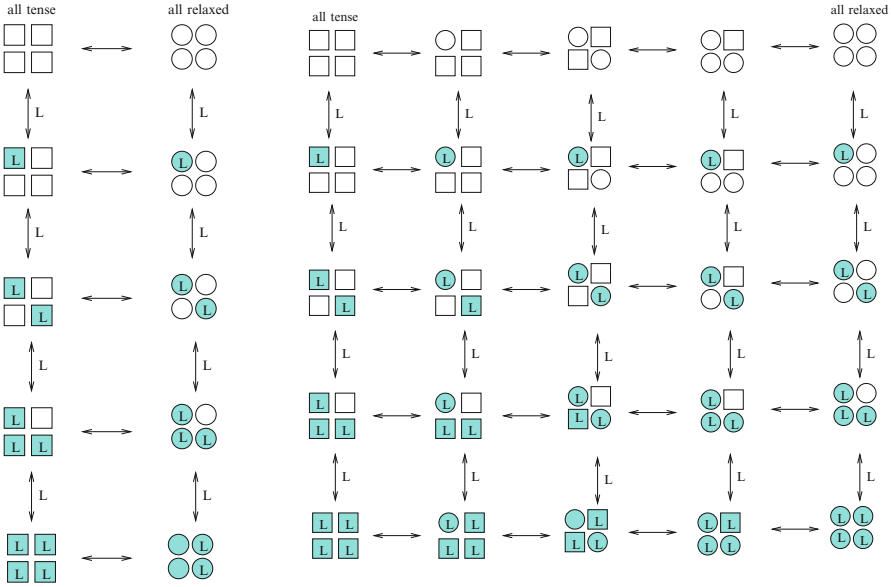
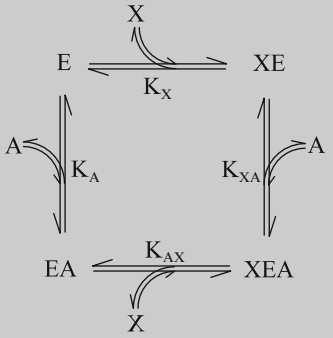


Fig. 7.5 Two possible models for cooperativity. *Left*: In the concerted model all subunits in a protein are either in the T- (*squares*) or in the R-state (*circles*). The probability of the protein being in the R-state increases as more and more substrate is bound (filled vs. empty symbols). *Right*: In the sequential model each subunit may individually switch between T- and R-states. The more subunits are in a given state, the higher the probability is for the remaining subunits to follow suit. Note that the concerted model may be viewed as the limiting case of the sequential. The initial binding of oxygen is to the α-subunits; binding to the β-subunits occurs only after T to R transition



The change in free energy of XEA-formation ΔG^{XEA} must be independent of whether A or X binds first, and binding constants and free energy are related. Thus any effect that A has on the binding of X must be reciprocated by an

(continued)

effect of X on the binding of A (principal of **reciprocity**). Thus we can define the allosteric coupling constant Q_{AX} as:

$$Q_{AX} = \frac{K_A}{K_{XA}} = \frac{K_X}{K_{AX}} \tag{7.2}$$

Q_{AX} compares how A binds to E in the absence and in the presence of saturating concentrations of X [5]. There are three possible cases:

$Q_{AX} > 1$ binding of the effector X increases the affinity of the protein E for the ligand A.
$Q_{AX} < 1$ binding of the effector X decreases the affinity of the protein E for the ligand A.
$Q_{AX} = 1$ binding of the effector X has no effect on the affinity of the protein E for the ligand A (no allostery).

Structurally, binding of both A and X results in conformational changes of E. However, allosteric-specific conformational changes are only those that require concurrent binding of both, or are prevented when both are present. For example, a steric clash between changes caused by binding of A and by binding of X would result in negative allostery ($Q_{AX} < 1$). In actual practice, there are usually multiple communication pathways between the substrate and effector site of a protein, which may involve different types of interactions, not only steric clashes.

7.2.2.1 Equation for Cooperative Binding

A. HILL proposed a mathematical model for cooperative binding in 1910 [9], long before the molecular causes of this phenomenon were discovered. He started with the assumption that several ligand molecules (L) bind to a protein (P):

$$P + nL \rightleftharpoons PL_n \tag{7.3}$$

Then the equations for the association constant K_a and binding Θ become:

$$K_d = \frac{[P] * [L]^n}{[PL_n]} \tag{7.4}$$

$$\Theta = \frac{[L]^n}{[L]^n + K_d} \tag{7.5}$$

The latter equation can be linearised by rearrangement followed by taking logarithms:

$$\log\left(\frac{\Theta}{1-\Theta}\right) = n * \log([L]) - \log(K_d) \tag{7.6}$$

This is called the HILL-equation, and a diagram of $\log(\frac{\Theta}{1-\Theta})$ vs $\log([L])$ is called HILL-plot. The function $f(x) = \log(\frac{x}{x_{\max}-x})$ is called logit(x). Note that from the above equations the half-saturation point $K_{0.5}$ is related to the K_d: $K_d = K_{0.5}^n$.

When HILL plotted the oxygen binding data for hæmoglobin in this way, he found that the line that best fitted the data had n=3, rather than the number of binding sites, which is, of course, 4. The number of binding sites and the Hill coefficient unfortunately both have the letter "n" as symbol. I have used different fonts to distinguish the two. It can even have (and frequently has) noninteger values.

0 $\leq n <$ **1** signifies **negative cooperativity**, where binding of a ligand to one site impedes the binding at other sites.

n**=1** means that there is no cooperation at all (binding follows hyperbolic kinetics).

1 $< n <$ **n** positive cooperativity, binding of one ligand to one site facilitates binding to the other sites.

n**=n** means complete cooperation (all sites are either filled or empty; no molecules with only some sites filled are allowed).

$n >$ **n** has never been observed, this situation is considered impossible.

7.2.3 Other Factors Involving Oxygen Affinity of Hæmoglobin

7.2.3.1 2,3-Bisphosphoglycerate as Heterotropic Allosteric Modulator

2,3-BPG binding reduces oxygen affinity of hæmoglobin [1], but has little effect on the HILL-coefficient. 2,3-BPG binding and release allows our bodies to adapt quickly to different environmental oxygen partial pressures. For example, if we climb onto a mountain, air pressure and therefore oxygen partial pressure is reduced. Under these conditions, 2,3-BPG concentration in blood will rise within a couple of days; this leads to a drop in the oxygen affinity of hæmoglobin, makes oxygen delivery to the tissues more efficient but has little effect on oxygen saturation in the lung (see Fig. 7.6).

> Hypoxia caused by lung damage or other pathological situations will also increase blood 2,3-BPG.

Each hæmoglobin tetramer has one binding site for 2,3-BPG, a pocket formed between the β-subunits in the T-state, lined with positive charges that can interact

with the negative charges of the phosphate groups of 2,3-BPG. In the R-state, this pocket is too small to accommodate 2,3-BPG, hence binding of 2,3-BPG into the pocket stabilises the T-state of hæmoglobin.

There is a direct medical application for this knowledge: erythrocytes from donated blood used to be stored in acid-citrate-dextrose medium ("ACD-blood"), where they are stable but lose much of their 2,3-BPG. As a result the hæmoglobin in transfused erythrocytes had an abnormal high affinity for oxygen and could not participate in the oxygen supply of tissue until the 2,3-BPG level had been restored (about 24–48 h after transfusion). This deprived the patients of most of the benefit from transfusions.

Today **hypoxanthine** is added to the storage medium, which is split to inosine and ribose in the erythrocytes. Ribose can be converted to 2,3-BPG, maintaining its level in the cells.

7.2.3.2 BOHR-Effect

In tissues oxygen is used to burn food to carbon dioxyde and water. The carbon dioxyde needs to be removed, and this too is done by hæmoglobin. As a first step, carbonic anhydrase converts CO_2 into the much more soluble bicarbonate ion:

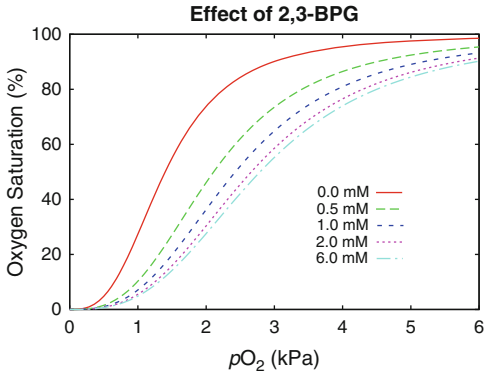


Fig. 7.6 *Left*: 2,3-BPG lowers the affinity of hæmoglobin for oxygen. Increased 2,3-BPG in blood is an adaptation to lower environmental pO_2, for example, in people living at high altitude or suffering from certain lung diseases. Plotted using data from [10]. *Right*: Indios running after the *Tren a las Nubes* near the viaduct La Polvorilla in Salta province, Argentina. The altitude is 4220 m above sea level. Whilst the passengers of the train are nearly incapacitated by altitude sickness, the natives run after the train from one station to the next, to continue selling their goods. The difference is mainly caused by different [BPG] in the blood of tourists and natives

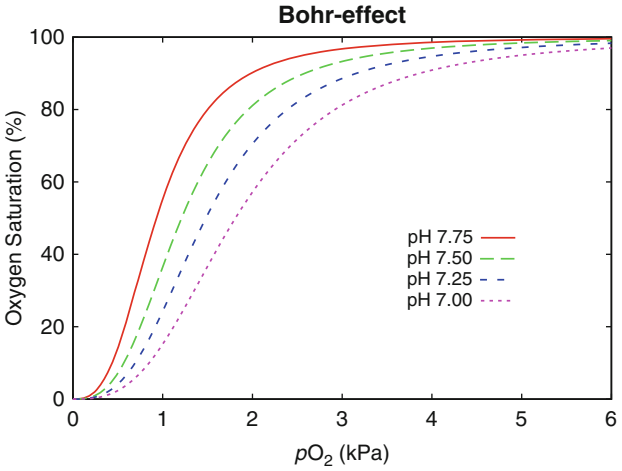


Fig. 7.7 BOHR-effect. Oxygen binding to hæmoglobin is *p*H -dependent; oxygen affinity is higher in the lungs than in muscle. Plotted using data from [10]

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^- \tag{7.7}$$

As protons are produced in this reaction, venous blood has a somewhat lower *p*H (7.2) than arterial (7.6), even though the effect is mitigated by the buffering capacity of blood. This is important, because a drop in *p*H lowers the affinity (but not the HILL-coefficient) of hæmoglobin for oxygen, and increases that for bicarbonate. This is called the BOHR**-effect**[1] [2]. Hæmoglobin actually releases one proton for each molecule of oxygen bound, thus decreasing *p*H interferes with oxygen binding (Fig. 7.7).

Molecular Mechanism of the BOHR-Effect
One of the most important binding sites for protons is His-146 of the β-subunit. When protonated, this His can form an ionic bond to Asp-94, stabilising the T-state of hæmoglobin. Because of that ionic interaction His-146 is stabilised in the protonated form, giving it a high pK_a in the T-state. However, several other amino acids in addition to His-146/Asp-94 are also responsible for the BOHR-effect.

CO_2 binds to the α-amino group of all four subunits, forming carbamino-hæmoglobin. In this reaction too protons are produced, increasing the BOHR-effect.

[1]Named after CHRISTIAN BOHR, physiologist and father of the physicist NIELS BOHR.

Additionally, this gives the amino termini a negative charge, making them available for salt bridge formation:

$$\text{Protein}-\text{NH}_2 + \text{CO}_2 \rightleftharpoons \text{Protein}-\text{NH}-\text{COO}^- + \text{H}^+$$

7.3 Hæmoglobin Related Diseases

7.3.1 Subunit Composition of Human Hæmoglobin

As already mentioned, normal adult hæmoglobin (HbA) consists of four subunits, two α- and two β-subunits. One α- and one β-subunit are held together tightly by ionic and hydrophobic bonds, forming a protomer. Two such protomers then form the hæmoglobin molecule. The protomers are held together mostly by hydrogen bonds. This allows the protomers to move with respect to each other during oxygen binding and release.

However, HbA is only one member of the hæmoglobin family. During embryonal development the fetus needs to get its oxygen supply from the mother. In order for this to work, its hæmoglobin has a different composition with a higher affinity for oxygen (Table 7.2).

During the first 2–4 weeks of gestation embryos produce **embryonal hæmoglobin** (Hb Gower 1) ($\zeta_2\epsilon_2$) in the yolk sac. This is followed by hæmoglobins Gower II ($\alpha_2\epsilon_2$) and Portland ($\zeta_2\gamma_2$). Oxygen supply of the embryo during this phase is from the amniotic fluid.

By week 14 of gestation the embryonal hæmoglobins have been replaced by the **Fetal hæmoglobin (HbF)** ($\alpha_2\gamma_2$). By this time the embryo has implanted and oxygen is supplied by the placenta. HbF is produced in the liver. The γ-subunit lacks some of the positive charges found in the β-subunit, resulting in weaker binding of 2,3-BPG and therefore higher oxygen affinity (oxygen affinity of HbF is similar to HbA striped of 2,3-BPG).

Production of hæmoglobin β-subunits starts at about the 6th month of gestation, and HbA will have completely replaced HbF about 3 months *post partum*.

Normal adult blood contains not only HbA, but also about 2 % HbA_2 and 3–9 % glycated HbA (HbA_{1c}). HbA_2 has the composition $\alpha_2\delta_2$. Its production starts about 3 months *post partum*. The physiological significance of its presence is not known.

HbA_1 is produced from HbA by a spontaneous, nonenzymatic reaction (see Fig. 7.8) between the N-terminal Val of the β-subunits with various compounds:

HbA_{1a} glucose phosphate + fructose phosphate

(continued)

HbA_{1b} pyruvate
HbA_{1c} glucose, $\approx 80\,\%$ of HbA_1

Of these, only HbA_{1c} has clinical importance. The rate of HbA_{1c}-formation, and therefore its accumulation during the 120 days life span of an erythrocyte and therefore its concentration in blood, is proportional to the blood glucose concentration. Thus the determination of $[HbA_{1c}]$ is valuable to check the long-term average blood sugar concentrations in **diabetics**. If the assay is performed according to the guidelines of the International Federation of Clinical Chemistry & Laboratory Medicine, then

$$[\bar{\text{gluk}}\text{ose}](\text{mM}) = 1.84 \times \text{HbA}_{1c}(\%) - 0.01$$

In humans the information for the hæmoglobin subunits α 1, α 2 and ζ is located on chromosome 16, for β, δ, γ and ϵ on chromosome 11. Each of these genes contains 3 coding segments (exons) and 2 noncoding segments (introns), which have to be spliced out of the mRNA before the proteins can be synthesised on the ribosomes. Synthesis of the hæmoglobin subunits is normally synchronised, so that equal amounts of α- and β-subunits are produced.

7.3.2 Inherited Diseases Relating to Hæmoglobin

Most countries have a mandatory screening program in which all newborns are tested for hæmoglobin variants, especially HbS. In the US, this identifies about 10.000 children per year with unidentified mutations in hæmoglobin. It is recommended that such children are reinvestigated at age 1 a. In addition, screening of the parents may help making a prognosis for the child; if one of them is also affected but healthy, then the mutation is unlikely to cause much harm (see table 7.3 for a list of more common hæmoglobin variants).

7.3.2.1 Sickle Cell Anæmia

Sickle cell anæmia is an autosomal recessive disorder (i.e., the gene needs to be inherited from both parents for the disease to become manifest) which affects mostly people from Africa and their descendants. It is caused by a Glu6Val mutation in the gene for the β-subunit (β^S). The resulting hæmoglobin is called HbS. The substitution of a charged amino acid with a hydrophobic one creates a hydrophobic patch, which is exposed in deoxy-hæmoglobin, but hidden when oxygen is bound.

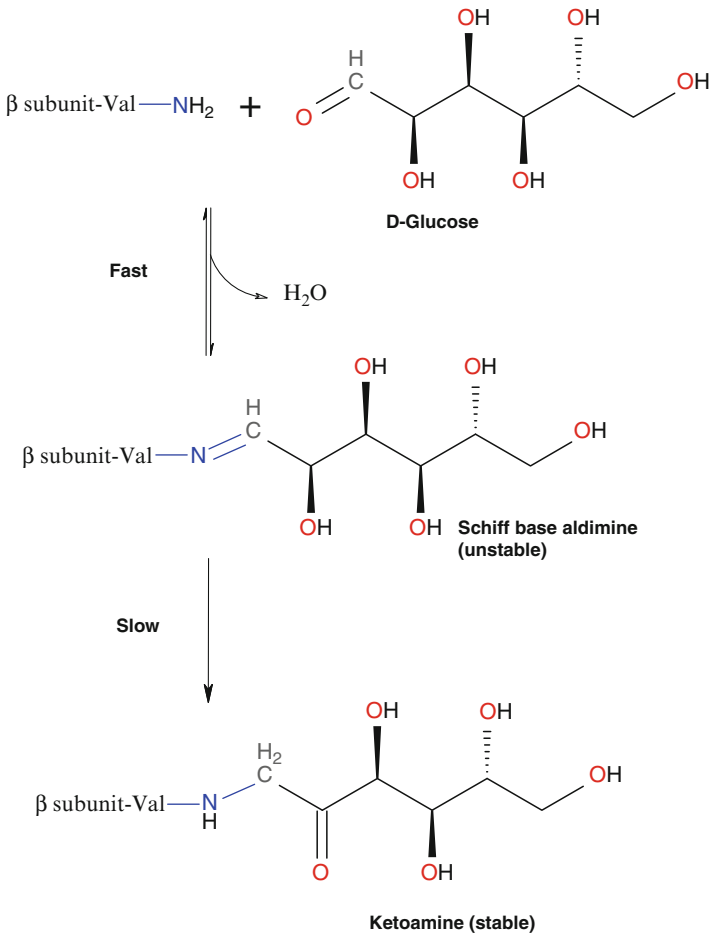


Fig. 7.8 Formation of glucated hæmoglobin (HbA_{1c}). The aldehyde group of glucose reacts with the N-terminal amino group of the β-subunit of hæmoglobin, forming an unstable SCHIFF-base. This intermediate undergoes AMADORI rearrangement to a stable ketosamine. Unlike glycoproteins, glucated proteins are formed spontaneously, without enzymes

[Hæmoglobin] in erythrocytes is quite high (35 % w/v), and if a large proportion of the molecules have such a hydrophobic patch, they aggregate, forming lumps in the erythrocytes. The erythrocytes take on an unusual shape, which gives the disease its name.

As a consequence, the lifetime of the erythrocytes in sickle cell disease is reduced from 120 to 20 days, leading to **anæmia** (lowered erythrocyte concentration in blood). This happens in particular when less oxygen is bound to the hæmoglobin (exertion, pregnancy, high altitudes, increased CO_2-concentration, or decreased blood *p*H). Sickled erythrocytes aggregate in the capillaries, blocking them and

Table 7.2 Normal hæmoglobin isoforms. Kinetic data from [10]

Name	Stage	Composition	p50 (kPa)	Hill coeff.	K_d BPG (mM)
Hb Gower 1	Embryo	$\zeta_2\epsilon_2$	0.47	2.3	0.5
Hb Gower 2	Embryo	$\alpha_2\epsilon_2$	0.73	2.4	0.5
Hb Portland	Embryo	$\zeta_2\gamma_2$	0.79	2.3	6.0
HbF	Fetus	$\alpha_2\gamma_2$			
HbA	Adult	$\alpha_2\beta_2$	1.31	2.9	0.45
HbA_2	Adult	$\alpha_2\delta_2$			
Hb Bart	In α-thalassæmia	γ_4			
Hb H	In α-thalassæmia	β_4			

causing pain and tissue death from oxygen deprivation (**infarct**). Susceptibility for infections is increased due to low tissue oxygen; patients receive prophylactic penicillin against pneumococci. Clinical aspects of sickle cell disease have been described in [13]. Some countries with a high incidence of HbS have mandatory screening programs for newborns; this includes most states in the USA. Because the Glu6Val mutation results in a changed isoelectric point of the β-chain, IEF or cellulose acetate electrophoresis may be used to identify both carriers and affected infants. In addition, vanishing of the 5'-CCTxAGG-3'-site cut by *Mst*II restriction DNase in the HbS-gene leads to a restriction fragment length polymorphism (RFLP).

It has been found that heterozygous individuals (one normal and one β^S gene), who have both HbA and HbS in their erythrocytes, are less prone to infection with *Plasmodium falciparum*, the causative agent for **malaria**. This parasite must spend part of its life cycle in erythrocytes, feeding on hæmoglobin. Because the presence of HbS decreases the erythrocyte life span (even in heterozygotes, although less than in homozygotes), the parasite cannot complete its life cycle in such individuals, giving resistance. In addition, parasites produce acids as metabolic products, the resulting BOHR-effect reduces the oxygen affinity of hæmoglobin. Thus in infected erythrocytes oxygen saturation of hæmoglobin is lower; they sickle more readily than noninfected cells and can be selectively removed by the spleen.

7.3.2.2 Hæmoglobin C Disease

The cause of HbC-disease is quite similar to HbS-disease, but Glu-6 is replaced by Lys rather than Val. In other words, a negatively charged amino acid is replaced by a positive one. Homozygous patients will show mild anæmia, but do not suffer from hæmolytic crises and usually do not require treatment.

Table 7.3 Hæmoglobin variants

Name	Chain	Mutation	Functional consequences (homozygote)
Hb Torino	α	42→Val	Hæmolysis improved by splenectomy
Hb Ferrara	α	47→Gly	Mild, intermittent hæmolysis
Hb Hasharon	α	47→His	Mild, intermittent hæmolysis
HbM Boston	α	58→Tyr	Cyanosis, methæmoglobinæmia
Hb Ann Arbor	α	80→Arg	Hæmolysis improved by splenectomy
Hb Etobicoke	α	84→Arg	Benign
HbM Iwate	α	87→Tyr	Cyanosis, methæmoglobinæmia
Hb Chesapeake	α	92→Leu	Increased oxygen affinity, polycythæmia
HbJ Capetown	α	92→Gln	Increased oxygen affinity, polycythæmia
Hb Titusville	α	94→Asn	Reduced oxygen affinity, cyanosis
Hb Dakar	α	112→Glu	Benign
Hb Bibba	α	136→Pro	Hæmolysis not improved by splenectomy, HEINZ bodies
HbS	β	6→Val	Sickle cell anæmia
HbC	β	6→Lys	Mild anæmia
Hb Leiden	β	6del Glu	Mild, intermittent hæmolysis
Hb Sogn	β	14→Arg	Benign
Hb Freiburg	β	23del Val	Mild, intermittent hæmolysis
Hb Genova	β	28→Pro	Hæmolysis improved by splenectomy
Hb Tacoma	β	30→Ser	Benign
Hb Hammersmith	β	42→Ser	Hæmolysis not improved by splenectomy, HEINZ bodies
Hb Louisville	β	42→Leu	Mild, intermittent hæmolysis
HbM Saskatoon	β	63→Tyr	Cyanosis, methæmoglobinæmia
HbM Osaka	β	63→Tyr	Cyanosis, methæmoglobinæmia
HbM Saskatoon	β	63→Tyr	Cyanosis, methæmoglobinæmia
Hb Zurich	β	63→Arg	Mild, intermittent hæmolysis
Hb Bristol	β	67→Asp	Hæmolysis not improved by splenectomy, HEINZ bodies
Hb Shepherd's Bush	β	74→Asp	Hæmolysis improved by splenectomy
Hb Seattle	β	76→Glu	Mild, intermittent hæmolysis
Hb Providence	β	82→Asp,Asn	Reduced oxygen affinity, cyanosis
Hb Agenogi	β	90→Lys	Reduced oxygen affinity, cyanosis
HbM Hyde Park	β	92→Tyr	Cyanosis, methæmoglobinæmia
Hb Gun Hill	β	del 91–97	Mild, intermittent hæmolysis
Hb Malmo	β	97→Gln	Increased oxygen affinity, polycythæmia
Hb Yakima	β	97→Gln	Increased oxygen affinity, polycythæmia
Hb Kemp	β	99→Asn	Increased oxygen affinity, polycythæmia
Hb Beth Israel	β	102→Ser	Reduced oxygen affinity, cyanosis
Hb Kansas	β	102→Thr	Reduced oxygen affinity, cyanosis

Table 7.3 (continued)

Name	Chain	Mutation	Functional consequences (homozygote)
Hb Ypsilanti	β	108→Asp	Reduced oxygen affinity, cyanosis
HbD Punjab	β	121→Gln	No anæmia
Hb Wien	β	130→Asp	Hæmolysis improved by splenectomy
Hb Olmsted	β	141→Arg	Hæmolysis not improved by splenectomy, HEINZ bodies
Hb Hiroshima	β	143→Asp	Increased oxygen affinity, polycythæmia
Hb Bethesda	β	145→His	Increased oxygen affinity, polycythæmia
Hb Rainier	β	145→Cys	Increased oxygen affinity, polycythæmia
HbE	β	261→Lys	Mild microcytic anæmia

7.3.2.3 Hæmoglobin SC Disease

This disease occurs when a patient inherits the HbS mutation from one and the HbC mutation from the other parent. The course of the disease is much milder than in HbS disease, but a (potentially fatal) hæmolytic crisis may occur in special situations (*e.g.,* during childbirth or surgery).

7.3.2.4 Thalassæmias

As mentioned before, the synthesis of the α- and β-subunits of hæmoglobin is normally synchronised, so that equal amounts are produced. Thalassæmias are caused when this synchronisation no longer works and either not enough α-subunit (α-thalassæmia) or not enough β-subunit (β-thalassæmia) is produced. Production of this subunit can be **absent** (α^0- and β^0-thalassæmias) or **reduced** (α^+- and β^+-thalassæmias). Irrespective of the gene affected, patients with severe thalassæmia suffer from microcytic, hypochromic anæmia, bone marrow hyperplasia leading to thicker bones, splenomegaly that may require removal of the spleen, increased sensitivity to infection (low tissue O_2, spleen damage), growth retardation, heart problems, and, if untreated, early death. As these patients require regular blood transfusion, iron accumulates and results in hæmochromatosis. Thalassæmias are most common in malaria-infested areas due to heterozygote advantage.

Because people carry the information for the α-subunit in four copies (two on each of the chromosomes 16), it can occur in a variety of degrees. An individual with one defect copy is called a **silent carrier** and will not show overt symptoms. Individuals with two defect copies are said to have the **thalassæmia trait** and with three defect copies **hæmoglobin H disease** (mild to moderate anæmia). Individuals with four defect copies will die during embryonal development, when the embryonal hæmoglobin is replaced by the fetal one. The abortus will have high

levels of γ-tetrameres (Hb Bart) or β-tetrameres (HbH), which are soluble but show hyperbolic oxygen binding curves with no cooperativity. As discussed above, they are useless as oxygen carriers. **α-thalassæmia** is most commonly the result of large deletions from the gene.

In **β-thalassæmia** the α-subunits are produced at a normal level, but β-subunit production is reduced, most commonly due to splicing defects. In the absence of the β-subunits normal hæmoglobin cannot form. Instead the α-subunits precipitate as monomers on the membranes of erythrocyte precursors, which die as a consequence. As there are only two copies of the β-globin gene, affected individuals either have the thalassæmia trait (**β-thalassæmia minor**, no treatment required) or **β-thalassæmia major** (COOLEY anæmia, with both copies defect). The latter individuals appear perfectly healthy at birth, where the fetal hæmoglobin is still present, but become rapidly anæmic as HbF should be replaced by HbA. Originally these patients were treated with regular blood transfusions, but the resulting iron overload resulted in death from **hæmochromatosis** (iron poisoning) in their mid-twenties. Chelation therapy (see Fig. 7.9) can reduce this problem, but is burdensome to the patient resulting in poor compliance. Today **bone marrow transplantation** offers a better prognosis.

δ-thalassæmia results in the inability (δ^0) or reduced ability (δ^+) to form HbA_2. This is clinically unobtrusive, as only 2–3 % of hæmoglobin is HbA_2. However, diagnosis of β-thalassæmia is sometimes made by the increased presence of HbA_2; if the patient also is δ^+ this assay would give a false-negative result [3].

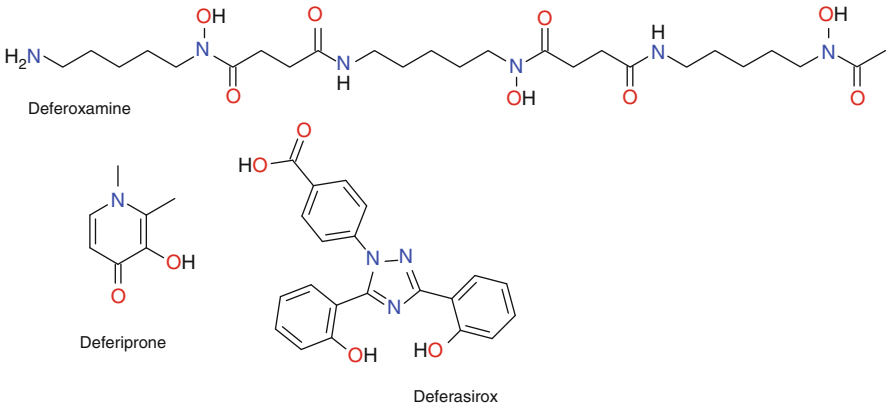


Fig. 7.9 Pharmaceuticals used as Fe-chelators in patients receiving frequent blood transfusion. Each unit of blood contains 250 mg Fe. Iron is not excreted and in patients with thalassæmias this iron cannot be used. This leads to chronic iron poisoning (hæmochromatosis) with destruction of heart, liver, and pancreas. Although chelation therapy can save the lives of these patients, all these drugs have significant side effects

7.4 Exercises

7.4.1 Problems

7.1. J.F. ENGELHARD found in 1825 that hæmoglobin of all species investigated contained about 0.5 % iron. What is the molecular mass of hæmoglobin?

7.2. Oxygen binding to hemoglobin (in % of maximum) is a sigmoidal function of oxygen concentration (in kPa partial pressure). Assuming a HILL coefficient $n = 2.9$ and $K_{0.5} = 13$ kPa, what is the saturation of hemoglobin in working muscle, where the oxygen concentration is 1 kPa?

A) 0 %
B) 7 %
C) 20 %
D) 50 %
E) 100 %

7.4.2 Solutions

7.1 Iron has a molecular mass of 55.8 Da. By the rule of three, if 55.8 Da is equivalent to 0.5 %, then 100 % is equivalent to $\frac{100\,\% \times 55.8\,\text{Da}}{0.5\,\%} = 11\,200\,\text{Da} = 11.2\,\text{kDa}$ per atom of iron bound; ENGELHARD had no way of determining how many iron atoms are in one molecule of hæmoglobin. Determination of the molecular mass of hæmoglobin by modern methods yields 4×16.7 kDa.

7.1

$$\begin{aligned}\theta &= \frac{\theta_{\text{max}} \times [O_2]^n}{K_{0.5} + [O_2]^n} \\ &= \frac{100\% \times 1^{2.9}}{13 + 1^{2.9}} \\ &= \frac{100\%}{14} \\ &\approx 7\%\end{aligned}$$

References

1. R. Benesch, R.E. Benesch, Intracellular organic phosphates as regulators of oxygen release by haemoglobin. Nature **221**, 618–622 (1969). doi:10.1038/221618a0
2. C. Bohr, Absorptionscoëfficienten des Blutes und des Blutplasmas für Gase. Skand. Arch. Physiol. **17**, 104 (1905). doi:10.1111/j.1748-1716.1905.tb00547.x

3. M.J. Bouva, C.L. Harteveld, P. van Delft, P.C. Giordano, Known and new delta globin gene mutations and their diagnostic significance. Haematologica **91**(1), 129–132 (2006). URL http://www.haematologica.org/content/91/1/129.long
4. J.-P. Changeux, S. Edelstein, Conformational selection or induced fit? 50 years of debate resolved. F1000 Biol. Rep. **3**(19), (2011). doi:10.3410/B3-19
5. A.W. Fenton, Allostery: an illustrated definition for the 'second secret of life'. TIBS **33**(9), 420–425 (2008). doi:10.1016/j.tibs.2008.05.009
6. P.R. Gardner, Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases. J. Inorg. Biochem. **99**(1), 247–266 (2005). doi:10.1016/j.jinorgbio.2004.10.003
7. P.R. Gardner, A.M. Gardner, L.A. Martin, A.L. Salzman, Nitric oxide dioxygenase: An enzymic function for flavohemoglobin. Proc. Natl. Acad. Sci. USA **95**(18), 10378–10383 (1998). URL http://www.pnas.org/content/95/18/10378.full.pdf+html
8. M.T. Gladwin, J.H. Crawford, R.P. Patel, The biochemistry of nitric oxide, nitrite, and hemoglobin: role in blood flow regulation. Free Radical Biol. Med. **36**(6), 707–717 (2004). doi:10.1016/j.freeradbiomed.2003.11.032
9. A.V. Hill, The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. J. Physiol. **40**, 190 (1910)
10. O. Hofmann, R. Mould, T. Brittain, Allosteric modulation of oxygen binding to the three human embryonic hæmoglobins. Biochem. J. **306**, 367–370 (1995). URL http://www.biochemj.org/bj/306/0367/3060367.pdf
11. D.E. Koshland, G. Nemethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry **4**, 365–385 (1966). doi:10.1021/bi00865a047. URL http://www.chem.umass.edu/~rmweis/Chem728/papers/P15_Koshland_et_al_Biochemistry_1966.pdf
12. J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model. J. Mol. Biol. **12**, 88–118 (1965). doi:10.1016/S0022-2836(65)80285-6
13. G.R. Serjeant, B.E. Serjeant, *Sickle Cell Disease*, 3rd edn. (Oxford University Press, Oxford, 2001). ISBN 978-0-1926-3036-0

Chapter 8
Enzyme Kinetics: Special Cases

Abstract Both hyperbolic and sigmoidal kinetics result in a gradual change of enzymatic activity with substrate concentration. Activation cascades, where upon a signal one enzyme activates another, which in turn activates yet another, allow switch-like, all-or-nothing responses. A cell either enters mitosis, or it doesn't. A liver cell either does glycolysis or it does gluconeogenesis. This behaviour more than signal amplification is the physiological role of such cascades.

In an enzyme following HMM-kinetics, enzyme activity rises gradually with substrate concentrations. For many physiological situations this is appropriate; the enzyme activity adapts to the metabolic needs of the cell. Indeed, the K_m of many, if not most, enzymes has evolved to be near the physiological concentrations of their substrates.

In this chapter we look at some cases where such a gradual response would be inadequate, where instead a steep, switch-like, **all-or-nothing response** is required.

Binding of a few hormone molecules to the receptors on the membrane of a cell can affect considerable changes in the metabolism of the cell. For example, binding of **mitogens** (growth factors) can make a cell enter mitosis (see Fig. 8.1). A cell, however, should be either in mitosis, or not. A gradual transition between these two states is impossible. The switching between **mutually exclusive metabolic states** by hormones (*e.g.*, **insulin** and **glucagon**) would be another example.

You have already learned how the association of subunits into oligomeric enzymes with cooperating binding sites can produce S-shaped control curves, which are steeper than hyperbolic curves. However, even steeper curves can be obtained by activation cascades.

8.1 Activation Cascades

The usual explanation given for such cascades is that they allow amplification of the signal. If each enzyme activated 100 substrates, than the Receptor/G-protein/MAPKKK/MAPKK/MAPK/MAP cascade in Fig. 8.2 would allow amplification by a factor of $100^5 = 10^{10}$. This explanation is incorrect on two grounds:

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_8

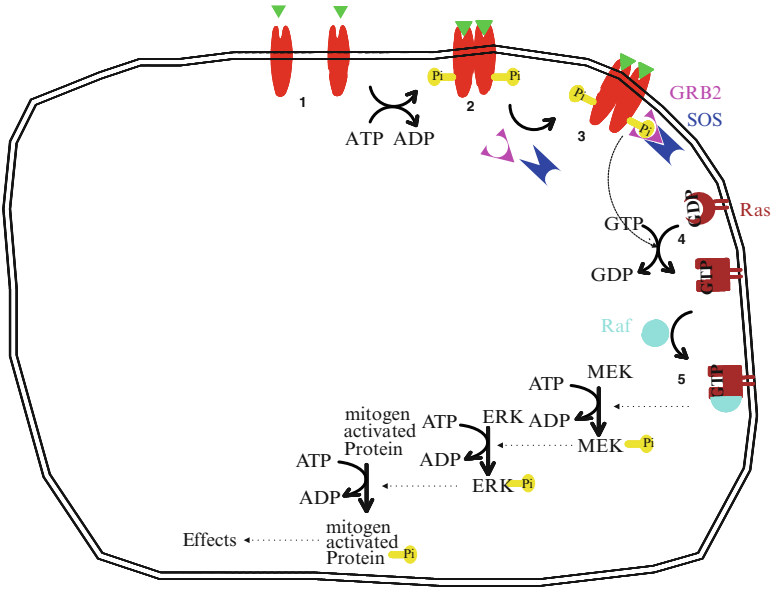


Fig. 8.1 Binding of mitogens results in the phosphorylation of so-called **mitogen activated proteins (MAP)**, which control entry of the cell into mitosis [1]. Mitogens bind to cell surface receptors (1), which dimerise and auto-phosphorylate each other on tyrosine residues (2). These phospho-tyrosine residues serve as binding sites for GRB2 and SOS (3). Bound SOS acts as a GDP/GTP exchange factor for the small G-protein Ras, converting the inactive GDP-Ras into active GTP-Ras (4). This in turn binds and activates Raf, which is a Ser/Thr kinase (**MAPKKK**) (5). Activated Raf will phosphorylate and thus activate MAP-kinase-kinases (**MAPKK**) like MEK, which will in turn phosphorylate MAP-kinases (**MAPK**) like ERK, which act on MAPs. Not shown are the GTP-hydrolysis stimulating factors that turn off Ras and the protein phosphatases that remove the activating phosphate groups from the other proteins. For more details see textbooks on molecular cell biology

- Such amplification factors are unmanageable by the cell and would lead to constant erroneous activation of MAPs due to noise.
- Experimentally determined amplification factors for this system are on the order of 10^1–10^2. Such factors could be achieved also by a single enzyme.

The reason for cascades is probably not amplification, but the shape of the resulting **response curves**. A cell should be either stimulated into mitosis, or not. Thus rather than the usual hyperbolic response curves something steeper, like an S-shaped HILL-curve, is required.

Assume we have a phosphorylatable substrate in the presence of a kinase and a phosphatase. The left column of the Fig. 8.2 shows this situation for low substrate concentrations, where the enzyme activity is **linearly** related to [S]; the right column shows high concentrations with **hyperbolic** curves.

The curves were drawn under the assumption that the total substrate concentration (sum of phosphorylated and dephosphorylated substrate) does not change over

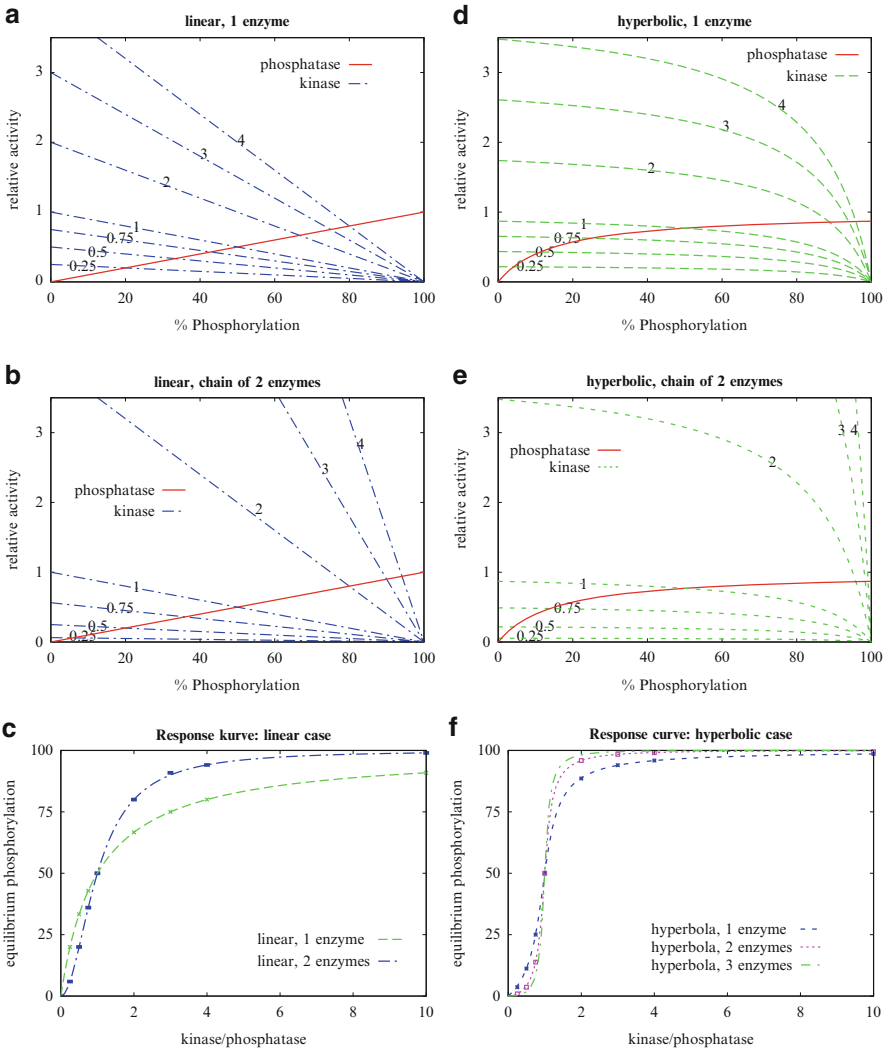


Fig. 8.2 Analysis of the response curves for enzyme cascades. For details see text

time. Only the fraction of substrate phosphorylated changes. Under this condition if all the substrate is dephosphorylated, the kinases will be maximally active, whilst the phosphatases have no substrate to work on and therefore their activity will be zero. In the case of fully phosphorylated substrate the opposite is true.

It is easy to see that if phosphatase and kinase are present at the same activity, an **equilibrium** will be established where half the substrate is phosphorylated. At this x-coordinate the lines of phosphatase and kinase activity in Fig. 8.2a intersect.

If we keep the phosphatase activity constant but change the kinase activity (say, after hormone binding), the equilibrium between phosphorylated and dephosphorylated substrate, and as a result the intersection between the lines denoting phosphatase and kinase activity will also change.

Now look at Fig. 8.2e. It shows how the equilibrium phosphorylation changes as a function of the ratio of kinase/phosphatase activity. If the kinase activity is regulated directly by some input stimulus we get a **hyperbolic response curve**.

Now let us assume that our kinase is regulated by a cascade: the stimulus acts on an enzyme that stimulates the kinase with an amplification factor of two. Under this condition, a doubling of the input will lead to a fourfold increase of kinase activity (plotted in the middle row of Fig. 8.2). The resulting response curve is S-shaped, and it is the steeper the higher the number of enzymes in the cascade is.

In case of a linear v versus [S] relationship (left column of Fig. 8.2) the result will be a HILL-curve. If the v versus [S] relationship is hyperbolic, the response curve will also be S-shaped, but steeper than predicted by the HILL-equation.

For further discussion of this topic see [2].

8.2 Feedback-Networks

Many events of metaphase are controlled by **maturation promoting factor (MPF)** (see, e.g., [4] for a detailed discussion on MPF). MPF is a complex of two proteins, the Ser/Thr-kinase **Cyclin-dependent kinase (Cdk)1** and its control protein **Cyclin B**. Cdk1 can be phosphorylated on two sites: Tyr15 and Thr161. Phosphorylation on Thr161 is required for activity; the enzymes involved are not shown here for reason of clarity. Phosphorylation on Tyr15 inhibits the kinase activity. This phosphorylation is caused by a protein kinase called **Wee1** (Scottish wee = tiny, as yeast cells with mutated Wee1 divide early and never reach full size); the dephosphorylation is accomplished by the phosphatase **Cdc25**.

The point of this system is that both Wee1 and Cdc25 can be phosphorylated by MPF (or kinases activated by MPF). Wee1 becomes inactivated, and Cdc25 activated by phosphorylation. Thus the presence of a small amount of active MPF will, by phosphorylating these two proteins, stimulate its own production. This is called **positive feed-back**.

MPF-activity at the end of the cell cycle is destroyed by ubiquitin-dependent proteolysis of Cyclin B (see Fig. 11.19 on page 260). The ubiquitinating enzymes, however, are also activated by MPF; that is, high MPF activity results in its own destruction. This is called **negative feed-back**.

Positive and negative feedback together explain the sawtooth-like activity pattern of MPF, as can be shown in computer simulations.

Simulation of Feedback Systems

What are the properties of a feedback system?

It is frequently observed that equations for the rate of change of a parameter (i.e., a differential equation) can be written down much more easily than the integrated equation for the parameter itself; we have already encountered such systems of differential equations when we derived the HMM-equation (see page 115).

For example, the change in concentration of MPF in Fig. 8.3 is given by its formation from preMPF ($v_1 = [preMPF] * k_{\text{Cdc25}}$) and its destruction by Wee1 and UbE ($v_2 = -[MPF] * k_{\text{Wee1}} - [MPF] * k_{\text{UbE}}$). Thus

$$\frac{d[MPF]}{dt} = [preMPF] * k_{\text{Cdc25}} - [MPF] * k_{\text{Wee1}} - [MPF] * k_{\text{UbE}} \tag{8.1}$$

Similar equations can be written for all other participants in this system (whether shown in Fig. 8.3 or not). This results in a system of linked differential equations.

However, we are not really interested in the rate of change of the activities over time, but in the activities themselves. To get those, the system of differential equations needs to be integrated. With small systems, this can sometimes be done analytically, but systems as complex as MPF-regulation can be integrated only numerically on a computer.

$$\frac{dx}{dt} = k * x \approx \frac{\Delta x}{\Delta t} \tag{8.2}$$

$$x(t + \Delta t) \approx x(t) + \frac{\Delta x}{\Delta t} \tag{8.3}$$

In other words, if you know the concentrations of all components in a system at time t =0, you can calculate the concentrations after a small time increment ($t_1 = t_0 + \Delta t$) by calculating the change over the time period and adding this to the original value. The results can then be used to calculate concentrations at t_2 and so on. This is called EULER'S algorithm.

There is an error associated with this procedure, because it assumes that the rate of change is constant during the period Δt. The error can be minimised by using the average rate of change during this period, rather than the rate of change at its beginning. This is the principle of the RUNGE-KUTTA algorithm (a good implementation of this method in C is found in [3]). Solving systems of differential equations in this way is central to areas as diverse as enzyme kinetics, weather prediction, or airplane construction.

If these principles are applied to cell cycle control enzymes, the results mimic the laboratory observations closely; for example, sawtooth patterns of MPF-activity can be seen. For further details see [6].

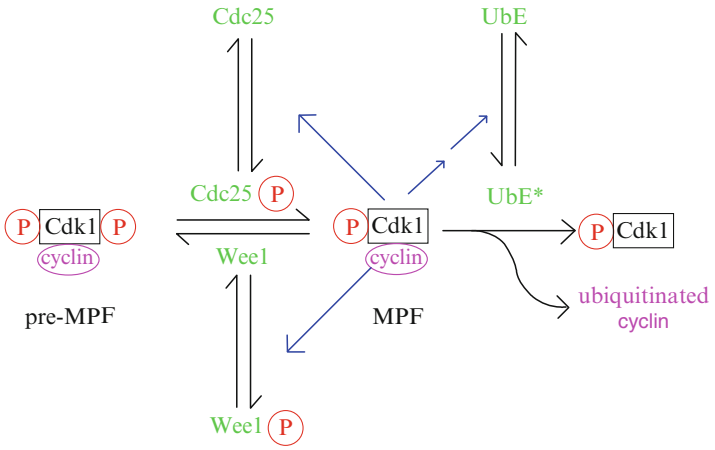


Fig. 8.3 Part of the feedback control of MPF-activity. For details see text

8.3 Multiple Phosphorylation

Another method used by cells to control response curves is exemplified by the *S. cerevisiae* Sic1 protein. This protein binds to the yeast S-phase cyclin/Cdk1 and inactivates it, thereby preventing DNA replication. At the G_1/S-transition Sic1 is phosphorylated (by another Cyclin/Cdk-complex), leaves the complex, and becomes ubiquitinated. At least 6 of the 9 phosphorylation sites of Sic1 need to be phosphorylated for this to happen.

Why 9, wouldn't one do? Ubiquitination starts by the binding of phosphorylated Sic1 to Cdc4, which is part of the ubiquitination machine. Other proteins can bind to Cdc4 after phosphorylation of only a single site. However, none of the phosphorylation site sequences in Sic1 matches the binding site of Cdc4 very closely. Thus multiple phosphorylation is required to increase the binding strength.

But why then does Sic1 not have a sequence that matches the binding site on Cdc4? By site-directed mutagenesis this question can be studied experimentally [5]. In short, such a mutated Sic1 does not work.

Assuming that each phosphorylation is accomplished by a single productive collision between Sic1 and its kinase, the first 5 events introduce a delay, before finally the 6th event marks Sic1 for destruction. Thus the requirement for multiple phosphorylation acts as a kind of timed fuse.

In addition, the first 5 sites act as a noise filter. Random phosphorylation by unspecific kinases may happen on one site, maybe even two. But all 6—that is unlikely. If the rate of unspecific phosphorylation is ϵ, the requirement for 6 phosphorylations reduces unwanted Sic1 destruction by a factor of ϵ^6. This is called **kinetic proof-reading**.

The same is true for phosphorylation by low levels of kinase. Only when the concentration of kinase has risen to a significant level will Sic1 be destroyed. A doubling of kinase concentration could then increase the rate of Sic1 destruction by $2^6 = 64$-fold, resulting in a step-like inactivation.

References

1. M. Atanasova, A. Whitty, Understanding cytokine and growth factor receptor activation mechanisms. Crit. Rev. Biochem. Mol. Biol. **47**(6), 502–530 (2012). doi:10.3109/10409238.2012.729561
2. J.E. Ferrell, Jr., Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem. Sci. **21**(12), 460–465 (1996). doi:10.1016/S0968-0004(96)20026-X
3. G. Heinzel, Beliebig genau: Moderne Runge-Kutta Verfahren zur Lösung von Differentialgleichungen. c't **8**(8), 172–185 (1992)
4. H. Lodish et al., *Molecular Cell Biology*, 7th edn. (W.H. Freeman and Company, New York, 2012). ISBN 978-1-4292-3413-9
5. P. Nash et al., Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. Nature **414**, 514–521 (2001). doi:10.1038/35107009
6. J.J. Tyson, B. Novak, G.M. Odell, K. Chen, C.D. Thron, Chemical kinetic theory: understanding cell-cycle regulation. Trends Biochem. Sci. **21**(3), 89–96 (1996). doi:10.1016/S0968-0004(96)10011-6

Chapter 9
The Flow of Metabolites Through Metabolic Pathways

Abstract All enzymes in a metabolic pathway operating in steady state have the same flux and maintain concentrations of all intermediates. The system will respond to perturbation in such a way that the perturbation is counteracted. This results in homeostasis. Cells can change the flux through a pathway by changing enzyme concentration, switching enzymes on or off by posttranslational modification, allosteric control, and catalytic cycling.

For each reaction in a pathway, the flux J of metabolites through the enzyme is equal to the velocity of the forward reaction v_f minus the velocity of the reverse reaction v_r

$$J = v_f - v_r \tag{9.1}$$

The flux through a pathway is controlled such that the organism's need for products is met. Any change in demand is passed through the entire pathway to maintain a steady state; that is, the metabolite concentrations along the pathway remain more or less constant.

We have seen that endproduct inhibition is an important method to control the flux through a pathway: as the concentration of the end product of a pathway increases, it allosterically shuts down the rate-limiting enzyme of that pathway. This is usually the enzyme catalysing the first committed step of that pathway (see Fig. 9.1), which:

- has a significant negative ΔG and is therefore expensive to reverse.
- occurs after all branch points leading to other products, so that end product inhibition by product H will not shut down production of another product X.

Due to this inhibition the concentrations of all metabolites before the regulated enzyme will increase; metabolites after the regulated enzyme will decrease. An increase of activity would have the opposite effects. This is called the **crossover theorem** [1].

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_9

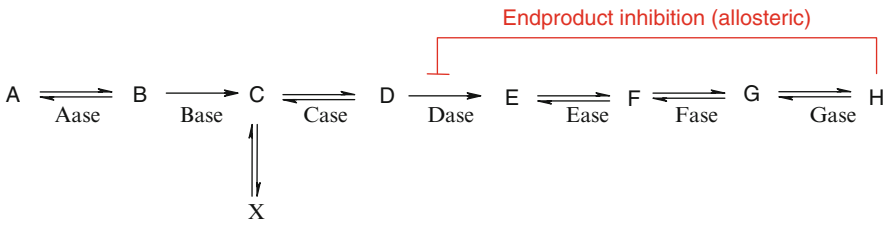


Fig. 9.1 Feedback-inhibition. The first committed step in the pathway leading from A to H is the Dase reaction. If Base were regulated, both the production of H and X would be affected. There is also no point in regulating Case as its reaction is reversible: D can be turned back into C without expenditure of metabolic energy. Turning E back into D—if possible at all—would require metabolic energy to bypass Dase (e.g., hexokinase/glucose-6-phosphatase)

9.1 Flux Control Theory

HEINRICH & RAPOPORT [2], and independent of them KACSER & BURNS [3], developed a more general approach to the problem. Each pathway is characterized by parameters such as equilibrium constants, MICHAELIS-constants or turnover numbers, which are determined outside the system. On the other hand, there are variables, most importantly the concentrations of metabolites, on which the system settles. The question then becomes: how does the system—that is, the metabolite pools—respond to a change of parameters? Which of them are the most important? To make this a quantitative question, we assume that a parameter P—say, the concentration of an inhibitor—changes. We measure the change in flux J. To get rid of units, we use scaled variables, dJ/J and dP/P. Then we get

$$\frac{dJ/J}{dP/P} = \frac{d\ln(J)}{d\ln(P)} = R_P^J \tag{9.2}$$

Recall that $dx/x = d\ln(x)$. R_P^J is called the **response coefficient**. It is the slope of the curve of J versus P *at the state the system is in*. In practice, we can't make infinitesimal changes and would measure the response δJ as function of a *small* change δP.

The response coefficient can be thought to consist of two parts:

Controllability coefficient $\pi_P^J = d\ln(v)/d\ln(P)$ is a property of the isolated enzyme. The change of enzymatic velocity v is determined by the type of response (hyperbolic vs. sigmoidal) and the K_i, as we have discussed in previous chapters.

Flux control coefficient C_P^J is a property of the whole system. Any change in the velocity of one enzyme in a pathway will affect the concentration of all metabolites, which will settle to new equilibrium values. This in turn will affect the overall flux through the pathway.

These coefficients are linked by $R_P^J = \pi_P^J \times C_P^J$.

Because the inhibition of one enzyme will increase the concentration of metabolites in front of it, and decrease those behind it (the crossover theorem), C_P^J is usually smaller than π_P^J. We recall that for a single reaction LE CHATELIER's principle states that a perturbation will result in a change of equilibrium that counteracts the perturbation. We now see that pathways in dynamic equilibrium have a similar property. Consider a situation $\mathsf{A} \xrightarrow{\mathsf{J}} \mathsf{B} \underset{v_r}{\overset{v_f}{\rightleftharpoons}} \mathsf{C} \xrightarrow{\mathsf{J}} \mathsf{P}$. If J through the rate-limiting Aase changes by ΔJ, then for Base:

$$J = v_f - v_r \tag{9.3}$$

$$\Delta J = \Delta v_f \tag{9.4}$$

$$\frac{\Delta J}{J} = \frac{\Delta v_f}{v_f}\frac{v_f}{J} = \frac{\Delta v_f}{v_f}\frac{v_f}{(v_f - v_r)} \tag{9.5}$$

$$v_f = \frac{V^f_{\max}[\mathsf{B}]}{K_{\mathrm{m}} + [\mathsf{B}]} \tag{9.6}$$

Because in most metabolic situations $[\mathsf{B}] < K_{\mathrm{m}}$ this simplifies to approximately $\frac{V^f_{\max}[\mathsf{B}]}{K_{\mathrm{m}}}$. Then

$$\Delta v_f = \frac{V^f_{\max}\Delta[\mathsf{B}]}{K_{\mathrm{m}}} \tag{9.7}$$

$$\frac{\Delta v_f}{v_f} = \frac{\Delta[\mathsf{B}]}{[\mathsf{B}]} \tag{9.8}$$

$$\frac{\Delta J}{J} = \frac{\Delta[\mathsf{B}]}{[\mathsf{B}]}\frac{v_f}{(v_f - v_r)} \tag{9.9}$$

$\frac{v_f}{(v_f - v_r)} = \frac{d\ln(v)}{d\ln[\mathsf{B}]}$ is called the **elasticity coefficient** ϵ. If the reaction is far from equilibrium, $v_r \approx 0$ and $\epsilon \approx 1$. For a given fractional change of the flux through Base an approximately equal fractional change in [B] is required. As the reaction approaches equilibrium, v_r approaches v_f and $\epsilon \to \infty$. Small changes in [B] will result in large changes in flux, which will return [B] to close to its original value. Therefore, all enzymes in a pathway operating near equilibrium have the same flux and maintain concentrations of all intermediates. This results in **homeostasis**. Elasticity can be defined not only for substrates, but also for effectors. This is how end-product inhibition works.

Remember: Flux control in biochemical reactions can be by limitation of

substrate in reversible reactions with rapid equilibrium between substrates and products

enzyme in irreversible reactions with slow regulated flow from substrates to products, which are far from equilibrium

An increase of [I] is equivalent to a decrease in (active) enzyme concentration. When e_i is the concentration of the i-th enzyme, then

$$\frac{d\ln(J)}{d\ln(e_i)} = C_i^J \tag{9.10}$$

is the flux control coefficient of the system with respect to the concentration of that enzyme. If, say, doubling the enzyme concentration resulted in a doubling of the flux, then $C_i^J = 1$. If this doubling of enzyme concentration had no effect at all on the flux, then $C_i^J = 0$. In a pathway of n enzymes usually several enzymes control the rate, thus the coefficients are somewhere between 0 and 1 and $\sum_{i=1}^{n} C_i^J = 1$ (additivity theorem). In practice, flux control coefficients of most enzymes in living organisms are small.

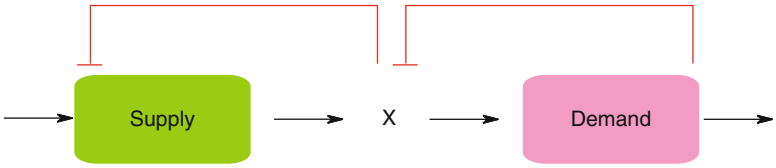
Inborn Errors of Metabolism
The vast majority of metabolic diseases are recessive; heterozygote carriers thus have 50 % of normal enzyme activity. In most cases, these people are phænotypically normal, as are their metabolite pools. This is a direct consequence of the low flux control coefficients of most enzymes. Cellular enzyme concentrations are in "metabolic excess" of what would be required to support the flux through the pathway; this adds stability to the system.

Experimental Methods
The enzyme concentration in an organism can be experimentally modified by genetic means, for example, in *E. coli* by placing the gene encoding the enzyme onto a plasmid under the control of *lac* promotor. Then exposing the bacteria to different concentrations of the nonmetabolisable inducer (IPTG) will result in different production rates of that particular enzyme without much effect on other proteins. In *Neurospora crassa* heterokaryons may be used for the same purpose. Another approach is to titrate the enzymatic activity in question by a membrane-permeable inhibitor.

Note that the analysis requires small changes!

9.1.1 Supply/Demand Analysis

Consider a situation where a metabolite X is produced by a supply pathway and utilised by a demand pathway. Both are treated as black box. X is a feedback inhibitor of the supply pathway, and the product of the demand pathway is a feedback inhibitor of the demand pathway:



An increase in supply would increase [X], and hence stimulate the demand pathway and slow down the supply pathway. An increase in demand would decrease [X], and hence stimulate the supply pathway. The elasticity of the responding block determines how much [X] has to change in order to reach a new steady state. The block with the lower elasticity and the higher flux control coefficient is in control.

We see that a high elasticity (low flux control coefficient) makes homeostasis easy to achieve. On the other hand, low elasticity (high flux control coefficient) allows us to fine-tune the flux through a pathway. Thus, if the supply pathway controls flux, the demand pathway will control homeostasis and *vice versa*. Feedback inhibition in this context allows homeostasis at low substrate concentrations that do not unduly increase the osmotic pressure inside the cell.

9.1.2 Mechanism of Flux Control

The rate-limiting step of a pathway must operate far from equilibrium and hence have a strongly negative ΔG (actual ΔG, not $\Delta G'^0$!). In such reactions $v_f \ll v_r$, therefore the elasticity coefficient is close to 1 and large changes in [substrate] have to occur to change the flux. Then the question arises as to how large changes of flux (*e.g.*, through glycolysis between exercising and resting muscle by a factor of 10^2) can be achieved without requiring concentration changes that are simply not observed in biological systems. Several mechanisms have been identified:

- allosteric control, for example by feedback inhibition (see Fig. 9.2)
- switching enzymes on or off by covalent modification, *e.g.*, phosphorylation (see Fig. 9.3)
- changes in enzyme concentration (transcription, translation and/or degradation)
- having forward and backward reactions performed by different enzymes (catalytic cycling)

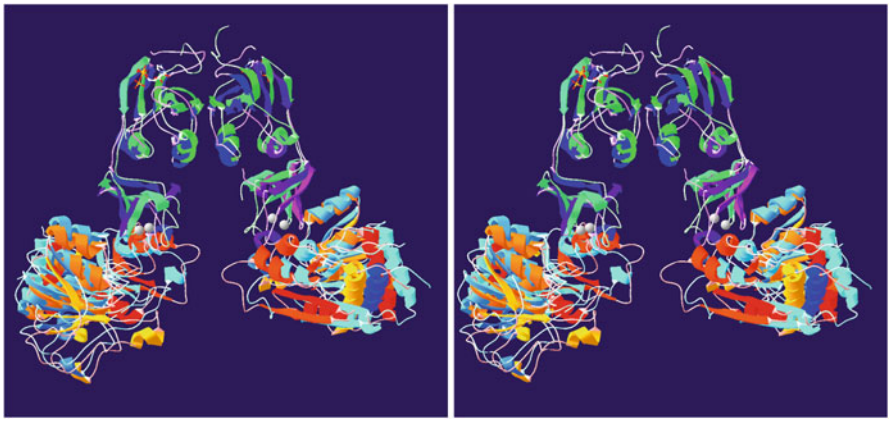


Fig. 9.2 Asp-carbamoyltransferase (here from *E. coli*) catalyses a regulated step of pyrimidine synthesis. It is inhibited by CTP (a pyrimidine), but activated by ATP (a purine). This ensures that pyrimidines and purines are synthesised in a 1:1 ratio for nucleic acid synthesis. This image shows the CTP-bound form of the regulatory (green to turquoise) subunit (T-state, PDB-code 1rab) and the R-state without ligand (PDB-code 1f1b, dark blue to purple. The catalytic subunits are red to orange and cyan to middle blue, respectively. Binding of CTP produces considerable conformational changes in the regulatory subunits, that are transmitted to the catalytic subunits by the finger-like protrusion (note the position of the Zn^{2+} ions)

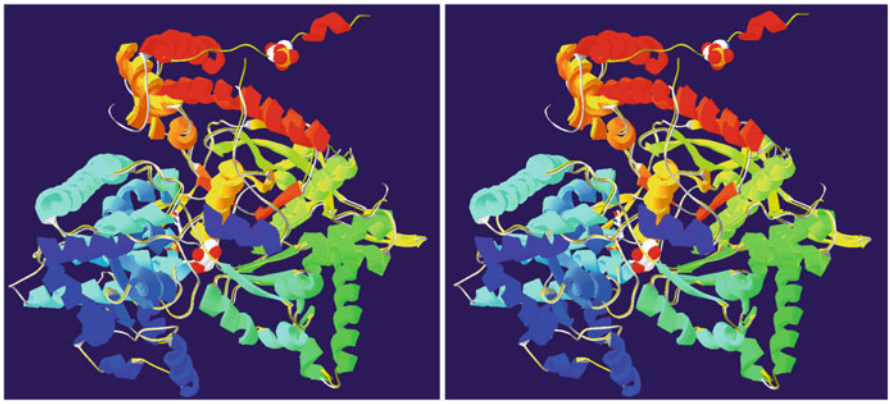


Fig. 9.3 Rabbit muscle phosphorylase in the phosphorylated (phosphorylase A, olive coils, PDB-code 1lwn) and unphosphorylated (phosphorylase B, grey coils, PDB-code 3zcv) form. Phosphorylase releases Glc-6-P from starch, using inorganic phosphate. Because of the low concentration of Glc-6-P in the cytosol the reaction is irreversible under physiological conditions. Phosphorylation of phosphorylase at Ser-14 causes the N-terminal α-helix to become disordered and move away from the protein (top of the figure). This affects both dimerisation and allosteric regulation by AMP, despite the small structural changes outside the phosphorylation site

The last point requires further explanation. Consider phosphofructokinase (PFK) as the rate-limiting enzyme of glycolysis. The reverse reaction is performed by Fru-1,6-bisphosphatase (FBPase):

$$\begin{array}{l} \text{Fru-6-P} + \text{ATP} \xrightarrow[\text{PFK}]{} \text{Fru-1,6-P}_2 + \text{ADP} \\ \text{Fru-1,6-P}_2 + \text{H}_2\text{O} \xrightarrow[\text{FBPase}]{} \text{Fru-6-P} + \text{P}_\text{i} \\ \hline \text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_\text{i} \end{array}$$

Both enzymes, when active at the same time, perform a **futile cycle**, with one high-energy phosphate lost per cycle. It was assumed that due to opposing effects of the allosteric regulators AMP and Fru-2,6-P_2, only one of these enzymes was active at any one time. This assumption proved wrong.

Assume that in a cell the $V_{\max}$ of PFK and FBPase are 100 and 10 relative units, respectively. Assume further, that a given increase of [AMP] (from 2 ADP $\xrightarrow{\text{adenylate kinase}}$ ATP + AMP) leads to a change of PFK activity from 10 to 90 % of maximal activity, and conversely the FBPase activity from 90 to 10 % of maximal activity. Then at low [AMP] the rates would be $v_f = 10$ and $v_r = 9$ and at high [AMP] $v_f = 90$ and $v_r = 1$. Thus AMP would increase the flux from $10 - 9 = 1$ to $90 - 1 = 89$. Instead of a 9-fold increase in flux, if both reactions were catalysed by the same enzyme, substrate cycling gives us an 89-fold increase in flux for the same change in [AMP]. Futile cycles are not futile at all, but effective mechanisms of metabolic control [4].

References

1. B. Chance, W. Holmes, J. Higgins, C.M. Connelly, Localization of interaction sites in multi-component transfer systems: Theorems derived from analogues. Nature **182**, 1190–1193 (1958). doi:10.1038/1821190a0
2. R. Heinrich, T.A. Rapoport, Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. Acta Biol. Med. Germ. **31**(4), 479–494 (1973)
3. H. Kacser, J.A. Burns, The control of flux. In *Symp. Soc. Exp. Biol.*, vol. 27, pp. 65–104 (1973). URL http://isites.harvard.edu/fs/docs/icb.topic94928.files/1002Reading/ControlofFlux.pdf
4. E.A. Newsholme, R.A.J. Challiss, B. Crabtree, Substrate cycles: their role in improving sensitivity in metabolic control. Trends Biochem. Sci. **9**(6), 277–280 (1984). doi:10.1016/0968-0004(84)90165-8

Part III
Special Proteins

Chapter 10
Protein Folding Diseases

Abstract Many important diseases including *morbus* ALZHEIMER, Prion diseases, type II diabetes, or some cancers involve protein misfolding. In all these diseases, called amyloidoses, intrinsically disordered proteins fold into β-helices. One misfolded protein molecule auto-catalytically converts other protein molecules, leading to aggregation. Intracellular aggregates are called inclusion bodies, and extracellular aggregates fibrils. It is unclear whether the aggregates themselves are the cause of the cell destruction that leads to disease, or whether the damage is caused by soluble intermediates. Thus, although all these diseases are caused by misfolding of different proteins in different cells, the underlying pathomechanism is identical. This raises hope that one day they may be treated in a similar manner.

10.1 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) contain relatively long coil segments, possibly interspersed with short segments of other secondary structures. Their tertiary structure is either an extended strand or a molten globule. When these disordered segments interact with other proteins, they bind under refolding to a different secondary and tertiary structure; the partner protein serves as a folding template. We do not yet fully understand these sequences; it appears that stretches of many identically charged amino acids keep that part on the surface of the protein as an extended coil, preventing the hydrophobic interactions required for folding. A large number of hydrophobic amino acids (more than 5 in a row) result in considerable entropy gains upon binding to a partner; these segments often form molten globules.

In enzymes intrinsic disorder is relatively rare (although examples do exist, *e.g.*, chorismate mutase), however, it is an important feature of many regulatory proteins (see Fig. 10.1). These regulating proteins may bind to many regulated proteins, and their structure is different with each partner. On the other hand, a regulated protein may bind different regulating, forcing each of them into similar structures. Thus the intrinsically disordered segments allow one-to-many and many-to-one relationships between regulating and regulated proteins and effectively dissociate specificity and affinity: folding upon binding results in loss of entropy; this reduces ΔG and hence increases K_d. As a result, binding is readily reversible, allowing quick termination

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_10

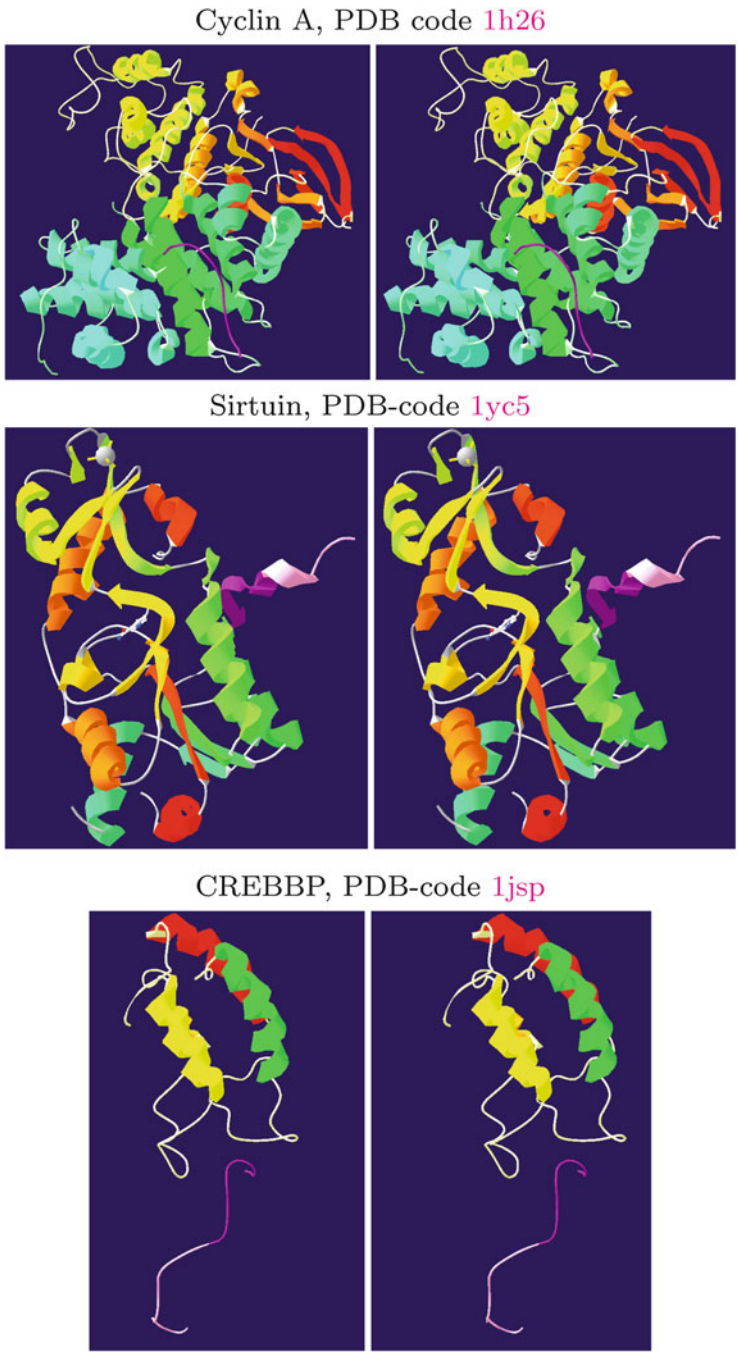


Fig. 10.1 Tumor supressor protein p53 C-terminal domain (residues S378-K386, deep purple) in contact with three different proteins. This domain is intrinsically disordered and assumes different structures with different binding partners

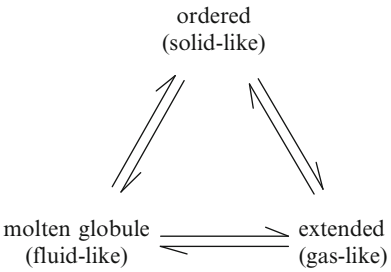


Fig. 10.2 Intrinsically disordered proteins exist in 3 states ("the trinity"): folded, molten globule, and disordered. Each of these may have its own biological function

of a signalling event. The surface area available for binding is larger (70 $Å^2$ per residue), and at the same time the flexibility to accommodate binding is increased. Binding is often regulated by posttranslational modification of proteins such as phosphorylation (see later); this can change K_d by several orders of magnitude and generates molecular switches (Fig. 10.2).

Disordered sequences are very sensitive to proteolytic attack; this creates technical problems for the researcher ("purification into nonexistence") and is certainly one of the reasons why the significance of these segments is only now more fully appreciated. It also leads to rapid degradation inside the cell, important for regulatory function.

The interaction between partners is very robustly encoded in the domains and maintained during evolutionary events which modify rather than abolish an interaction. Experiments indicate that early enzymes probably were more disordered and hence flexible, by becoming more structured enzymes specialised on a particular function and substrate. Because of their importance in regulation intrinsically disordered proteins are more frequent in metazoans (more than 500 have been described) than in protozoans, who have more than prokaryotes. An exception to that rule are the *Apicomplexa*, who have unusually many intrinsically disordered proteins. Important pathogens, including the malaria parasite, belong to that clade. How their pathogenicity depends on protein disorder is a topic of current research.

Amyloid formation—a process causing various debilitating diseases—can be understood in similar terms of intrinsic disorder (see later).

10.1.1 Protein-Only Elements of Inheritance

Proteins with alternative folding have been conserved throughout eukaryotic phylogenesis and actually serve very useful functions in a variety of organisms. In yeasts they confer new—often beneficial—phænotypes, and activation of a protein (Cytoplasmic polyadenylation element binding protein (CPEB)) by conversion to its prion conformation is required for memory formation in the snail *Aplysia* (reviewed in [22]).

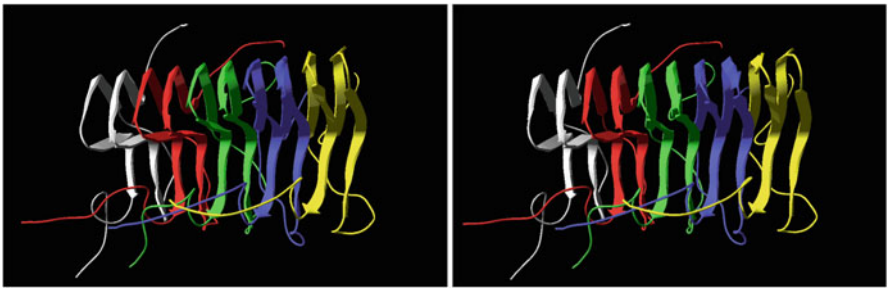


Fig. 10.3 Structure of amyloid formed by a yeast prion protein (HET-s fragment 218–289, PDB-code 2rnm), determined by solid state (magic angle spinning) NMR. Each protein forms a β-helix; those come together from different proteins (represented by different colours) to form extended parallel β-sheets

So far we have looked upon protein folding as driven by the reduction of ΔG to one stable, called **native**, structure. However, there are proteins that have a metastable native structure and that can fold into one or even several more stable structures with different biological properties.

The problem is that contact of some proteins with their alternative conformation results in their **autocatalytic** conversion into the alternative conformation, which is not only devoid of the biological function but also has a tendency to form **aggregates**, called **amyloids**. Surprisingly all amyloids seem to have basically the same structure, fibres made of β-helices. Many proteins that form amyloid tend to have more Gln and Asn, aromatic and β-branched chain and fewer polar amino acids than average proteins, however, this is not a universal feature (Fig. 10.3).

In yeasts there are several such proteins; the amyloid states are transferred to the daughter cells in mitosis and meiosis (dominant), inheritance is in a non-MENDELIAN, cytosolic manner. Some of these amyloids can be cured by growing cells in the presence of guanidinium hydrochloride (GuHCl), a compound which unfolds proteins. Low temperature also prevents the autocatalytic amyloid formation and can even break up amyloid fibrils; apparently amyloid formation is driven mainly by entropy.

10.2 Amyloidoses

Amyloids (a word coined by RUDOLF VIRCHOW [43] from the Greek words for "starch" and "similar", as amyloid gives a blue reaction with iodine) are involved in (many say the cause of) several severe human diseases [38]. It is, however, unclear whether the pathogenesis is really caused by the large amyloid aggregates or rather by soluble, oligomeric intermediates. The latter have been shown to solubilise phospholipid vesicles [28]. Trapping of other proteins in the aggregate—which then cannot perform their normal function—may also play a role. Another hypothesis is

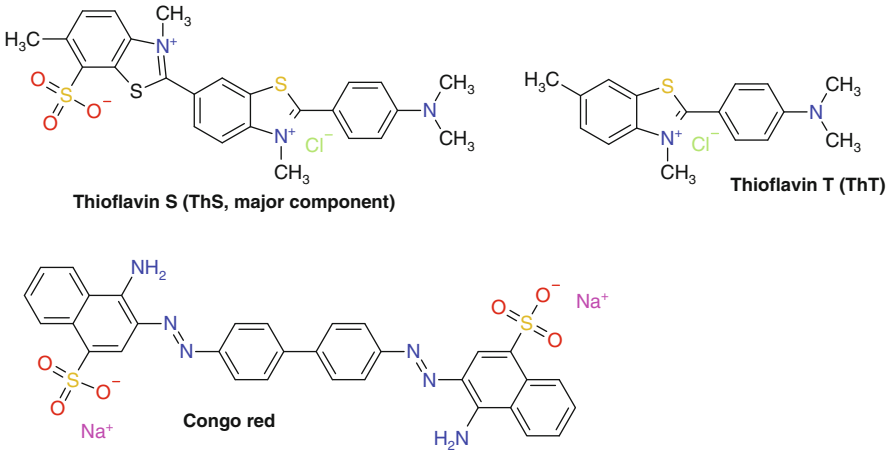


Fig. 10.4 Dyes that can be used to detect amyloid. The long-stretched molecules get ordered along the amyloid fibrils, leading to a change in optical properties. Congo red shows apple-green birefringence if stained histological sections are viewed between crossed polarisers. Thioflavins fluoresce after binding to amyloid; this also works in solution. ^{11}C-labelled derivatives of thioflavin have been used to detect amyloid deposits in the living body by PET

that the oligomeric aggregates stimulate the unfolded protein response, leading to a reduction in protein synthesis and thus to cell death [21].

Because the conversion of proteins into amyloid is a slow process amyloidoses tend to manifest in the **brain**, where cell turnover is low, at **advanced age**, and in the form of a slow **degeneration**. In our aging society this puts severe strain on public health budgets and the families of the affected. Because disease onset is past the usual age of reproduction there is little evolutionary pressure against these diseases.

Detection of amyloid in histologic sections often relies on the colour or fluorescence changes of dyes bound to them; examples include Congo Red and Thioflavin T (see Fig. 10.4). Binding of these dyes is saturable; the binding sites have been identified [29].

One possible mechanism by which amyloidoses may one day be treated is to employ small molecules that shift the equilibrium between folded and misfolded protein by binding specifically to the folded form. Such molecules would act as molecular chaperones; they have been dubbed **pharmacoperones** (see, e.g., [15]).

10.2.1 Prion Proteins and Prion Diseases

The so-called prion diseases (for more details see [23]) result in destruction of the brain; this is eventually fatal. The first such disease described (in the eighteenth century) was **scrapie** in sheep. It was noted that this disease could be spread to other

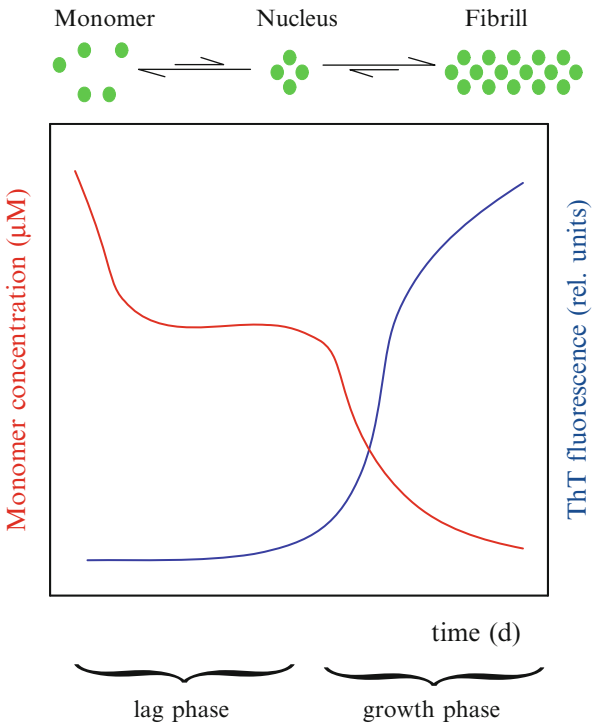


Fig. 10.5 *Top*: Amyloid aggregation resembles crystallisation in that very small aggregates are unstable; they dissolve faster than they grow. Once an aggregate has achieved a certain size it serves as nucleus for further addition of monomers. *Bottom*: Kinetics of amyloid aggregation. Initially, speed of aggregate formation (as measured by ThT fluorescence, *blue line*) is low, even though the monomer concentration (*red line*) is high enough to support aggregation (lag phase). Once stable nucleï have been formed they grow rapidly, using up the available monomer (growth phase). Eventually, a steady state between growth and dissolution of amyloid fibrils is achieved (steady-state, equilibrium phase). If a small amount of fibrils is transferred into a fresh batch of monomer solution, immediate aggregation—without lag-phase—is observed, as the fibrils serve as nucleï (not shown)

animal species by feeding them infected brain. One of the most remarkable features of the infectious agent of scrapie is its incredible resistance to destruction. A sheep head, buried for 3 years, proved still infectious when fed to mice. The agent cannot be inactivated with formalin (methanal), with UV light, or by heat sterilisation.

Similar observations were made with **kuru**, a disease prevalent in the Fore people of Papua, New Guinea [17, 20]. Kuru was spread by endocannibalism (ritual eating of dead relatives), until the colonial power (Australia) stomped out this habit in 1957. New cases are still reported from time to time, but only in older people who probably became infected while cannibalism was still rife. This points to a very long incubation period which can exceed 40 years. Victims first giggle and tremble uncontrollably (kuru actually means “the laughing death”), later they lose all awareness and control over body functions and finally die.

10.2.1.1 Pathomechanism

Pathologically, these and other diseases, collectively known as **spongiforme encephalopathies**, are very similar. They all result in the formation of intracellular vacuoles in brain tissue (at locations specific to a particular disease), which eventually cause death of the affected cell and fatal failure of brain function. These plaques contain an abnormal protein, PrP^{sc} (Prion protein scrapie, **prion** stands for proteinaceous infectious particle).

Researchers who tried to isolate an infectious agent from these plaques so far invariably have ended up with preparations which contained only PrP^{sc}, but no nucleic acids. Methods usually employed to destroy nucleic acids (nuclease treatment, UV-irradiation at 250 nm) fail to destroy prion infectivity, whereas reagents that react with proteins—such as Diethylpyrocarbonate (DEPC), which reacts with His (see Fig. 10.6)—inactivate it. This led TIKVAH ALPER to suggest that the causative agent is not a microorganism as in other diseases [2, 3]. S. PRUSINER has vigorously followed up on this idea, finally succeeding in the isolation of PrP^{sc} (Nobel Prize in 1997) [37]. Recently, WANG *et al.* [44] could demonstrate that recombinant prion protein produced in *E. coli* can cause prion disease in mice. In essence, this fulfills KOCH's postulates for prion diseases.

KOCH's Postulates
were established by ROBERT KOCH to prove a causal relationship between an agent and a disease. They are still in use, but certain limitations have to be kept in mind:

- Agent is present in affected, but not healthy individuals (although asymptomatic carriers can exist for many pathogens).
- Agent can be isolated from the infected host in pure form and cultured *in vitro*. However, there are cases such as whooping cough, where the characteristic signs appear only after the bacteria have been eliminated by the body's immune system. For some pathogens it has not yet been possible to establish suitable culture conditions.
- The isolated agent should cause identical disease in a suitable healthy model organism. Note, however, that certain disease may strike only in some, but not all individuals, for example, those with reduced immune status.
- The agent can be reisolated from the model organism and shown to be identical to the original isolate.

PrP^{sc} is derived from a protein normally present in the brain, PrP^{c} (c for cellular, the gene is located on chromosome 20). Both forms are membrane bound via a GPI-anchor (see page 368), but whilst PrP^{c} is bound to the plasma membrane PrP^{sc} occurs mainly inside the cell on lysosomal membranes.

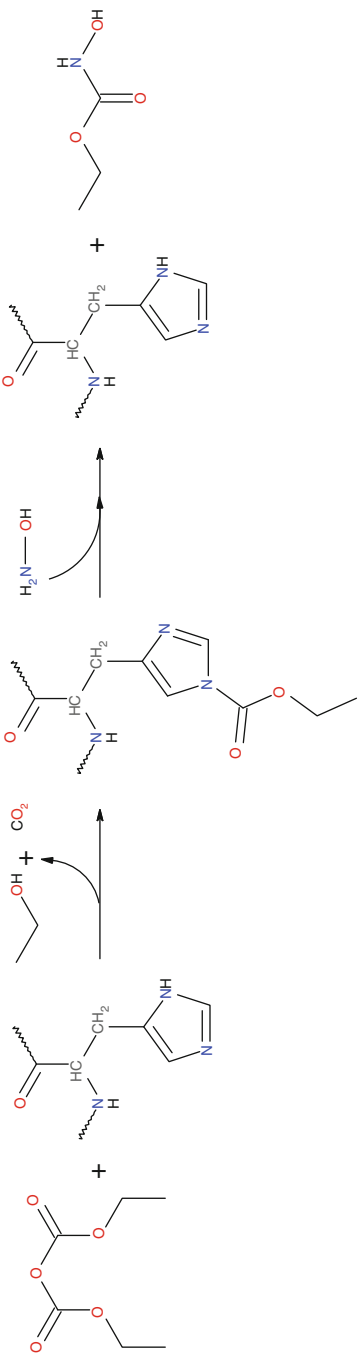


Fig. 10.6 PrP^{sc} can be made noninfectious by DEPC, a reagent that inactivates proteins (at His-residues), but not nucleic acids. Regeneration of the His-residue with hydroxylamine makes the protein infectious again. Reagents that inactivate nucleic acids have no effect on infectivity. This is one piece of evidence for the "protein only" hypothesis of prion transmission

Fig. 10.7 Stereo view of the normal PrP^c (bovine, residues 23–230, determined by NMR, PDB-code 1DX0). According to one hypothesis the entire N-terminal region (the two β-strands and the small α-helix turns into a sheet of 4 anti-parallel β-strands during conversion to PrP^{sc}. However, due to its tendency to aggregate no experimental structures are available for PrP^{sc} yet (recent review in [39])

The function of PrP^c is unknown, knock-out mice strains without the genetic information for this protein are resistant to infection with PrP^{sc}. Otherwise they appear normal, but show subtle changes in electrophysiological experiments. If copper is present during the *in vitro* folding of PrP^c the protein acquires superoxyde dismutase activity ($2O_2^- \cdot + 2H^+ \rightarrow O_2 + H_2O_2$). Thus it may play a role in the protection of the brain against damage from **reactive oxygen species (ROS)**.

Apparently PrP^{sc} causes an autocatalytic conversion of PrP^c to PrP^{sc}. Because these two forms of the protein have different conformations, they show different fragments after treatment with proteases and different sensitivities to urea and GuHCl. Indeed, different strains of PrP^{sc} also result in different proteolytic fragments; these fragments therefore can be used to identify them (Fig. 10.7).

Inherited encephalopathies apparently are caused by mutations in the PrP^c gene, which make spontaneous conversion to PrP^{sc} more likely. They are inherited in an autosomal dominant manner:

GERSTMANN-STRÄUSSLER-SCHEINKER**-disease (GSS)** [16] (OMIM #137 440) Patients have difficulty in coordinating their movement (ataxia, nystagmus, tremor), and lose their speech and finally control over body functions. Disease is protracted (1–11 a). Familiar cases have been described with F198S or P102L in the prion protein.

CREUTZFELDT-JAKOB **disease (CJD)** [11, 26] (OMIM #123400) Patients suffer from myoclonus, ataxia, hallucinations, loss of memory, change of personality and dementia, in the final stages akinetic mutism (decerebrate rigidity). Death usually occurs within 4–6 mo after first symptoms. In familial cases the patients are younger than in sporadic (40 vs. 65 a); mutations E200K and V210I in prion protein have been reported.

Fatal familial insomnia (FFI) (OMIM #600072) At age 40–50 patients suddenly fail to sleep. Hypnotics show no effect. Secondary to sleep deprivation dysautonomia develops: myosis; elevated blood pressure, heart rate, and temperature; profuse sweating; myoclonus; and impotency. As the disease progresses,

patients lose the ability to walk, keep their balance, control their sphincter, and speak. Hallucinations, panic attacks, agitation, and phobias develop, but unlike the other prion diseases patients retain their mental capacity until almost the end [31]. Death occurs between 0.5 and 3 a after onset. On autopsy one finds neuronal degradation and reactive astrocytosis in the anterior and dorsomedial thalamic nucleï, but without spongiosis. The disease is usually familial with D178N of the prion protein. However, FFI occurs only if amino acid 129 is Met; if it is Val, the D178N mutation leads to CJD. The reason for this selectivity is unclear [25]. Spontaneous cases of FFI have also been reported [30].

10.2.1.2 Orally Infectious Prion Diseases: Kuru, BSE, and vCJD

It appears that prions can be absorbed intact in the gastrointestinal tract (possible by the microfold (M) cells of the lymphatic system [33]) and then move through the body first in the Gut-associated lymphoid tissue (GALT) and then the enteritic and autonomous nervous system upward into the spinal cord and brain. There they start the autocatalytic conversion of PrP^{c} to PrP^{sc}, that leads to disease outbreak. Once Kuru patients die, they are eaten by their mates, spreading the disease further. Lymphatic (cecum, PEYER's patches, spleen, tonsils, bone marrow) and in particular neural tissue (spinal cord, brain, eyes) of infected people and animals contain the highest concentration of PrP^{sc} (10^6 and 10^9 infectious particles per g, respectively), in animal experiments 0.1 g of infected cow brain can transmit the disease to other animals.

Scrapie-infected sheep offal, converted to bone meal, had been used to feed cows in Britain without adequate precautions. This led to the appearance of **mad cow disease** or **bovine spongiforme encephalopathy (BSE)**; some humans eating infected beef (about 750 000 infected animals were converted to human food) have come down with "**vCJD**". Apparently prions that have crossed a species barrier once are much more prone to do so again. Because the incubation period is very long (at least 15 years) we do not yet know how many people ultimately will fall victim to this new disease. Infected cow offal, converted to bone meal, has been exported from Britain all over the world and fed to other animals. Export to third-world countries even continued after export to other EU countries was banned because of BSE. This may mean that BSE was spread in animals fed with bone meal containing food worldwide (not only cows and sheep, but also pigs, chicken, cats, dogs,...). One of the current concerns is that although scrapie in sheep does not appear to be infectious to humans, BSE can be transmitted to sheep and is then difficult to distinguish from scrapie.

Compared to classical CJD, the patients with vCJD are much younger (29 compared to 65 a on average; the youngest was 14). This may have to do with the more active immune system in children, which ironically may make them more susceptible to infection. Also, children tend to eat more junk food (hamburger, sausages). The course of the disease is protracted (14 months from outbreak to death compared to 4.5 months in classical CJD). First symptoms usually are psychiatric:

depression, aggression, and loss of memory. As of August 2013, 227 patients have been reported worldwide, 176 in Britain alone.

Vertical transmission of scrapie in sheep has been known to occur for some time, and there were some fears that it may occur for vCJD in cattle and humans. Possible vertical transmission (mother to unborn child) has been discussed in [19], but is now in doubt [32]. Several people who later succumbed to vCJD were blood donors, and infected blood can transmit the disease in animal experiments. As the blood was pooled with that from other donors for the production of blood products, several hundred people have been given potentially tainted medicines. Again newer studies suggest that transmission by transfusion, if it occurs at all, must be a rare event [4].

Another possible route for nosocomial infections are surgical instruments (1 case described) and transplants of neuronal tissue (*e.g. Dura mater* or corneas) or pharmaceuticals made from brain (*e.g.,* growth hormone). The infectious agent is difficult to destroy with sterilisation techniques currently in use.

As long as there are only relatively few cases, there will be no cure: the costs of developing a pharmaceutical are so high (according to industry figures 800 000 000 € on average, independent watchdog organisations have claimed that "only" 200 000 000 € are required) that the industry will do this only for widespread diseases, which allow these costs to be recovered.

It has been observed that in vCJD all victims thus far had a mutation of V129M in both copies of the PrP^c gene (MM genotype). However, kuru victims with VV and VM genotype are also observed, thus the MM genotype may simply reduce the incubation period (or kuru and vCJD may behave differently in this respect).

10.2.2 Neuronal Amyloidoses

Amyloidoses of the brain [21] can occur spontaneously or they can be inherited. In inherited disease age of onset is relatively early (30s or 40s); spontaneous disease takes longer to develop with onset in the 60s or later. An exciting recent finding is that amyloidoses may occur as a result of **somatic mosaicism** [5], that is, mutations that occur during cell divisions in early embryonal development and affect only descendants from this mutated cell. Especially in the brain, where cells are highly linked to each other, a small fraction of affected cells may suffice to induce disease. Neuronal amyloidoses affect not only the central, but also the enteric (gut) nervous system. ENS biopsies may therefore be useful for early diagnosis of these diseases [33, 41]. This, of course, will be really useful only once effective treatment options exist.

10.2.2.1 *Morbus* ALZHEIMER

Morbus ALZHEIMER (dementia) is named after ALOIS ALZHEIMER (German psychiatrist and neuropathologist, 1864–1915), who first described it in 1907 [1]. The disease is clinically characterised by the loss of (short-term) memory, excitation,

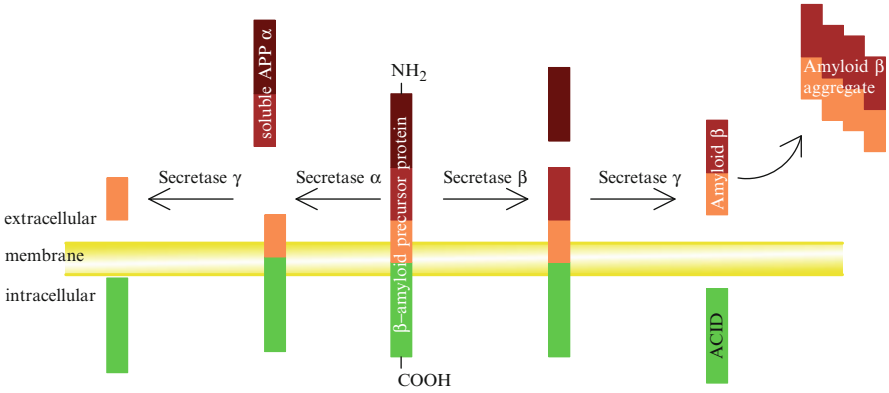


Fig. 10.8 Amyloid formation in ALZHEIMER disease

apathy, paranoia, depression, aggression with possible violence, progressing over loss of language, immobility, and incontinence to finally death. Usual age of onset is $\geq$ 65 a; this disease is spontaneous and age related. About 2 % of the population at age 65 a is affected, by age 80 a prevalence increases to 20 %. Beyond age 85 a the prevalence decreases again, as the patients rarely reach that age. With the growing number of elderly in advanced societies these patients already put a considerable strain on their families and the public health system; this is likely to increase in future. WHO estimates that there are currently about 29×10^6 patients living with ALZHEIMER disease, and that this number will increase to 106×10^6 in 2050 (when the incidence will be 1:85).

There is also an inherited form of the disease (OMIM #104300, accounting for about 5 % of cases) characterised by an early onset ($\leq$ 60 a). Strictly speaking, the case described by ALZHEIMER was early-onset (presenile dementia), the late-onset form should be called "senile dementia of the ALZHEIMER type (SDAT)", but this distinction is rarely made.

In ALZHEIMER disease the **β-amyloid precursor protein (APP)** is proteolytically cleaved extracellularly by β-secretase, then within the membrane by γ-secretase. The extracellular fragment resulting from the latter cleavage is the β-amyloid, which forms neuritic plaques (Fig. 10.10). If instead by β-secretase the APP is first cleaved by α-secretase, then the product resulting from γ-secretase cleavage cannot form plaques (see Fig. 10.8). Injection of β-amyloid containing brain extracts into the peritoneal cavity of mice transgenic for human APP leads, in due course, to an ALZHEIMER-like disease. β-amyloid, therefore, has prion-like properties [14]. Interestingly, the A673T mutation in the β-secretase cleavage site of APP protects Icelandic people from late-onset ALZHEIMER disease [27]. Formation of β-amyloid begins in the basal temporal neocortex and spreads from there along reproducible neuronal pathways throughout the neocortex [34] (Fig. 10.9). Our knowledge of the pathomechanism of ALZHEIMER disease has been reviewed recently [12].

The cytosolic part of APP left over after cleavage is called APP intracellular domain (AICD). The role of AICD in ALZHEIMER disease is controversial, as it appears short-lived. However, it interacts with about 20 identified partner

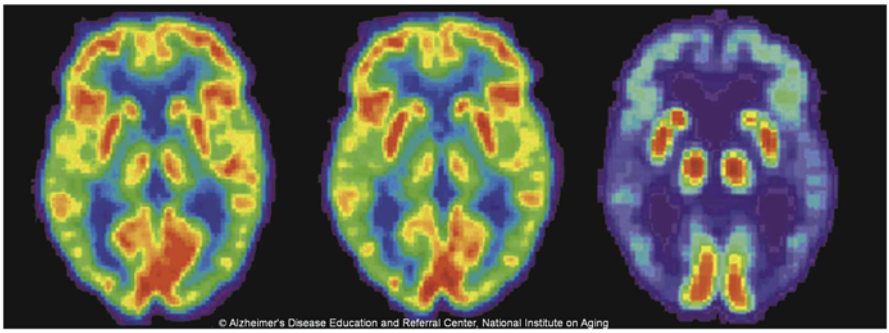


Fig. 10.9 Positron emission tomography (PET) scan of the brains of *left*: normal 20 a old, *middle*: normal 80 a old, *right*: 80 a old with ALZHEIMER disease. Colours denote metabolic activity in the brain (red = high to blue = low). Figure © Alzheimer's Disease Education and Referral Center, National Institute on Aging

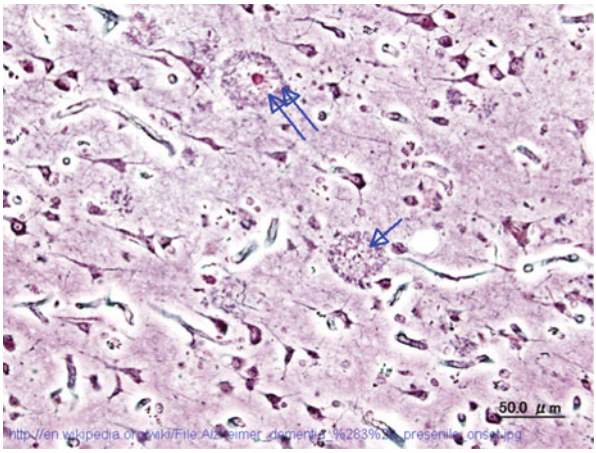


Fig. 10.10 Senile plaques (*arrow*), in part with amyloid core (*double arrow*), in the cerebral cortex of a patient with ALZHEIMER disease, silver stain. Source: http://en.wikipedia.org/wiki/File:Alzheimer_dementia_%283%29_presenile_onset.jpg

proteins. Amongst those is the histone acetyltransferase TIP60 (affects DNA/histone interactions and hence gene transcription) and the adapter FE65 (increases stability of AICD). The AICD/FE65/TIP60 complex ("AFT") can be found in the nucleus, where it may act as a transcription factor. In addition to the extracellular neuritic plaques, histology of brains from patients with ALZHEIMER disease also reveals neurofibrillary tangles inside the cells. These consist of hyperphosphorylated τ-protein (component of the cell skeleton). One of the potential target genes identified for AFT is glycogen synthase kinase 3β, which is one of the kinases responsible for τ-phosphorylation. If these results could be confirmed, then AFT would connect plaque and tangle formation. Another hypothesis is that β-amyloid serves as a template for the misfolding of τ-protein [34]. Misfolding of τ is also observed in some other degenerative neuronal diseases (tauopathies, see below).

Further investigation into the biochemistry of ALZHEIMER-brains reveals dysfunctional mitochondria which produce high amounts of **Reactive oxygen species (ROS)**. ROS are very toxic to cells. It is currently unclear what the pathomechanistic relationship among plaques, tangles, and ROS is.

In early onset (familial) ALZHEIMER disease several mutations have been found. The genes PSEN1 (AD3, on chromosome 14) and PSEN2 (AD4, on chromosome 1) code for presenilin 1 and 2, respectively, which are components of γ-secretase. Mutations in APP (AD1, on chromosome 21q) have also been found. There is also significant association between ALZHEIMER disease and the ϵ 4 allele of the ApoE protein (AD2, on chromosome 19), which is involved in the transport of cholesterol in blood. In addition, mitochondrial DNA-polymorphism and several other mutations in other genes have been described as associated with ALZHEIMER disease. Patients with DOWN **syndrome** (trisomy 21) are at increased risk for ALZHEIMER too; because of their general mental deficiency the onset of ALZHEIMER is particularly difficult to diagnose in these patients.

Because the cause of ALZHEIMER disease is unknown, prevention is in its infancy. The following recommendations, however, are widely agreed upon:

- control of blood pressure and [cholesterol]
- balanced diet with minerals and vitamins
- no smoking
- profession with high intellectual activity
- high physical activity
- limited TV consumption

Several authors have described a connection between ALZHEIMER disease and high aluminium ion intake (*e.g.* from cooking utensils) [10], however, this is now considered a red herring [6, 13]. Treatment of ALZHEIMER disease is possible with various pharmaceuticals that have come onto the market in the last couple of years, but all of them only reduce the symptoms, they do not slow down disease progression and do not change the final outcome. This includes the vitamins folic acid, cobalamin, and pyridoxin (possibly to reduce homocysteine levels); acetylcholine esterase inhibitors; NMDA-receptor antagonists and symptomatically used psychopharmaca.

10.2.2.2 Morbus PARKINSON (#168600)

The first description of this disease in modern medical literature was by the English physician JAMES PARKINSON in 1817 [36], but several ancient sources describe what appears to be PARKINSON disease, *e.g.,* in China the *Yellow Emperor's Internal Classics* from 425 BC and the *Ayurveda* in India (≈ 1000 BC).

In PARKINSON's disease (**shaking palsy**) the protein **α-synuclein** aggregates in the substantia nigra of the brain, forming LEWY**-bodies**. This leads to a failure of ER → GOLGI transport, resulting in the death of extrapyramidal cells in the *pars compacta* of the *substantia nigra*. These cells would normally produce **dopamine** which acts on *basal ganglia*. This results in a characteristic trembling, in slow movement,

and finally the cessation of movement. Patients show a characteristic, bend-forward posture when standing. Hallucinations, depression, and other psychiatric symptoms may be seen. The disease usually strikes between 50 and 60 a of age, more often in ♂than in ♀. Morbus PARKINSON is usually caused by gene duplication, but contact with toxic chemicals (rotenone (pesticide, [35]), glyphosat and paraquat (weed killers, [40, 42]), trichloroethylene (solvent, [18]), air pollution [7], or with amphetamines [8]) can show similar results. If the latter proves true, what are we doing to the children we treat with amphetamine derivatives (Ritalin, Adderall) for alleged attention deficit hyperactivity disorder (ADHD)?

Precursors of dopamine (L-Dopa (Levodopa)), dopamine agonists, or substances that interfere with dopamine breakdown (MAO-B or COMT inhibitors) are used to treat the disease. It is fascinating that the herbal remedies described in ancient Chinese sources contain substances that act as COMT-inhibitors.

10.2.2.3 Chorea HUNTINGTON (#143100)

HUNTINGTON disease [24] is caused by an expansion of CAG-repeats (base triplet encoding for Gln) from the normal 10–35 in the protein **huntingtin**. Poly-Gln > 40 amino acids form β-sheets which lead to aggregation. Inheritance is dominant autosomal with complete penetrance, chromosome 4p16.3. Huntingtin is required for endocytosis and hence for recycling of vesicle membranes after exocytosis. The death of brain cells in basal ganglia (putamen + caudate nucleus) reduces indirect inhibition of *globus pallidus internus*, resulting in activation of the thalamus and cortex. HUNTINGTON disease is characterised by jerky movements (choreoathetosis, choreia (Gr.) = dance), and cognitive and behavioural defects. There are several other trinucleotide expansion diseases in other proteins.

Not fully understood is the role of tissue transglutaminases in HUNTINGTON and also PARKINSON disease. Transglutaminases form isopeptide bonds between glutamine (R) and lysine (R') residues in proteins ($R{-}CO{-}NH_2 + H_2N{-}R' \rightarrow R{-}CO{-}NH{-}R' + NH_3$), the most well-known transglutaminase is factor VIII of the blood-clotting cascade. Tissue transglutaminase (tTG) is an enzyme found in all organs, including the brain. It has been found that **intra**molecular crosslinks in α-synuclein and tTG binding to synuclein increase as the disease progresses, but this may actually be a protective mechanism to reduce amyloid production by preventing the β-sheet formation.

10.2.2.4 Amyotrophic Lateral Sclerosis (ALS, LOU-GEHRING's Disease #105400)

This disease is named in the United States after HENRY LOUIS (LOU) GEHRING, a popular US baseball player (1903–1941). Another famous patient is astrophysicist STEPHEN HAWKING. It was first described by the French neuropathologist JEAN-MARTIN CHARCOT in 1874 [9]. ALS is characterised by progressive degeneration of upper and lower motor neurons in the lateral columns and anterior

horns of the spinal cord, resulting in fasciculations and atrophy of muscle. Limb weakness, dysphagia, dysarthria, hyperreflexia, positive BABINSKI sign, and exaggerated expression of emotion (pseudobulbar affect) are the most noticeable results. Subtle frontotemporal changes in cognitive function (decision making, attention, word-forming), or frontotemporal dementia (rare and usually familial) may also be present. In advanced stages, the patient may need noninvasive positive pressure ventilation or tracheotomy; the most common course of death is respiratory failure or pneumonia. ALS is caused by formation of inclusion bodies of superoxyde dismutase 1, poly-phosphorylated TAR DNA-binding protein or RNA-binding protein FUS (FUSed in sarcoma). The proteins are ubiquitinated. Inclusion bodies do *not* stain with Congo red or ThT (true amyloid?). Treatment: Riluzole (Rilutek®, Sanofi-Aventis) protects neurons by a variety of mechanisms (glutamate transporter, Na/Ca-transporter, PKC). Other drugs are in clinical trial.

There are several other, much rarer neuronal amyloid diseases; these are not covered here.

10.2.3 Amyloidoses in Other Organs

Although the amyloidoses of the central nervous system have attracted most attention, such diseases can form in any tissue. They may be the result of some mutation in a protein, or may be caused secondary to the excessive production of a protein in another disease.

10.2.3.1 Familial Visceral (OSTERTAG-)Amyloidoses (#105200)

These are caused by extracellular deposits of amyloid from various mutated proteins which interfere with organ function. The disease is autosomal dominant; the penetrance depends on the mutation. There is generally no inflammation in the affected tissues, but apoptosis may be seen. There is slow, but measurable turnover of deposited protein. Apart from the mutated protein, deposits also contain **amyloid P component (AP)**, a pentameric protein which is highly resistant to proteolysis and whose presence protects deposits from degradation. OSTERTAG-amyloidoses lead to proteinuria and progressive renal disease, petechia, organomegaly and bleeding (sudden and often fatal: liver, spleen), GI perforation, and sicca syndrome. Treatment is largely supportive; transplantation of kidneys, heart, or liver may be required.

Mutations in the following proteins may cause the disease:

Lysozyme (*153450), keratoconjunctivitis sicca and xerostomia, progressive nephropathy. Frequent cause of death is uncontrollable bleeding, especially liver. Heart not usually involved.

Apolipoprotein AI (ApoAI, +107680), nephropathy + hypertension, cardiomyopathy possibly leading to congestive heart failure, gut involvement, skin + laryngeal deposits (hoarseness), progressive neuropathy

Apolipoprotein AII (ApoAII, +107670), damages mostly the kidney

Fibrinogen A α-chain (+134820), Glu526Val mutation most common, mainly affects the kidney, but also liver, spleen, and adrenal.

Transthyretin (#105210) is a transporter for thyroxine in the blood and CSF. It causes a senile systemic amyloidosis in patients > 90 a, which is not usually clinically significant. However, a familial amyloidotic polyneuropathy is caused by > 100 mutations. Age of onset and penetrance (♂> ♀) varies; amyloid is formed in the heart, peripheral nerves, gastrointestinal tract, vitreous, lungs, and carpal ligament. Treatment is by liver transplant, as transthyretin is produced there.

10.2.3.2 Inherited, Localised Nonneuronal Amyloidoses

The following proteins are involved:

Gelsolin (#105120), Asp187Asn mutation leads to Finnish hereditary amyloidosis: slow progressive cranial neuropathy with lattice corneal dystrophy, sometimes renal and cardiac involvement, autosomal dominant, homozygous mutants with more severe phenotype known.

Atrial natriuretic factor (ANF, *108780), disease is caused by a frameshift mutation that leads to a protein with 12 additional amino acids.

Kerato-epithelin (#607541), causes hereditary lattice corneal dystrophy Avellino type. Patients are mostly of Italian descent. Several mutations known, which interfere with visual acuity. *Cave*: In patients with this disease LASIK is counter-indicated!

Lactoferrin is an antibacterial protein in milk and other secretions. In the eye it is produced by corneal basal cells. Corneal amyloidosis associated with trichiasis secondary to chronic eye infections is caused by a Glu561Asp mutation.

Medin is a 50 aa fragment of lactadherin (protein of unknown function) that causes amyloidosis in arterial media in ≈ 100% of Caucasians > 60 a. Clinical significance is limited.

α-Crystallins are small heat shock proteins (molecular chaperones) that keep all proteins (including themselves) in solution in the eye lens. Failure to do so leads to cataract, the cause of ≈ 1/2 of the 45×10^6 cases of blindness per year worldwide. In such cases the eye lens is replaced by an artificial one.

Lung surfactant protein C (*178620), is a proteolipid that stabilises DPPC at the air/liquid interface, using an α-helix through the DPPC double layer. Conversion of the α-helix to β-strand leads to pulmonary alveolar proteinosis. The problem is acquired (90 % of cases, *e.g.* smoking), secondary (after damage to macrophages) or congenital. Alveoli are filled with granular, eosinophilic PAS-positive material with degenerating macrophages. X-ray: ground-glass opa-

cifications with a superimposed interlobular septal and intralobular thickening. Treatment: Sequential bilateral whole-lung lavage.

Keratins (#171400), deposition of amyloids result in itching and scratching (friction amyloidosis) in interscapular region.

WISCOTT-ALDRICH**-protein** (#301000), mutation leads to immunodeficiency, thrombocytopenia, eczema, and recurrent infections. Deposition of complement and IgG on the basement membrane leads to glomerulonephritis. Bone marrow or cord blood transplant is curative, especially before age 5.

10.2.4 Amyloidoses Secondary to Other Diseases

The following proteins are involved:

Amylin (Islet amyloid protein, *147940), is cosecreted with insulin, reduces insulin effect on muscle, but not adipocytes. Amyloid is found in the pancreas of patients with type II diabetes or insulinoma between β-cells and capillaries → destruction of β-cells → insulin-dependence in advanced type 2 diabetes.

Immunoglobulin light chain (BENCE JONES protein) is seen in monoclonal plasma cell disorder. Immunoglobulin light chain is produced in excess, if the capacity of the kidney to handle it is exceeded, proteinuria and amyloid formation result in heart, kidney, peripheral nerve,and gastrointestinal and respiratory tract, but also everywhere else. The role of human herpesvirus 8 (HHV-8) in the etiology is controversial. Treatment: Chemotherapy, 4'-iodo-4'deoxydoxorubicin to solubilise amyloid (experimental), Thalidomide or Lenalidomide.

Serum amyloid A protein is an acute-phase protein, amyloidosis is seen in chronic inflammation, chronic infections, or some neoplasms. A mutation in this protein leads to familial Mediterranean fever: pain attacks (abdomen, joints, scrotum, chest) may be mistaken for surgical conditions—treat with NSAID.

β-microglobulin is produced as component of MHC-I and normally removed from serum by the kidney. In dialysis patients the protein can accumulate and cause amyloidosis: carpal tunnel syndrome, flexor tenosynovitis, subchondral bone cysts and erosions, pathologic fractures, macroglossia, GI bleeding, and congestive heart failure. Treatment: NSAID, intra-articular prednisolone, conservative treatment of joints, modern (polysulphone) dialysis membrane, β-M absorption with Lixelle-columns, renal transplantation.

Calcitonin can form amyloid in parafollicular C-cells secondary to medullary carcinoma of thyroid.

Prolactin (#600634), is overproduced in prolactinoma, a benign pituitary tumor present in 6–25 % of US population, but clinically significant only in 14:100 000. Symptoms in ♀ from elevated hormone levels.

Insulin Injection-localised amyloidosis: abscess-like, needs surgical extraction. May also occur in concentrated insulin solutions and reduce their efficacy.

10.3 Exercises

10.3.1 Problems

10.1. A 45-year-old male visited his family physician because he could not sleep. Hypnotics were without effect. The following signs were noted: myosis, hypertension, tachycardia, and elevated body temperature with diaphoresis. Over the following months the patient developed dream-like states, dysarthria, myoclonus, and impotency. He lost the ability to walk, keep his balance, and control his sphincter. However, his ability to think and understand did not diminish until after 15 mo the patient fell into a coma and finally died quite suddenly. Whilst the patient was able to talk about his suffering initially, at the end only the desperate look in his eyes told caretakers and family that he still fully understood what was happening to him. Autopsy showed neuronal degradation with reactive astrocytosis limited to the anterior and dorsomedial thalamic nuclei without spongiosis or inflammation.

What is the most likely diagnosis?

A Astrocytoma
B Myasthenia gravis
C Morbus ALZHEIMER
D Fatal familial insomnia
E New variant CREUTZFELD-JAKOB disease

10.2. A 45-year-old male visits his doctor because of involuntary movements in arms, legs, and face. He is very worried because the (protracted and finally fatal) disease of both his father and paternal grandfather had started with the same symptoms. Upon molecular investigation you find that the repeat-length of a CAG-stretch in the gene for a protein involved in endocytosis is 155 (normal up to 35).

The most likely diagnosis is:

A Chorea HUNTINGTON
B ALZHEIMER disease
C Kuru
D Fatal familial insomnia
E CREUTZFELDT-JAKOB-disease

10.3.2 Solutions

10.1 The description is typical for a case of FFI, notice also that the name of the disease is very descriptive.

10.1 HUNTINGTON disease caused by CAG-expansion in Huntingtin → poly-Glu amyloid formation

References

1. A. Alzheimer, Über eine eigenartige Erkrankung der Hirnrinde. Allg. Zeitschr. Psychiat. Psychisch-gerichtliche Med. **64**, 146–148 (1907). doi:10.1002/ca.980080612. URL http://www.alzforum.org/papers/uber-eine-eigenartige-erkrankung-der-hirnrinde
2. T. Alper, The exceptionally small size of the Scrapie agent. Biochem. Biophys. Res. Commun. **22**, 278–284 (1966). doi:10.1016/0006-291X(66)90478-5
3. T. Alper, W.A. Cramp, D.A. Haig, M.C. Clarke, Does the agent of Scrapie replicate without nucleic acid? Nature **214**, 764–766 (1967). doi:10.1038/214764a0
4. Anonymous, National Creutzfeldt-Jakob disease surveillance unit scientific report. Technical report, National Creutzfeldt-Jakob disease surveillance unit, University of Edinburgh, Edinburgh, 2007/08. URL http://www.cjd.ed.ac.uk/data.html
5. J.A. Beck, M. Poulter, T.A. Campbell, J.B. Uphill, G. Adamson, J.F. Geddes, T. Revesz, M.B. Davis, N.W. Wood, J. Collinge, S.J. Tabrizi, Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease. Hum. Mol. Genet. **13**(12), 1219–1224 (2004). doi:10.1093/hmg/ddh134
6. S.C. Bondy, The neurotoxicity of environmental aluminum is still an issue. NeuroToxicology **31**(5), 575–581 (2010). doi:10.1016/j.neuro.2010.05.009
7. L. Calderón-Garcidueñas, M. Franco-Lira, A. Mora-Tiscareño, H. Medina-Cortina, R. Torres-Jardón, M. Kavanaugh, Early Alzheimer's and Parkinson's disease pathology in urban children: Friend versus foe responses—it is time to face the evidence. BioMed Res. Int. page Article ID 161687 (2013). doi:10.1155/2013/161687
8. R.C. Callaghan, J.K. Cunningham, J. Sykes, S.J. Kish, Increased risk of Parkinson's disease in individuals hospitalized with conditions related to the use of methamphetamine or other amphetamine-type drugs. Drug Alcohol Depend. **120**(1–3), 35–40 (2012). doi:10.1016/j.drugalcdep.2011.06.013
9. J.M. Charcot, De la sclérose latérale amyotrophique. Prog. Med. **2**(325), 341–453 (1874)
10. D.R. Crapper, S.S. Krishnan, S. Quittkat, Aluminium, neurofibrillary degeneration and Alzheimer's disease. Brain **99**(1), 67–80 (1976). doi:10.1093/brain/99.1.67
11. H.G. Creutzfeld, Ueber eine eigenartige herdförmige Erkrankung des Zentralnervensystems. Ztschr. ges. Neurol. Psychiat. **57**, 1–18 (1924)
12. B. Da Costa Diasa, K. Jovanovica, D. Gonsalvesa, S.F.T. Weiss, Structural and mechanistic commonalities of amyloid-β and the prion protein. Prion **5**(3), 126–137 (2011). doi:10.4161/pri.5.3.17025
13. R. Doll, Review: Alzheimer's disease and environmental aluminium. Age Ageing **22**(2), 138–153 (1993). doi:10.1093/ageing/22.2.138
14. Y.S. Eisele, U. Obermüller, G. Heilbronner, F. Baumann, S.A. Kaeser, H. Wolburg, L.C. Walker, M. Staufenbiel, M. Heikenwalder, M. Jucker, Peripherally applied aβ-containing inoculates induce cerebral β-amyloidosis. Science **330**, 980–982 (2010). doi:10.1126/science.1194516
15. A.C.M. Ferreon, M.M. Moosa, Y. Gambin, A.A. Deniz, Counteracting chemical chaperone effects on the single-molecule α-synuclein structural landscape. Proc. Natl. Acad. Sci. USA **109**(44), 17826–17831 (2012). doi:10.1073/pnas.1201802109
16. J. Gerstmann, E. Sträussler, I. Scheinker, ÃIJber eine eigenartige, hereditär-familiäre Erkrankung des Zentralnervensystems, zugleich ein Beitrag zur Frage des vorzeitigen lokalen Alterns. Z. Neurol. **154**, 736–762 (1936). doi:10.1007/BF02865827
17. C.J. Gibbs, H.L. Amyx, A. Bacote, C.L. Masters, D.C. Gajdnsek, Oral transmission of Kuru, Creutzfeldt-Jakob disease, and Scrapie to nonhuman primates. J. Infect. Dis. **142**(2), 205–208 (1980). doi:10.1093/infdis/142.2.205
18. S.M. Goldman, P.J. Quinlan, G.W. Ross, C. Marras, C. Meng, G.S. Bhudhikanok, K. Comyns, M. Korell, A.R. Chade, M. Kasten, B. Priestley, K.L. Chou, H.H. Fernandez, F. Cambi, J.W. Langston, C.M. Tanner, Solvent exposures and Parkinson disease risk in twins. Ann. Neurol. **71**(6), 776–784 (2012). doi:10.1002/ana.22629

19. S.M. Gore, W.R. Gilks, J.W. Wilesmith, Bovine spongiform encephalopathy maternal cohort study—exploratory analysis. J. Roy. Stat. Soc. C (Appl. Stat.) **46**(3), 305–320 (1997). doi:10.1111/1467-9876.00071
20. W. Hadlow, Scrapie and Kuru. Lancet **274**(7097), 289–290 (1959). doi:10.1016/S0140-6736(59)92081-1
21. M. Halliday, H. Radford, G.R. Mallucci, Prions: Generation and spread versus neurotoxicity. J. Biol. Chem. **289**(29), 19862–19868 (2014). doi:10.1074/jbc.R114.568477
22. R.D. Hawkins, E.R. Kandel, C.H. Bailey, Molecular mechanisms of memory storage in aplysia. Biol. Bull. **210**(3), 174–191 (2006). URL http://www.biolbull.org/content/210/3/174.full.pdf+html
23. B. Hörnlimann, D. Riesner, H. Kretzschmar, *Prionen und Prionkrankheiten* (de Gruyter, Berlin, New York, 2001). ISBN 978-3-1101-6361-2
24. G.S. Huntington, On chorea. Med. Surg. Reporter **26**, 317–321 (1872). URL http://en.wikisource.org/wiki/On_Chorea
25. W.S. Jackson, Selective vulnerability to neurodegenerative disease: the curious case of prion protein. Dis. Models Mech. **7**(1), 21–29 (2014). doi:10.1242/dmm.012146
26. A. Jakob, Über eigenartige Erkrankungen des Zentralnervensystems mit bemerkenswertem anatomischen Befunde. Z. ges. Neurol. Psychiat. **64**(1), 147–228 (1921). doi:10.1007/BF02870932
27. T. Jonsson, J.K. Atwal, S. Steinberg, J. Snaedal, P.V. Jonsson, S. Bjornsson, H. Stefansson, P. Sulem, D. Gudbjartsson, J. Maloney, K. Hoyte, A. Gustafson, Y. Liu, Y. Lu, T. Bhangale, R.R. Graham, J. Huttenlocher, G. Bjornsdottir, O.A. Andreassen, E.G. Jönsson, A. Palotie, T.W. Behrens, O.T. Magnusson, A. Kong, U. Thorsteinsdottir, R.J. Watts, K. Stefansson, A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature **488**(7409), 96–99 (2012). doi: 10.1038/nature11283
28. R. Kayed, Y. Sokolov, B. Edmonds, T.M. McIntire, S.C. Milton, J.E. Hall, C.G. Glabe, Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases. J. Biol. Chem. **279**(45), 46363–46366 (2004). doi:10.1074/jbc.C400260200
29. B. Keshet, J.J. Gray, T.A. Good, Structurally distinct toxicity inhibitors bind at common loci on β-amyloid fibril. Protein Sci. **19**, 2291–2304 (2010). doi:10.1002/pro.509
30. J.A. Mastrianni, R. Nixon, R. Layzer, G.C. Telling, D. Han, S.J. DeArmond, S.B. Prusiner, Prion protein conformation in a patient with sporadic fatal insomnia. New Engl. J. Med. **340**(21), 1630–1638 (1999). doi:10.1056/NEJM199905273402104
31. D.T. Max, *The Family That Couldn't Sleep—Unravelling a Venetian Medical Mystery* (Portobello, London, 2007). ISBN 978-1-84627-089-5
32. K.L. Murray, Vertical transmission of variant CJD. J. Neurol. Neurosurg. Psychiat. **76**(9), 1318 (2005). URL http://jnnp.bmj.com/content/76/9/1313.full.pdf+html
33. G. Natalea, M. Ferruccia, G. Lazzeria, A. Paparellia, F. Fornai, Transmission of prions within the gut and towards the central nervous system. Prion **5**(3), 142–149 (2011). doi:10.4161/pri.5.3.16328
34. J.M. Nussbaum, M.E. Seward, G.S. Bloom, Alzheimer disease: A tale of two prions. Prion **7**(1), 14–19 (2013). doi:10.4161/pri.22118
35. F. Pan-Montojo, M. Schwarz, C. Winkler, M. Arnhold, G.A. O'Sullivan, A. Pal, J. Said, G. Marsico, J.-M. Verbavatz, M. Rodrigo-Angulo, G. Gille, R.H.W. Funk, H. Reichmann, Environmental toxins trigger PD-like progression via increased alpha-synuclein release from enteric neurons in mice. Sci. Rep. **2**, Article number 898 (2012). doi:10.1038/srep00898
36. J. Parkinson, *An Essay on the Shaking Palsy* (Sherwood, Neely and Jones, London, 1817). URL http://books.google.de/books?hl=de&lr=&id=3ZsTAQAAMAAJ&oi=fnd&pg=PA1&dq=Parkinson+1817&ots=LxtGwIJbcH&sig=DAtrM5IwPrFOgReHXUWPDpNKGNE#v=onepage&q=Parkinson%201817&f=false
37. S.B. Prusiner, Prions. Proc. Natl. Acad. Sci. USA **95**, 13363–13383 (1998). doi:10.1073/pnas.95.23.13363

38. S.B. Prusiner, Biology and genetics of prions causing neurodegeneration. Annu. Rev. Genet. **47**, 601–623 (2013). doi:10.1146/annurev-genet-110711-155524
39. J.R. Requena, H Wille, The structure of the infectious prion protein: Experimental data and molecular models. Prion **8**(1), 60–66 (2014). doi:10.4161/pri.28368
40. A. Samsel, S. Seneff, Glyphosate's suppression of cytochrome P450 enzymes and amino acid biosynthesis by the gut microbiome: Pathways to modern diseases. Entropy **15**(4),1416–1463 (2013). doi:10.3390/e15041416
41. S. Semar, M. Klotz, M. Letiembre, C. Van Ginneken, A. Braun, V. Jost, M. Bischof, W.J. Lammers, Y. Liu, K. Fassbender, T. Wyss-Coray, F. Kirchhoff, K.-H. Schäfer, Changes of the enteric nervous system in amyloid-β protein precursor transgenic mice correlate with disease progression. J. Alzheimer's Dis. **36**(1), 7–20 (2013). doi:10.3233/jad-120511
42. M. Thiruchelvam, B.J. Brockel, E.K. Richfield, R.B. Baggs, D.A. Cory-Slechta, Potentiated and preferential effects of combined paraquat and maneb on nigrostriatal dopamine systems: environmental risk factors for Parkinson's disease? Brain Res. **873**, 225–234 (2000). doi:10.1016/S0006-8993(00)02496-3
43. R. Virchow, Ueber eine im Gehirn und Rückenmark des Menschen aufgefundene Substanz mit der chemischen Reaction der Cellulose. Virchows Arch. path. Anat. Physiol. Klin. Med. **6**(1), 135–138 (1854). doi:10.1007/BF01930815
44. F. Wang, X. Wang, C.-G. Yuan, J. Ma, Generating a prion with bacterially expressed recombinant prion protein. Science **327**, 1132–1135 (2010). doi:10.1126/science.1183748

Chapter 11
Immunoproteins

Whoever desires peace should prepare for war
*(*PUBLIUS FLAVIUS VEGETIUS RENATUS*: On military matters (around 390 AD))*

Abstract Our bodies have efficient defence systems against invading pathogens and against cancer cells. We distinguish cellular from humoral and innate from acquired systems. Humoral defences consist of antibacterial polypeptides, Pathogen-associated molecular pattern (PAMP)-receptors, complement systems, and antibodies. Antibacterial polypeptides include enzymes such as lysozyme and phospholipase A which digest bacterial cell walls and membranes, and the pore-forming defensins. PAMP-receptors recognise molecules found on certain groups of pathogens such as lipopolysaccharide in GRAM-negative bacteria, lipoteichoic acid in GRAM-positive bacteria or mannose-ending glycoproteins in yeasts. They opsonise pathogens for the complement system. The complement system is a cascade of proteolytic enzymes that produce anaphylactic peptides and membrane pores that lyse cells marked as foreign either by PAMP-receptors or by antibodies. Antibodies are proteins that specifically recognise foreign material and opsonise it for the complement system and for the cellular immune system. They are produced from a limited number of building blocks that are randomly recombined and subjected to hypermutation. Thus an almost infinite variety of antibodies can be produced from a limited number of genes. Acquiring specific antibodies requires about two weeks, during this time the body has to rely on innate immunity. Once an antibody response has been mounted, the immune system remembers that pathogen for the rest of our life. Artificial exposure to antigens from important pathogens (immunisation) can protect us from infections and is, after sanitation, mankind's second most effective weapon against infectious diseases.

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_11

11.1 Overview

The human body is a nutrient-rich environment and would be colonised rapidly by virus,[1] bacteria, and other parasites if it were not protected by an immune system. This system not only recognises and destroys invading parasites (and even cancer cells), but does not attack the body's own cells. Both properties are equally important for our survival. An attack of our immune system against our own cells is called **autoimmune disease**. Such diseases tend to be debilitating, protracted, and eventually fatal. Examples are *myasthenia gravis* (antibody production against the muscle acetylcholine receptor), Systemic *lupus erythematosus* (SLE) (reaction against nuclear antigens), or *diabetes mellitus* type 1 (destruction of the β-cells in the islets of LANGERHANS in the pancreas).

We look at the immune system from the biochemical point of view, and focus on the proteins involved. Later in your studies you will take immunology courses that take a broader perspective. For a historical overview, see Table 11.1.

There are three different defense systems in our body: the barrier of our skin and mucous membranes, the **innate**, and the **acquired** immunity. An **antigen** is anything that triggers the acquired immune system. These systems are not isolated but closely intertwined. As we see, the main function of the adaptive immune system is to help the evolutionary older innate system deal with a particular antigen, just as a targeting scope can make a gun more effective.

The **innate immunity** is our first line of defense against pathogens that have breached the barrier provided by our skin. It consists of antibacterial polypeptides that digest bacteria or form holes in their membranes, phagocytic cells that devour any foreign material, and the complement system that lyses them. Because this system is inborn and does not require training, it can respond quickly to a new pathogen, but its efficiency is much lower than that of the acquired immune system.

Within about two weeks of encountering a new antigen the body mounts a specific immune response against it. This **acquired immunity** is highly specific and very effective, and it also results in a permanent memory of that antigen. If the antigen is encountered a second time, the specific immunity can be mounted much faster. For that reason we can get many diseases only once ("childhood diseases"). By artificially introducing us to antigens of important pathogens active **immunisation** shortens the time to mount a specific immune response. After sanitation immunisation is our second most important weapon against infectious diseases, which led to the eradication of smallpox (see Fig. 11.1).

[1] The word virus is of Latin origin, the original meaning was "saliva of rabid dogs". It is noncountable (like in English water or information); "viruses" is therefore etymologically incorrect. It is, however, possible to count virus particles (virions) or virus species.

Table 11.1 Key historic events in immunology

Event	Author
Survivors of bubonic plague can nurse victims without becoming infected again	THUCYDIDES 430 BC
Use of cowpox virus as smallpox vaccine	JENNER 1798 [30]
Germ theory of disease	HENLE 1840 [25]
Bacteria as cause of anthrax, KOCH's postulates	KOCH 1876 [32]
Use of attenuated pathogens for vaccination	PASTEUR 1881 [45]
Macrophages	MEČNIKOV 1894 [38]
Serum therapy	KITASATO & V. BEHRING 1890 [3]
Antibody specificity	EHRLICH 1891 [14, 15]
Complement system	BORDET 1896 [5]
Complement fixation test to detect infections	WASSERMANN *et al.* 1906 [55]
Major histocompatibility complex (MHC)	GORER, LYMAN & SNELL 1948 [22]
Antibodies produced by plasma cells	FAGREAUS 1948 [18]
Clonal selection theory	JERNE, BURNET [8, 31]
Thymus involved in cellular immunity	MILLER 1961 [39]
Embryonal exposure leads to tolerance (clonal deletion)	MEDAWAR 1961 [37]
Rats without lymphocytes have no adaptive immunity	MCGREGOR & GOWANS 1963 [36]
Enzyme-linked immunosorbent assay ELISA	ENGVALL & PERLMANN 1971 [17]
Monoclonal antibodies	KÖHLER & MILSTEIN 1975 [33]
Somatic rearrangement in Ig-genes	HOZUMI & TONEGAWA 1976 [27]
Somatic recombination of immunoglobulin genes	BRACK *et al.* 1976 [6]
HIV as cause of AIDS	BARRÉ-SINOUSSI *et al.* 1983 [2]

11.1.1 Cells of the Immune System

The cells that make up our immune system are originally produced in the bone marrow; they are derived from the same multipotent stem cells that give rise to erythrocytes and platelets. Some of these cells, however, migrate to other lymphoid organs during their development for final differentiation.

Primary **lymphoid organs** are bone marrow (and liver during early embryonal development), thymus, and in birds the *bursa fabricii* (an organ associated with the end gut). Secondary (peripheral) lymphoid organs are tonsils, spleen, lymph nodes, PEYER's patches and the appendix.

Apart from the erythroblast and megakaryoblast cell lines, which are not involved in the immune system, the multipotent stem cell of the bone marrow gives rise to the following cell lines (see also Table 11.2):

- The lymphoblast cell line develops into **lymphocytes**, the principal carriers of the acquired immune system:

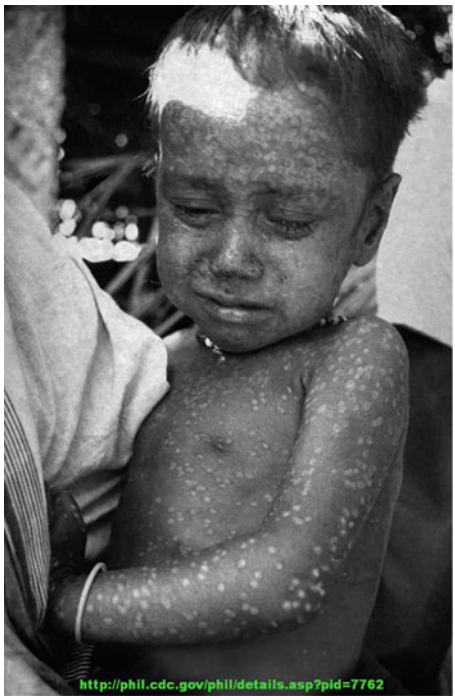


Fig. 11.1 The world's last reported case of smallpox, the two-year-old RAHIMA BANU from Bangladesh (1975). Starting with Bavaria in 1807 immunisation of all children was mandated worldwide; this terrible disease was declared extinct in 1980. It is hoped that extermination of polio will follow soon. The image, shot by S. O. FOSTER, comes from the CDC Public Health Image Library

– **T-lymphocytes** migrate from the bone marrow to the thymus for final differentiation, before they are released into the bloodstream. We distinguish the $CD8^+$ T-killer- (T_k) and the $CD4^+$ T-helper (T_h) cells. There are four known types of T-helper cells:

  T_H1 help macrophages digest bacteria enclosed in their phagosomes.
  T_H2 produce cytokines that increase mucous production during parasitic infection. This strengthens the barrier function of our mucosal epithelia and is also part of the allergic response.
  T_H17 produce cytokines that direct neutrophiles to a site of infection.
  T_{FH} stimulate antibody production in B-cells.

– In birds the **B-lymphocytes** differentiate in the *bursa fabricii*. A lot of effort was spend on the search for the bursa-equivalent in mammals; it now appears that in mammals the B-cells do not move to another organ, but develop completely in the bone marrow before they are released into the bloodstream. Once stimulated with antigen, a B-cell can develop either into an

Table 11.2 Properties of blood cells

	Erythrocyte	Platelet	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte
Size (μ m)	6–8	2–3	10–12	10–12	9–10	7–8	14–17
Number per mL	4–6 Mio	150 000-400 000	2800–5300	70–420	0–70	1400–3200	140–700
Differential leucocyte count (%)	–	–	40–75	1–6	< 1	20–45	2–6
Duration of development (d)	5–7	4–5	6–9	6–9	3–7	1–2	2–3
Half-life period (d)	120	7	0.3	0.5	0.25	Variable	Month to years

antibody-producing **plasma cell** or into a long-lived **memory cell**, which enables our body to mount a specific immune response much faster when it encounters an antigen a second time.

- **Natural killer (NK-) cells** kill virus-infected and cancer cells marked by antibodies (IgG1 and IgG3) on their surface. They are also important in innate immunity, killing cells with a low level of MHC-I, or cells expressing foreign carbohydrates. Killing occurs by release of granzymes and perforin from cytotoxic granules into the immune synapse between NK and target cells. Perforins ensure the delivery of granzymes into the target cell; granzymes are Ser-proteases that induce apoptosis by cleaving and hence activating caspases.

- The myeloblast cell line leads to granulocytes, which are responsible for innate immunity, although some of them help in acquired immunity as well. All granulocytes have lobed nuclei.
  - **Basophils** have granules that interact with basic dyes such as methylene blue; they will appear blue in blood smears stained with the usual GIEMSA-type of stain. These granules contain heparin (an anticoagulant) and mediators of inflammation such as histamine, which increase the permeability of blood vessels and attract other cells to a site of infection. These mediators are responsible for the typical signs of inflammation: *calor* (heat), *rubor* (redness), *tumor* (swelling) and *dolor* (pain, [49]). In connective tissue **mast cells** are found instead, which have similar function as basophils.
  - **Neutrophils** contain many lysosomes (primary granules) and smaller secondary granules. Both stain only weakly with GIEMSA-stain. Neutrophils are phagocytic cells that engulf and digest foreign material. They have few mitochondria and therefore an anaerobic metabolism, an adaptation to the low oxygen environment in tissue. They have a limited capability for regeneration and die after a single burst of activity, forming the principal component of pus. Phagocytosis by neutrophils is increased when pathogenes are covered with antibodies.
  - **Eosinophils** are packed with large, uniformly sized granules that bind acidic dyes such as eosin, giving them a dark pink colour in blood smears. Eosinophils are highly phagocytic for antigens covered with IgE; they are thought to be involved in defense against large parasites like worms. They are also involved in **allergic** responses.

- The monoblast cell line gives rise to **monocytes**, which spend only a short time in the blood before settling in tissues as **macrophages** (in connective tissue, often abbreviated MΦ), KUPFFER-cells (in liver), microglia (in brain), and dendritic cells of the reticuloendothelial system *etc.* These cells are collectively known as mononuclear phagocytes and form part of the innate immune response; they also present peptides from digested pathogens on their surface, allowing the acquired immunity to develop (see the section on MHC, pp 258). Some have additional specific functions, for example, microglia cells which are part of the blood-brain barrier. Plasmacytoid dendritic cells (pDCs) produce antiviral interferons.

11.2 Humoral Immunity: Immunoglobulins

Immunoglobulins are highly soluble proteins found in blood, lymph, and interstitial fluid. They are Y-shaped molecules: the two short (N-terminal) ends of the Y interact with the antigen, and the long F_c-end with effector cells (see Fig. 11.2).

Antibodies can protect against infection in four different ways:

Neutralisation occurs when antibody binding prevents the function of an antigen. This process is of particular importance against bacterial toxins and against virus infections. Viruses bind to and enter cells of their host in order to be replicated by them. Special surface proteins on the virion bind to proteins expressed on the surface of the host cell to start this process. If antibodies bind to the viral proteins, the virus can no longer multiply. Some bacteria produce very powerful toxins, for example, diphtheria, cholera, botulinum, or tetanus toxin. Antibodies against these toxins do not prevent infection, but a large part of the pathogenicity.

Opsonisation means that pathogens covered with antibodies are devoured more efficiently by phagocytic cells such as macrophages and neutrophils. These have special receptors for the F_c-ends of antibodies.

Complement activation occurs when antibodies bound to pathogens start a chemical reaction of soluble proteins in our blood, that lyse pathogens. The function of complement is described in detail on page 249.

Production of reactive oxygen species catalysed by antibodies may lead to direct killing of pathogens. This enzymatic activity of antibodies (which appears to be independent of antibody specificity) has been discovered only recently [56] and needs further investigation.

11.2.1 Structure of Immunoglobulins

Antibodies consist of several copies of a characteristic motive, a sandwich of two antiparallel β-pleated sheets connected by a disulphide bond (see Fig. 11.3). This **Ig-fold** is found in many proteins in the immune system and some other proteins with different function.

There are some differences in the sequences even of the constant regions of antibodies between individuals. About 20 such **allotypes** (from Greek: different marker) have been identified in human heavy, a few more in the light chain. Some such allotypes are enriched in specific populations; they may be a cause of different disease susceptibilities in different individuals, but this has not been confirmed. Allotypes are named after the chain in which they occur; *for example*, G1m(3) and G1m(17) are allotypes of the γ 1-chain with Arg and Lys, respectively, at position 214. Km(1) and Km(3) are allotypes with Val, Leu, or Ala, Val in positions 153 and 191 of the κ-chain. An individual heterozygous for, say, G1m(3) and G1m(17) will

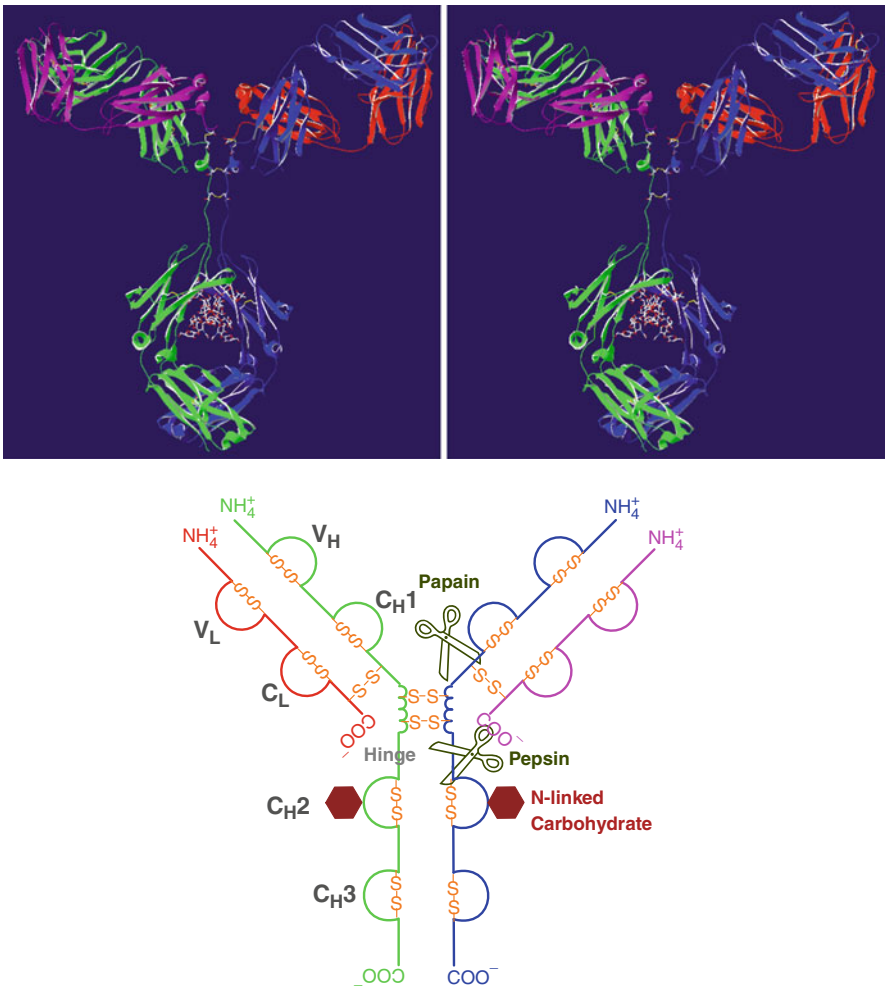


Fig. 11.2 IgG is the archetypical immunoglobulin. It consists of 2 heavy (*blue and green*) and 2 light chains (*purple and red*), connected by disulphide bonds. The light chains consist of 2 immunoglobulin motives each, and the heavy chains of 4. The latter are decorated with N-linked oligosaccharides. *Top*: Stereo view of the crystal structure of IgG (PDB-code 1FC2). *Bottom*: Pepsin can cleave IgG so that the antigen-binding site is still dimeric, but the F_c-end is removed. Papain, on the other hand, cleaves IgG above the disulphide bonds holding the heavy chains together. This results in F_{ab}-fragments with only one antigen binding site. Thus the effects of antigen cross-linking and of interactions of the F_c-end with other molecules can be experimentally dissociated

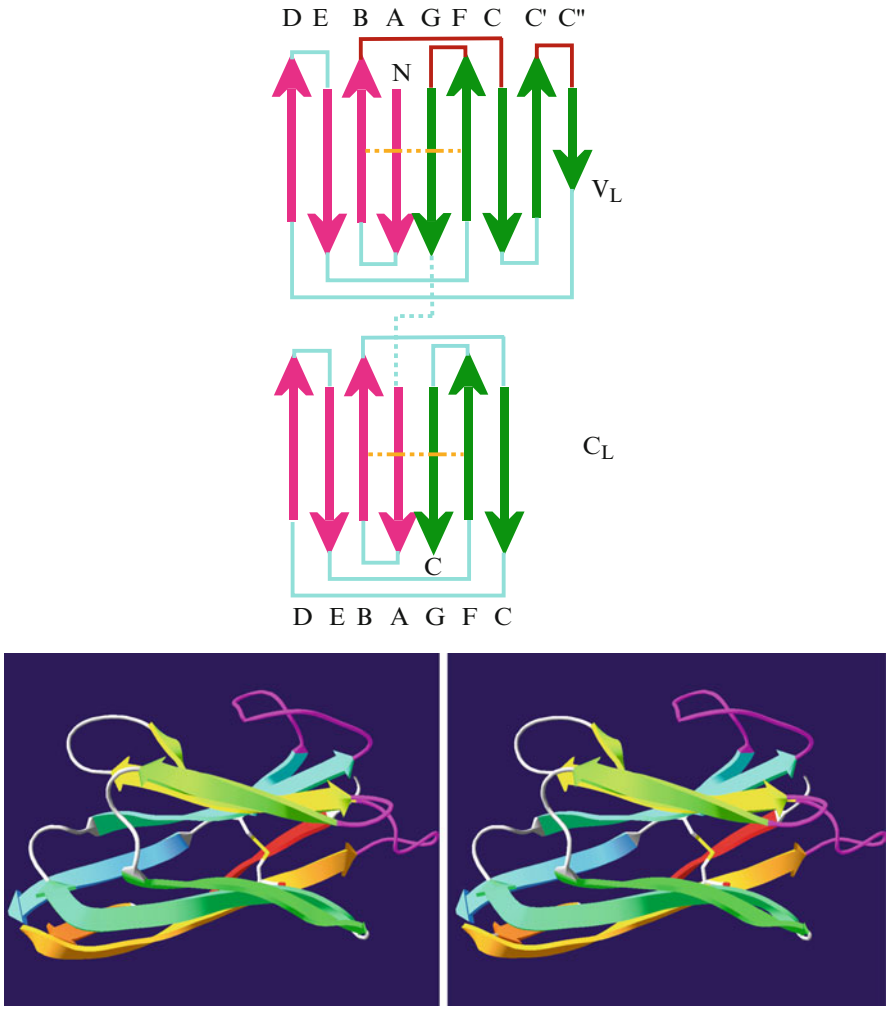


Fig. 11.3 *Top*: The segments of the immunoglobulins all have the same structure, which is known as Ig-fold and is also found in other molecules: it consists of 2 antiparallel β-plaited sheets stacked on top of each other (*purple and green*). The loop segments connecting the β-sheets are shown in cyan, the hypervariable loops that form the antigen-binding side in red. The disulphide bonds are shown in yellow. *Bottom*: Crystal structure of a **nanobody** (PDB-code 3eak, image taken from [9]). Some animals (Camelidae like llamas, dromedaries, or camels, and selachians like sharks and rays) have antibodies that consist of only a single chain. Nanobodies are the variable domains of these antibodies; they have exciting possibilities as laboratory tools and pharmaceuticals. The two β-sheets linked by a disulphide bond that make up an immunoglobulin-domain are clearly visible. The hypervariable loops are shown in purple

express both allotypes. However, due to allelic exclusion (see below), any particular immunoglobulin will be of only one of these allotypes.

Allotype Determination in Medicine

Bone marrow graft monitoring if donor and recipient produce different allotypes
Forensic determination in blood or semen stains
Paternity testing if possible fathers have different allotypes

Interactions between antigen and antibody may be based on ionic, hydrogen, and hydrophobic bonds.

11.2.1.1 There Are Five Immunoglobulin Isoforms

Apart from IgG, there are four other immunoglobulins produced in our body (see Fig. 11.4): IgA, IgD, IgE, and IgM. Some of these occur in several subclasses. The reason for this diversity in molecular structure is the specialisation for different functions, as shown in Table 11.3.

From the clinical point of view, IgE deserves special mention because it is this antibody isoform that is involved in **allergic reactions**. IgE helps protect our body against large parasites, for example, worms. A person with an allergy produces IgE antibodies against an inappropriate target, for example, against a pharmaceutical (penicillin being a notorious example). The ability of antigen-bound IgE to interact with eosinophils and in particular mast cells leads to the rapid release of inflammatory mediators, in the worst case resulting in an **anaphylactic shock**.

IgA is unusual in that it can be secreted across mucous membranes. This can be significant against pathogens that invade through mucous membranes like HIV.

There are two light-chain isoforms (κ and λ), which can occur in any of the Ig-classes. No functional differences between these isoforms has been found to date.

11.2.2 How Is the Large Number of Ig-Molecules Obtained?

It is estimated that our body can produce about 10^{15} different antibodies. This is a very large figure, in particular as our genome contains only about 21 000 different

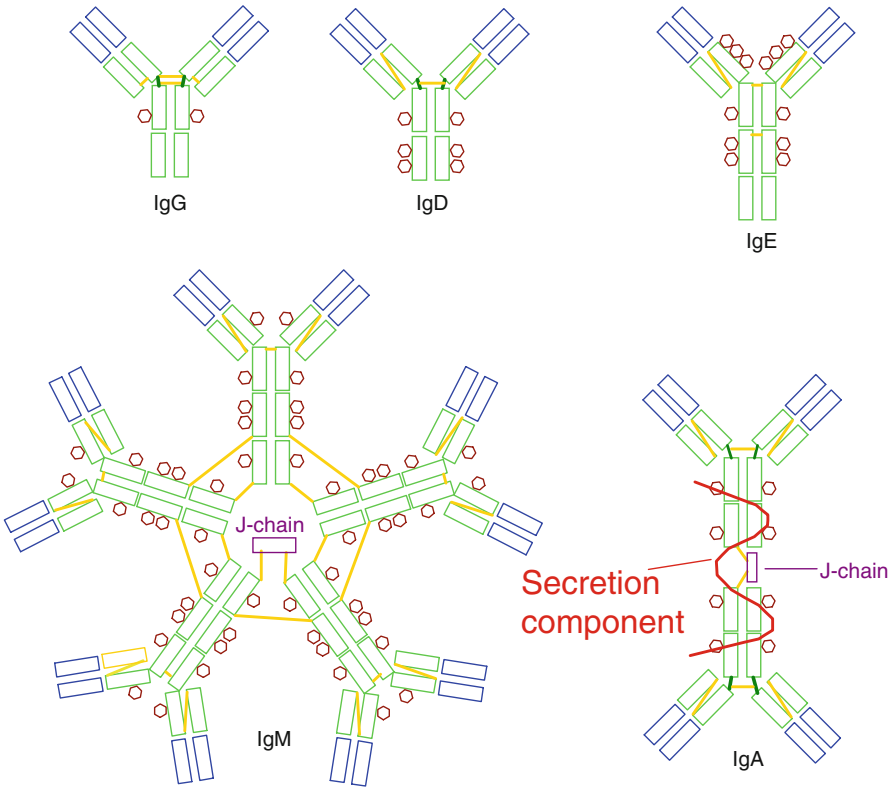


Fig. 11.4 The 5 different immunoglobulins are composed of similar units: variable (*blue*) and constant (*green*) domains, forming heavy and light chains. These are held together by disulphide bonds (*yellow*). The proteins are decorated by sugar side-chains in N-glycosidic bond (*purple hexagons*). IgA and IgM additionally contain a small J-chain (*purple*), IgA also a secretion component

protein-coding genes. Clearly it is impossible to have one separate gene for each antibody. How then are antibodies encoded?

The genetic information for antibodies is contained not in a single chunk, but in small building blocks (see Fig. 11.5). Each of these building blocks is polygenic, and as a B-cell matures, these building blocks are brought together by random **somatic recombination**, resulting in a unique sequence.

The resulting antibodies are presented on the surface of the cell via a short transmembrane segment (removed from soluble antibodies by alternative splicing of mRNA). Antigen binding is signalled into the cell by tyrosine kinases and the cell starts to proliferate. During proliferation mutations are introduced into the sequence of the variable domains, adding to the number of different antibody molecules. Those cells whose mutations lead to higher affinity antibodies have a higher chance of being stimulated into further division by renewed antigen contact, resulting

Table 11.3 Structure and function of immunoglobulin isoforms. Protein A is isolated from the cell walls of pathogenic strains of *Staphylococcus aureus*, Protein G from *Streptococcus ssp.* These proteins bind to the F_c-end of antibodies, resulting in a coat of host protein surrounding the bacteria. This makes them less prone to phagocytosis by macrophages and neutrophils. These proteins have become very useful in the laboratory for the isolation of antibodies and antibody-antigen complexes. Protein L from *Peptostreptococcus magnus* serves the same function, but binds to the κ-light chain and hence to all antibody isoforms. Note that antibody binding to proteins A and G is specified for human antibodies, other species can react differently

	IgA1	IgA2	IgD	IgE	IgG1	IgG2	IgG3	IgG4	IgM
Heavy chain	α_1	α_2	δ	ϵ	γ_1	γ_2	γ_3	γ_4	μ
Light chain	κ, μ	κ, μ	κ, μ	κ, μ	κ, μ	κ, μ	κ, μ	κ, μ	κ, μ
# of protomers	1 or 2	1 or 2	1	1	1	1	1	1	5
Molecular mass (kDa)	160	160	184	188	146	146	165	146	970
Serum level (mg/ml)	3.0	0.5	0.03	5×10^{-5}	9	3	1	0.5	1.5
Half-life (d)	6	6	3	2	21	20	7	21	10
Virus neutralisation	Strong	Strong	No	No	Strong	Strong	Strong	Strong	Weak
Classical complement	No	No	No	No	Moderate	Weak	Strong	No	Strong
Alternative complement	Weak	No	No	No	No	No	No	No	No
$M\Phi$ + Neutrophils	Yes	Yes	No	Yes	Yes	No	Yes	No	No
Mast cells + basophile	No	No	No	Yes	No	No	No	No	No
Natural killer cells	No	No	No	No	Strong	No	Strong	No	No
Passes placenta	No	No	No	No	Strong	Weak	Moderate	Very Weak	No
Secretable	Yes	Yes	No	No	No	No	No	No	Weak
Binds to Protein A	Weak	Weak	No	Weak	Strong	Strong	No	Strong	Weak
Binds to Protein G	No	No	No	No	Strong	Strong	Strong	Strong	No

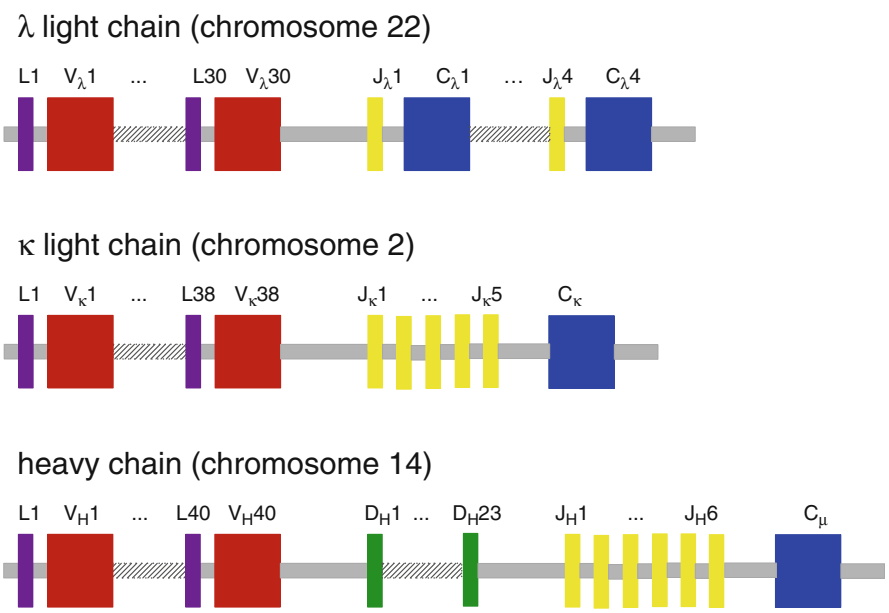
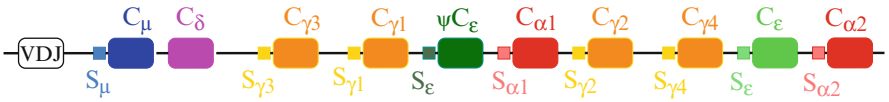


Fig. 11.5 Organisation of Ig-genes on human chromosomes. The genetic information for the λ-light chain is on chromosome 22. There are about 30 different L+V and 4 different J+C regions, resulting in 120 different combinations. The information for the κ-light chain is on chromosome 2. About 40 L+V regions, 5 J-regions and one C-region result in 200 possible combinations. Because each antibody contains either λ- or κ-chains, 120 + 200 = 320 combinations are possible for the light chains. The heavy chain is encoded on chromosome 14. About 40 L+V, 23 D, 6 J, and one C region result in 5520 possible combinations for the heavy chain. Thus the genetic information in our germ line can directly encode 320 × 5520 = 1 766 400 different variable domains



Human immunoglobulin heavy chain gene organisation (in B-cell)

Fig. 11.6 All B-cells start their life producing IgM and IgD, which are made by alternative splicing of mRNA. Once stimulated by antigen binding to the membrane-bound antibody, a B-cell may perform an **isoform switch** by recombination between the switch regions in front of the heavy chain genes. The sequence of the S_{μ}-region is $[(GAGCT)_n GGGGGT]_m$ with n =3–7 and $m \approx 150$. The other switch regions have similar sequences. Note that there is no S_{δ}-region. There appears to have been a gene duplication event, of a unit with two γ-, an ϵ, and an α-gene. One of the ϵ-genes is a pseudo-gene (ψC_{ϵ}), thus only one IgE isoform is expressed in humans

in the selection of antibodies with higher and higher affinities. This is called **affinity maturation**. If creationists argue that nobody ever has observed random mutation and selection—that is, evolution—at work, then here is a counterexample (Fig. 11.6).

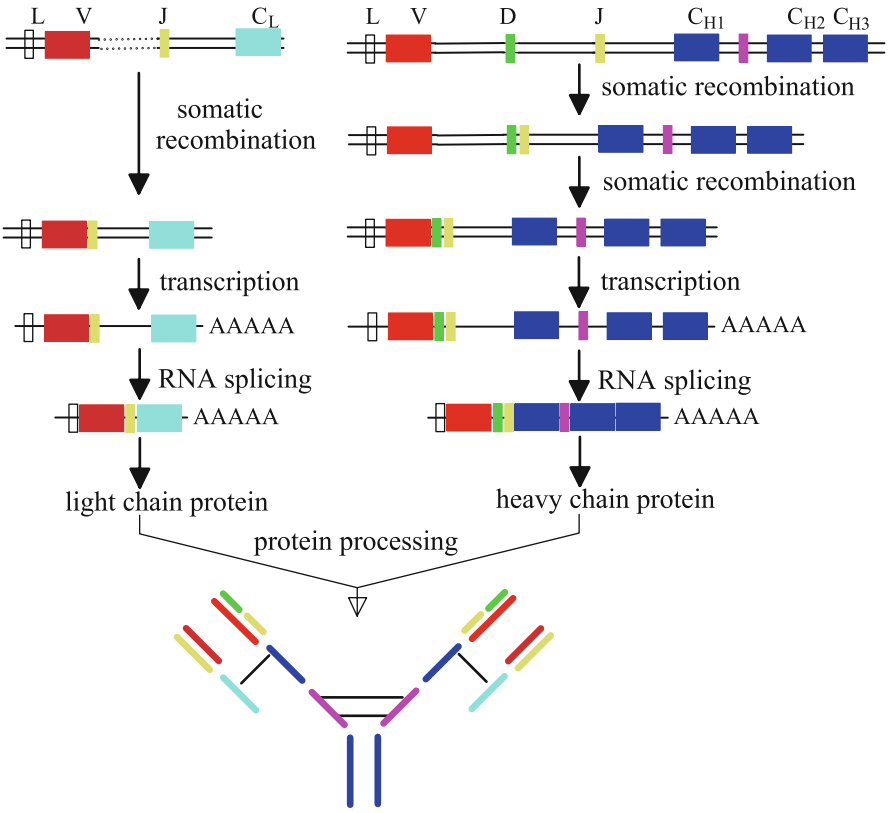


Fig. 11.7 Somatic recombination joins the V-, D-, and J-regions. mRNA-splicing then joins the L- and C-regions. The final antibody is made by joining heavy and light subunits during protein processing

These two processes together lead to the large number of antibodies in our body. Joining the various regions together is done by a combination of somatic recombination, RNA-splicing, and protein-processing (see Figs. 11.7 and 11.8).

Somatic recombination is done in such a way that a variable number of random bases (P and N nucleotides) are added between the regions, greatly increasing the number of possible antibodies (see Fig. 11.9).

A B-cell is, like all other mammalian cells except gametes, diploid. Thus there are two heavy chain and four light chain genes (two κ and two λ), but only one of each gets expressed. This selection is called **allele exclusion**. Recombination occurs first in the heavy chain locus, D- and J-segments recombine on both chromosomes, but only one of them also recombines the V-segment. Recombination of the light chains starts in one κ-locus, if that does not lead to a functional light chain, the second κ-locus will recombine, then the λ-loci, one after the other.

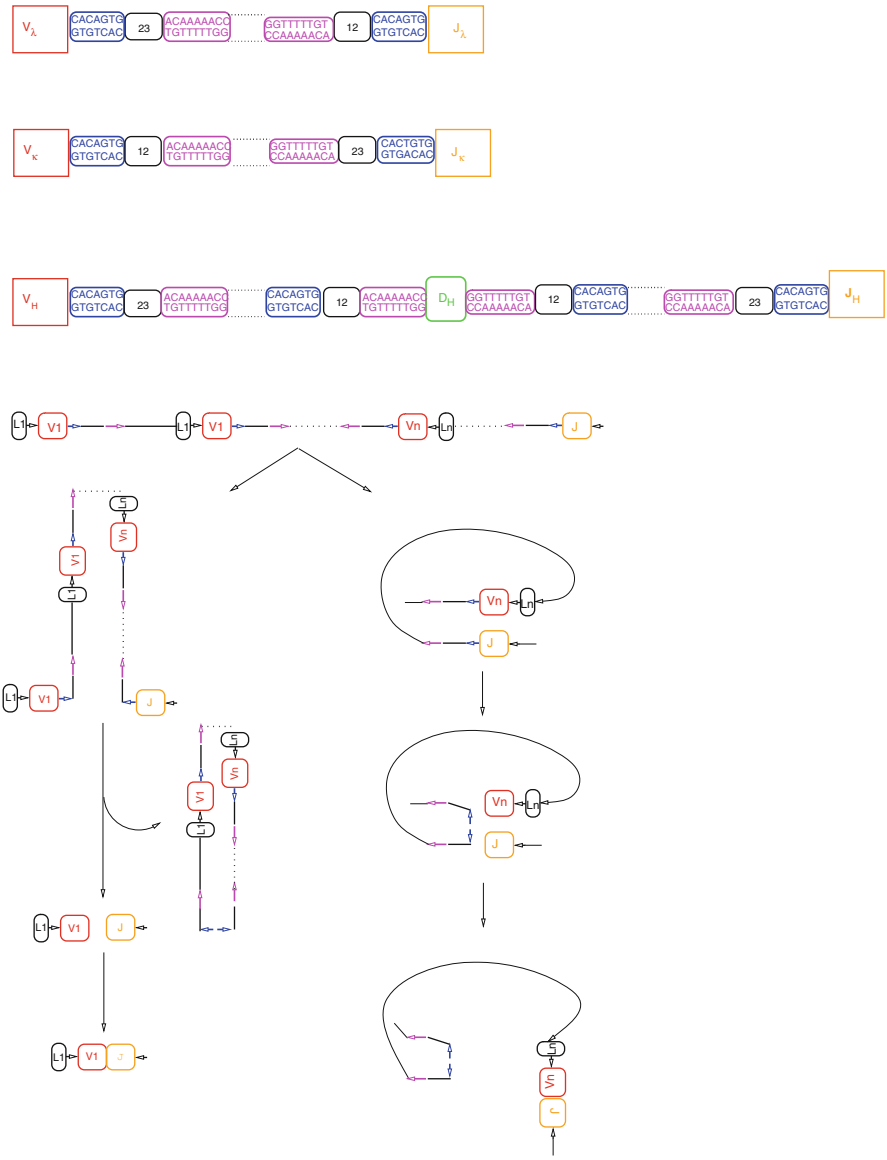


Fig. 11.8 *Top*: The variable regions are flanked by conserved heptamer (7 bases) and nonamer (9 bases) regions, which are needed for recombination. Between those are regions of either 12 or 23 bases. A segment with a 12-base region is joined with a region with 23 bases during recombination. *Bottom*: There are two possible mechanisms of recombination: one results in the loss of the intervening sequences, the other does not

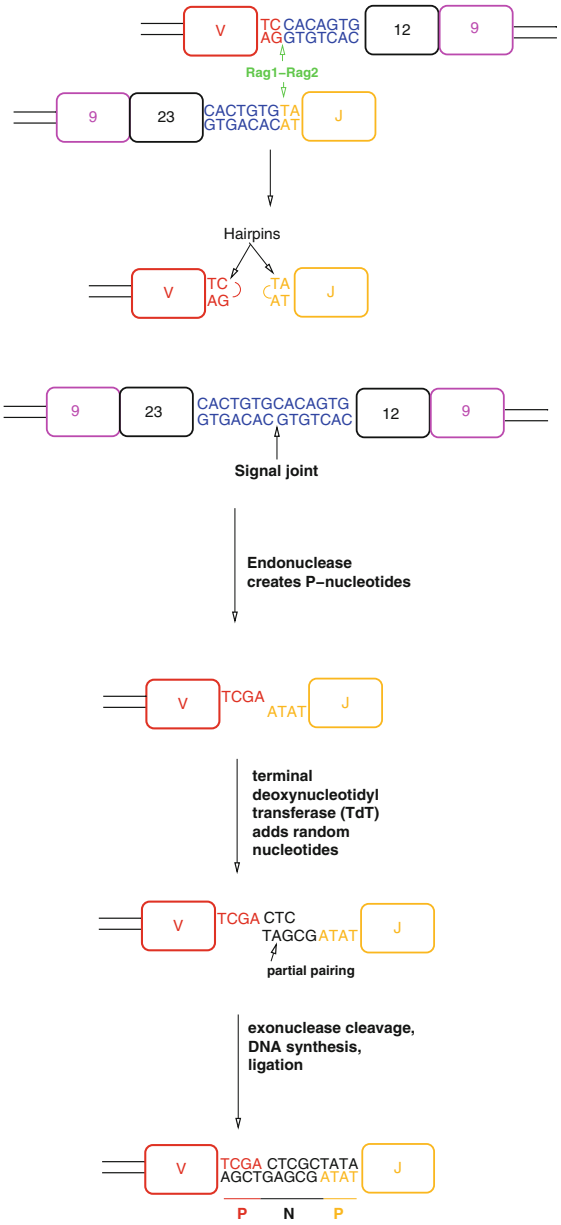


Fig. 11.9 Somatic recombination of Ig-genes requires a set of specialised enzymes. (1) First Rag1-Rag2 (Recombination Activating Gene) recognises the recombination signal and cuts one strand of the DNA at the end of the heptamer sequence. (2) The cut ends then react with the uncut strand, cutting it and forming a hairpin. The two heptamers are ligated into a signal joint. (3) The hairpins are cleaved by endonuclease at a random site, to generate the P- (palindromic) nucleotides. (4) Deoxynucleotidyl-transferase (TdT) adds a variable number of N- (nontemplate) nucleotides, until partial pairing of the ends is possible. (5) Exonuclease trims of any nonpaired ends. (6) The ends are ligated together

Once a functional IgM molecule is produced and expressed on the cell surface, neighbouring stroma cells send out a signal that leads to suppression of Rag1-Rag2 expression in the B-cell, and therefore to an end of recombination. Thus B-cells express only one antibody **idiotype** (sequence of the variable loops). If neither of the recombination events leads to a functional antibody, the cell dies.

Following Affinity Maturation in the Lab
Each clone of B-cells has unique sequences of P- and N-nucleotides, which can be used to follow its descendants through affinity maturation and isoform-switch in experimental studies.

11.2.3 Time Course of Antibody Response

If an antigen enters our body for the first time, it will encounter only few B-cells that by chance express on their surfaces antibodies against this antigen. Binding of antigen to these surface antibodies stimulates these cells to proliferate. At the same time, the cells will start producing soluble rather than membrane-bound antibodies (by alternative splicing), which they secrete into the bloodstream. Such **plasma cells** are protein factories, filled with rough ER. Each plasma cell produces about 2000 antibody molecules per second. This is probably the reason for the short life span of such cells (Fig. 11.10).

These antibodies will invariably be of the **IgM** isoform, their affinity to the antigen will probably be low. However, IgM is pentameric, and there are 10 antigen binding sites per molecule. Most pathogens have highly repetitive antigens on their surface, for example, the coat proteins of virus or the glycosaminoglycan subunits in the bacterial wall. Depending on the distance between antigenic sites and the flexibility of antigens and antibodies multiple binding will occur. Once several of the binding sites of an IgM molecule have bound to a pathogen, release of the antibody requires simultaneous dissociation of all binding sites, which is an unlikely event. Thus the apparent binding strength of an antibody, called its functional affinity, is higher by several orders of magnitude than its single-site affinity would suggest (up to 10.000-fold). This is called the **avidity** (from Latin "greedy") of an antibody (see Fig. 11.11).

Determination of Functional Affinity
Diluted serum samples are pipetted into ELISA-wells (see Fig. 11.13 on page 248) coated with antigen, in duplicate. After binding, wells are washed

(continued)

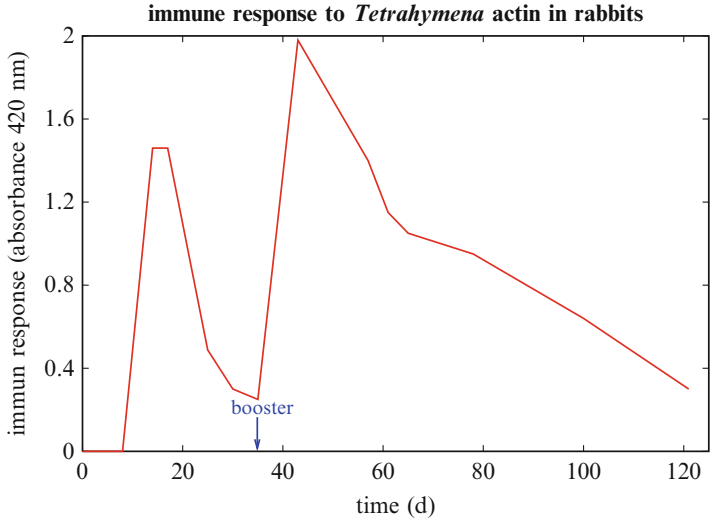


Fig. 11.10 Time course of the antibody response. A rabbit was injected with antigen (here actin from *Tetrahymena ssp.* in complete FREUND's adjuvant) and the concentration of antibodies against actin was measured by ELISA for total rabbit immunoglobulin. A booster injection (with incomplete FREUND's adjuvant) was given on day 36. Note that the response in ELISA is not linear; the change in antibody concentration is much larger than the change in absorbance would suggest. After first contact it takes about 2 weeks for the rabbit to develop a measurable antibody titre against the antigen, and the concentration of antibodies drops rapidly afterwards. However, memory cells remain in the rabbit's blood; a second contact with the antigen results in a quick, long-lasting, strong response. Repeated booster injections could maintain the high antibody concentration (not shown)

with either denaturant (6–8 M urea or 0.5–1 M guanidinium hydrochloride) in washing buffer or with washing buffer alone. Then the amount of bound total immunoglobulin (all subclasses) is determined in both wells. The urea denatures proteins and removes weakly bound antibodies, whilst strongly bound antibodies stay on the antigen. Thus less antibody will be bound in the wells treated with urea than in the wells treated with washing buffer only. Then the avidity index is:

$$\mathrm{AI} = \frac{\mathcal{E}_u}{\mathcal{E}_b} \times 100\,\% \tag{11.1}$$

with $\mathcal{E}$ the measured signal (*e.g.*, optical absorbance, fluorescence or radioactivity) in urea- ($\mathcal{E}_u$) and buffer-treated ($\mathcal{E}_b$) samples, respectively [52].

(continued)

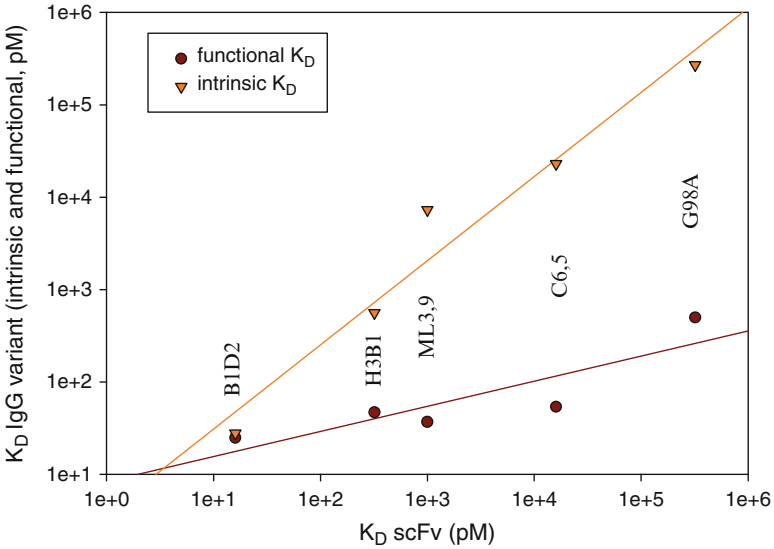


Fig. 11.11 Affinity and avidity of antibodies. In this study [53], single-chain variable fragment (scFv) antibodies (fusion protein of V_H and V_L, one binding site) against human Her2/*neu* were produced by genetic engineering. These had different dissociation constants for the antigen. Then these variable domains were introduced into IgG1 (two binding sites) and both the intrinsic and the functional dissociation constants (K_D) were measured. The intrinsic (K_D) of the full IgG was very similar to that of the corresponding scFv. The functional (K_D), however, was much lower, corresponding to higher affinity

For more reliable results, binding curves (absorbance vs. antibody dilution) are compared, rather than single absorbance values [28]. Alternatively, avidity is determined with several different concentrations of denaturant. Note that the outcome of this type of assay is determined not only by avidity, but also by affinity. The latter can be measured separately using F_{ab}-fragments of the antibody in question. For a discussion of the biophysical basis of avidity see [13, pp. 32–41]

As the B-cells proliferate, somatic hypermutation in the variable region of the immunoglobulin gene creates a large number of different antibodies; those cells that produce higher affinity antibody (on their surface) have a higher chance to be activated by antigen binding and to proliferate further. Thus antibodies with higher and higher affinity are created. After affinity maturation cells switch from producing IgM to other isoforms, in particular **IgG**. IgG has only 2 binding sites; its avidity is only about 10-fold higher than its affinity. However, because of its higher affinity, this is sufficient. Antibody-antigen complexes have dissociation constants of between about 10 µM and 1 pM.

Medical Uses of Functional Affinity Measurements
Early in infections, the antibodies produced have low affinity but high avidity (IgM). In chronic infections, high-affinity, low-avidity antibodies (mainly IgG) are produced. In addition, the intrinsic affinity of antibodies increases during maturation. The distinction between acute and chronic infections is made by comparing the avidity index obtained for a patient sample with cut-off points established by the test manufacturer.

Avidity of a single antibody isoform (say, IgG) in long-term infection (say, with HIV) can change over time if the steric orientation of the antigen binding sites changes to allow all binding sites to bind to the pathogen at the same time. Such changes may occur in response to escape mutations in the pathogen [19].

If the antigen is no longer present, the antibody-producing plasma cells die within a few days, and the serum antibody concentration drops quickly to undetectable levels. However, not all descendants of the stimulated B-cells developed into plasma cells. Others became **memory cells**, which have a life span of many years. If these encounter the same antigen again, they can mount a rapid, strong, high affinity and long-lasting antibody response (see Fig. 11.10). For this reason we can get many diseases only once: if we overcome the initial infection, we are protected from reinfection for the rest of our life. Some pathogens, however, in particular virus such as influenza, can overcome this protection by rapidly changing their surface proteins.

11.2.4 Immunisation

The observation that one contact with many pathogens is enough to result in long-term protection against them is used for the purpose of **immunisation** (for a demonstration of its effectiveness, see Fig. 11.12). In principle, there are three ways to accomplish immunisation:

- **Passive** immunisation is achieved by transferring antibodies from one individual to another. Only the proteins, but not the memory cells are transferred, therefore protection lasts only a few days. This method is used to intervene with an acute crisis (e.g., by injecting **anti-snake venom** antibodies raised in animals such as horse after a snake bite). Another example for this technique is the transfer of serum from a patient who just

(continued)

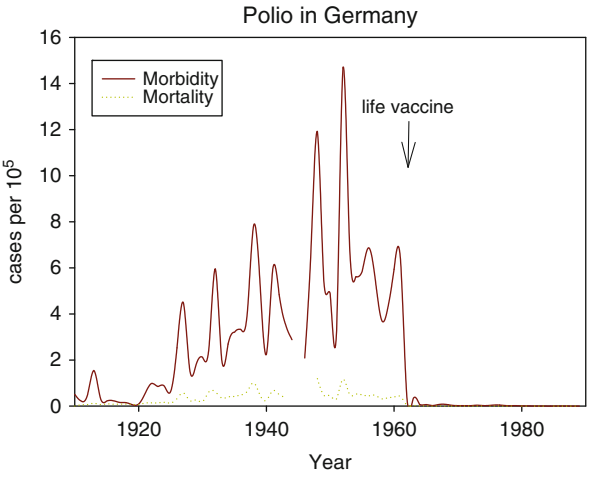


Fig. 11.12 Morbidity and mortality from polio (*Poliomyelitis epidemica anterior acuta*) in Germany (data from [46], numbers from West and East Germany were combined during 1949–1989). Polio is a picorna virus infection that in a natural environment is spread by the fecal-oral route to infants where it causes a bland intestinal infection followed by lifelong immunity. In modern society hygiene delays infection to school age, when it may spread in 1 % of cases to the α-motoneurons of the spinal cord and to brain nerves IX and X. This causes paralysis of limbs or, even worse, the muscles of breathing. The disease was first described by HEINE in 1840 [24]. The morbidity from polio increased during the first half of the twentieth century (with a short observation gap after World War II), until mandatory immunisation (developed by ALBERT SABIN) was introduced in 1962. This led to a drastic drop in incidence. Today polio is still endemic in Nigeria, the Democratic Republic of Congo, Chad, India, Pakistan, Tajikistan, Afghanistan, and Myanmar. Because humans are the only reservoir immunisation may lead to eradication of polio as it did for smallpox

survived an infection to another patient suffering from the same infection (**hyperimmune serum**).

- **Active** immunisation is achieved by injecting the antigen so that a lasting (in ideal cases lifelong) protection is achieved. Dead pathogens or their purified components can be used as **vaccines**, however, even better results are often achieved by pathogens which are still alive, but have lost the ability to cause disease (**attenuated** pathogens). Genetic engineering allows the production of vaccines where surface markers of very dangerous pathogens, which are not available in attenuated form, have been inserted into relatively harmless organisms such as **vaccinia virus**.[a] Antigens are often injected together with an **adjuvant**, a substance that increases the immune response of our body to the antigen. **Anti-idiotypic antibodies**; that is, antibodies against the variable loops of an antibody, are also

(continued)

considered as possible active vaccine. The idea is that the antigen binding site of an antibody is like a mold of the original antigen. An anti-idiotypic antibody would therefore resemble that antigen and could be used in cases where attenuated pathogens are not available (currently (autumn 2014) discussed for vaccination against Ebola virus).

- **Adoptive** immunisation requires the transfer of spleen cells from an immunised individual to another. So far, this procedure has been used only for experimental purposes.

11.2.5 Monoclonal Antibodies

The high affinity and specificity of antibodies makes them ideal tools in research, diagnostics, and even treatment of disease. It would therefore be useful if a particular antibody could be produced in large quantities.

As we have seen, antibodies are produced by plasma cells, which have a life span of only a few days. They are also **terminally differentiated** and no longer divide. At any given time, there are many different plasma cells active in our bodies; the immunoglobulins in our plasma are directed against many different antigens.

MILSTEIN & KÖHLER have found a way around these problems (NOBEL-price 1984 together with JERNE). They isolated spleen cells from a mouse immunised against a particular antigen. They fused these cells, which included antibody-producing cells, with cells from a **myeloma** of a mouse from the same strain. Some of the fusion products had both the ability to produce antibodies against the antigen in question (from the spleen cells) and to divide indefinitely (from the tumor cells). Such cells were isolated and cloned. Because each B-cell produces only antibodies with a particular variable region (i.e., a particular **idiotype**), the antibodies produced by such clones are homogeneous; they can be isolated from the culture medium. Such antibodies are called **monoclonal** because they are produced from a single cell clone.

It is of course more difficult to get spleen cells and matching myeloma cells from human. However, using techniques of molecular biology, it is possible to replace the genetic information for the constant region of IgG in those cloned mouse cells by that for human IgG, so that the resulting cells will produce **humanised** antibodies. Such antibodies can be injected into patients repeatedly without causing an immune reaction against mouse IgG, and are therefore useful as **therapeutic agents**.

[a]Vaccinia is the causative agent of cowpox, and is relatively harmless in immune-competent humans. It was used for immunisation against **smallpox** (see Fig. 11.11) until this terrible disease was declared extinct, one of the greatest successes of man's fight against disease.

By the same method it is also possible to change the isoform of an antibody, which can be useful in allergy research.

Indeed modern molecular biology techniques allow the whole process of somatic recombination, hypermutation, and selection to be done entirely *in vitro* producing the DNA for a perfectly good antibody (from any species) against any antigen without the use of research animals. This DNA is then integrated into a cultured cell line for antibody production, or even into bacteria such as *E. coli*. This allows researchers to raise antibodies, for example, against self-antigens, that could not be raised in animals.

Nomenclature of Pharmaceutical Antibodies
Pharmaceuticals based on monoclonal antibodies have names ending on -mab, with the letters in front denoting origin:

omab mouse
imab primates
ximab chimera: variable part mouse, constant part human
zumab humanised antibody; only the antigen binding loops are mouse
umab human

Most of these antibodies are directed against cancer, but some are used against autoimmune, infectious, protein folding, or inherited diseases.

11.2.6 Laboratory Uses of Antibodies

Because antibodies bind specifically and with high affinity to antigens, the antibody-antigen reaction can be used to detect either antibodies or antigens, both in research and in clinical laboratories. Considerations for assay development have been reviewed in [11].

Figure 11.13 shows the basic principles involved in immunological assays. The exact procedure followed depends on the sort of information required. Assays that detect only the presence of an antigen are usually simple, fast, and cheap, however, if additional information such as the molecular mass of the antigen is required, costs can increase significantly. It is therefore not uncommon to use simple assays for screening purposes and the more expensive ones for closer investigation of only those samples that were identified in the screen.

The following main assay types are used:

Competitive: The sample is added together with a known amount of labelled antigen to a known amount of antibodies. The labelled antigen binds to the antibody, resulting in a defined signal. If the sample also contains (unlabelled) antigen, this would compete with binding of the labelled antigen molecules, reducing the signal measured in a concentration-dependent manner. Competitive

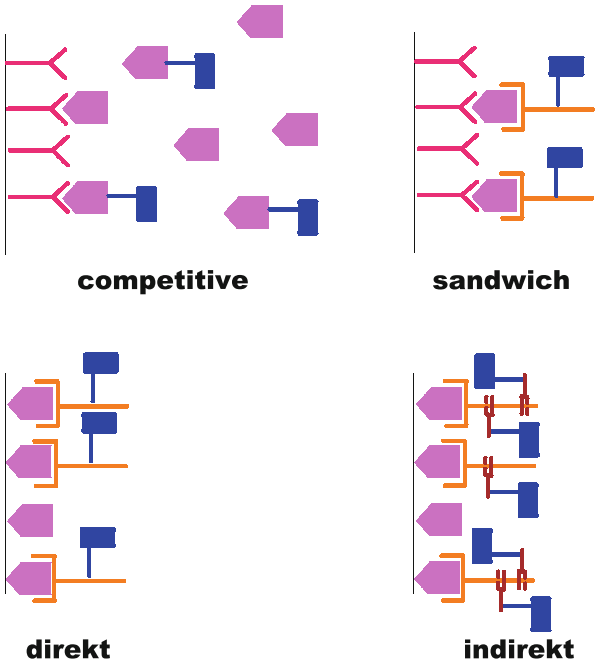


Fig. 11.13 Basic principles of immunological assays. For details see text

assays are often used for the detection of small molecules, such as drugs. The problem with that type of assay is to label the antigen without destroying antigen-antibody interaction.

Sandwich: Antibodies are fixed on a solid support, then incubated with the antigen. Antigen binding creates binding sites for a second antibody, directed against a different epitope of the antigen, thus the amount of second antibody bound is proportional to the concentration of antigen. This procedure works only with fairly large antigens (with several epitopes); in addition it requires antibody pairs, which are not always available.

Direct: The labelled antibody binds to the antigen, which is fixed on a solid support. The amount of antibody bound and hence the signal depends on the amount of antigen present. This type of assay is often used against large antigens including proteins, virus, bacteria, or even whole cells.

Indirect: assays work similar to the direct ones, however, the primary antibody is unlabelled. After it has bound to the antigen, a secondary antibody directed against the constant part of the primary antibody bears the label. Because each primary antibody can bind several secondary ones this leads to an increase in sensitivity. Additionally, the complicated labelling procedure needs to be performed only once in the secondary antibody (obtained from large animals like sheep or horse by manufacturers), which can then be used to detect several primary antibodies (obtained in the lab from small animals such as rats).

Various different labels may be used to detect the antibody; the most important are radioactive isotopes, fluorescent molecules, and enzymes. Note that the assays described above can easily be set up to detect antibodies, rather than the antigens.

Development of immunological assays for drugs, disease marker proteins, or for pathogens is a multibillion € industry, and has especially revolutionised clinical laboratories. Fluorescence-activated cell sorter (FACS) (see Fig. 11.14) can be used to not only count, but also isolate cells expressing a particular marker for further use. In the last couple of years it has been complemented by the detection of nucleic acids, for example, by Polymerase chain reaction (PCR).

Fast Assays for Small Molecules

The **Enzyme multiplied immunoassay technique (EMIT)** works a bit like a competitive assay, in that labelled antigen from the test kit and unlabelled antigens present in the sample compete for a limited amount of antibody. The label is an enzyme, which is inactivated when the antigen is bound to an antibody. Sample, labelled antigen, and antibody are mixed, given time to react (in some cases, just 1 min), and then the substrate of the enzyme is added; enzymatic activity is detected by the amount of colour produced. If the sample contained little of the unlabelled antigen, most of the labelled one will be bound by the antibodies, the enzyme is inactivated, and no colour is produced. If the sample contained a lot of unlabelled antigen, then it competes with the labelled antigen, the enzymatic activity is high, and colour is produced. No separation between bound and unbound antigens is required, hence this type of test is very quick.

Such assays are used, for example, in the emergency department to test unconscious patients for exposure to street drugs. They can also be used to monitor patients on medications with a small therapeutic window (difference between minimal effective and toxic dose) such as digoxin or coumarin analogues.

11.3 Destroying Invaders: The Complement System

Blood and interstitial fluid contain a group of proteins that can destroy cell membranes. Collectively, they are known as **complement system** [43]. The name indicates that these proteins complement and enhance the disease-preventing activity of antibodies and immune cells. There are about 30 different proteins in the complement system; they are produced mostly in the **liver**, but also in parenchyma of the **kidney** (endothelial and epithelial cells of glomeruli, proximal tubular and mesangial cells). Complement production in the renal tubular interstitium, where

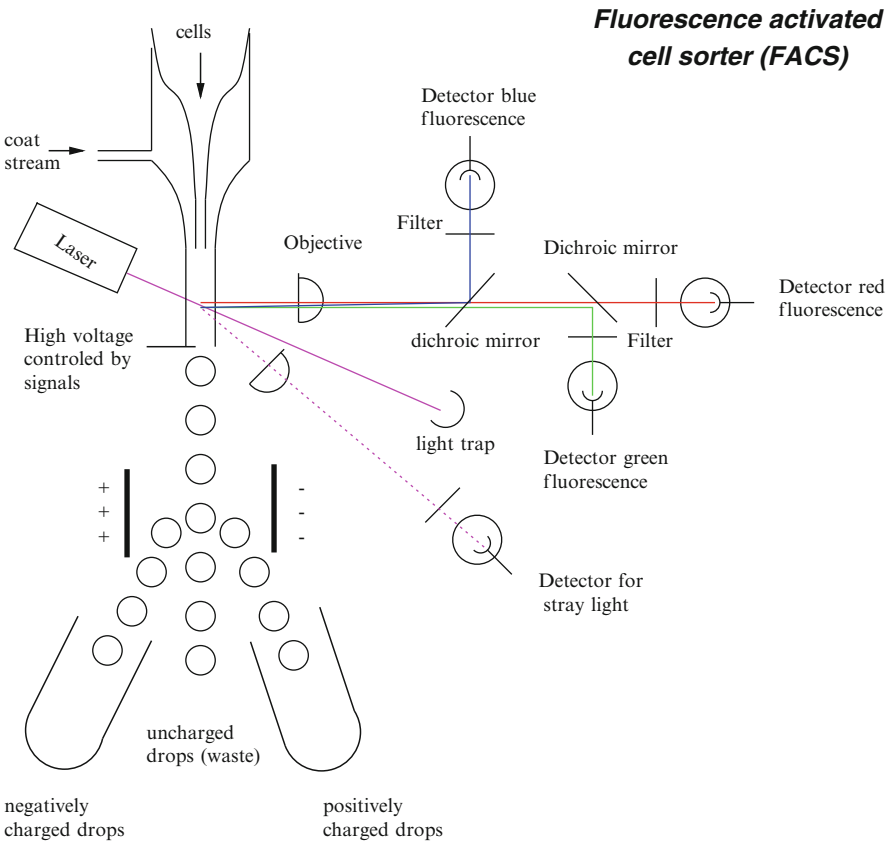


Fig. 11.14 A Fluorescence-activated cell sorter (FACS) can be used both to quantify and to isolate a particular cell type. Cells are incubated with fluorescently labelled antibodies against markers for that cell type, say, CD4 for T_h-cells. Then the cell suspension is pressed through a nozzle that allows passage of only a single cell at a time. The nozzle is mounted inside a tube through which buffer flows at a higher velocity than that of the cell suspension (coat stream). Thus the cells are passed under a microscopic objective like pearls on a string. They are illuminated by a laser, the resulting fluorescent light is separated by wavelength, and the intensity measured by photomultipliers. Modern instruments can routinely look at five wavelengths simultaneously; in addition they measure stray light as an indicator of cell size. Once the stream of cells has passed the microscope it exits as a stream of droplets, each droplet containing a single cell. If a cell expressing the desired marker(s) has been detected, the droplet containing it is charged by a high voltage pulse and redirected by electrical fields into collection vessels. Thus it is possible to isolate even very rare cells from large volumes of sample, for example, pluripotent stem cells from peripheral blood. This image was taken from [9]

little contact with blood plasma occurs, may protect against infections ascending through the urinary tract. C1q is synthesised mainly in **mononuclear phagocytes**, and CFD in **adipocytes**.

Obviously, such proteins are very dangerous, and their activity needs to be tightly controlled. The early components of the complement system are highly specific serin-proteases; activation of the complement system is a chain of protein cleavage reactions. Apart from activating another protease, each cleavage results in the production of a small peptide, which acts as **inflammatory mediator (anaphylatoxin)**.

Complement

- **opsonises** cells for destruction by macrophages and neutrophils
- catalyses the formation of the **membrane attack complex** from components C5–C9. A complex of C5–C8 catalyses the oligomerisation of C9, which forms pores of about 10 nm diameter in the cell membrane.
- causes **inflammation** by production of peptides with **anaphylactic** activity.

Unfortunately, components are numbered in order of their discovery rather than their place in the cascade. This makes studying the complement system "a little" confusing.

11.3.1 How Is Complement Activated?

There are four sequences of complement activation, known as **classical**, **alternative**, **lectin**, and **extrinsic** pathway, respectively.

The **classical pathway** (see Fig. 11.15) is started by IgM or IgG molecules bound to a pathogen.

In the **lectin pathway** the antibody-antigen complex as starting point is replaced by either Mannan binding lectin (MBL) or by a ficolin. These bind MBL associated serine protease (MASP) which cleaves C4 and C2 just as C1s does in the classical pathway.

> Deficiencies in MBL (OMIM #614372) or MASP (OMIM #613791) lead to increased respiratory tract infections during early childhood, when the protection of maternal antibodies is gone, but the child's own acquired immunity not yet fully developed [10].

There are three known types of ficolins in humans:

M-ficolin (*FCN1*) recognises acetylated sugars [21] and dying host cells.

L-ficolin (*FCN2*) recognises acetylated sugars such as GlcNAc or GalNAc, in particular lipoteichoic acids from GRAM-positive bacteria [34, 35].

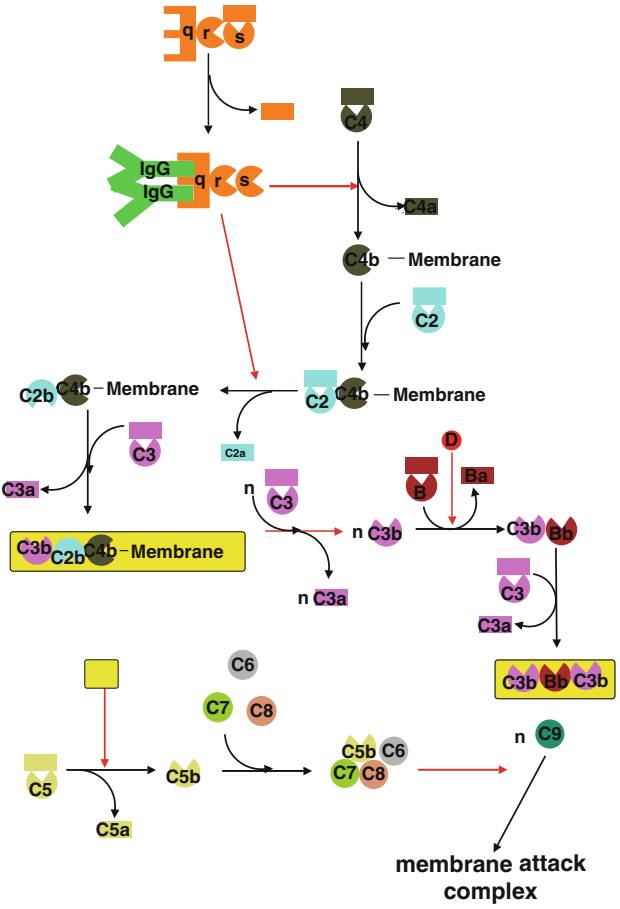


Fig. 11.15 Complement activation. Top Classical pathway: The C1q,r,s complex can bind to immunoglobulins bound to antigen. At least 1 IgM or 2 IgG are required. Binding of the hexameric C1q protein to Ig results in C1r cleaving and activating C1s. Activated C1s can cleave C4 into C4a (a weak inflammatory peptide) and C4b, which binds to the membrane. This binding is covalent: cleavage of C4 exposes a reactive **thioester** bond (between Glu and Cys) in C4b, which reacts with OH- or NH_2-groups in proteins and carbohydrates forming ester or amides and the free thiol-group of Cys. C4b then recruits C2, which in the C4b-bound state is also cleaved by C1s into C2a (inflammatory peptide with moderate activity) and C2b. The C2b,4b complex recruits C3, which is cleaved to C3a (inflammatory peptide) and C3b. C3b can stay bound as C2b,3b,4b complex, but as each C2b,4b complex cleaves up to 1000 C3 molecules, noncomplexed C3b is also formed. Middle Alternative pathway: C3b binds factor B, which is then cleaved by factor D, a soluble protease. This results in the formation of a C3b,Bb complex, which also has C3-convertase activity (i.e., splits C3 into C3a and C3b). This leads to a rapid autocatalytic flooding of the pathogen's surface with C5-convertase. *Bottom*: C2b,3b,4b, and $C3b_2Bb$ both have C5-convertase activity; they recruit C5 to the membrane, where it is split into C5a (a highly active inflammatory mediator) and C5b. C5b then binds C6, C7, and C8; the latter inserts into the membrane and catalyses the oligomerisation of 10-16 C9 molecules into a large transmembrane pore. In endothelial cells, C9 induces the release of chemokines and the production of adhesion molecules, and hence inflammation

H-ficolin (*FCN3*) recognises D-fucose and galactose. It protects the respiratory tract against infections with *Aerococcus viridans* and is involved in clearing late apoptotic cells [26, 54]. In SLE auto-antibodies against H-ficolin (Hakata-antigen) may be found. A deficiency in this protein has been described (OMIM #613860).

The **alternative pathway** starts with spontaneous cleavage of C3 into C3a and C3b, the latter inserts into membranes. Our own cell membranes have C3b-inactivating components, but some pathogens lack those. Thus C3b can hydrolyse B and form the C3b,Bb-complex, which leads to activation of further C3-molecules. The C3b,Bb complex is the functional homologue of C2b,4b in the classical pathway; C2 and B are encoded next to each other in the MHC-locus. Both complexes can bind C3b to form a C5-convertase (C2b,3b,4b or $C3b_2Bb$).

In the **extrinsic pathway** C5 is activated by proteins of the coagulation cascade, factors IXa, Xa, and XIa, plasmin, and thrombin. Once the C5-convertases have been formed, alternative and lectin pathways proceed just as the classical.

Note that the conversion of complement molecules works as a catalytic cascade quite similar to what is seen in hormone signalling (see Fig. 8.1 on page 186). A small number of C1-molecules bound to antigen-antibody-complexes leads to the rapid production of large amounts of C3b and C5b, flooding the surface of the invader. The cascade leads to a strong, all-or-nothing response.

Cleavage of complement molecules converts a water-soluble protein without enzymatic activity into a membrane-bound Ser-protease and a soluble peptide (anaphylatoxin). This protective mechanism is also observed, for example, with digestive enzymes.

In evolution the alternative pathway probably developed first as part of our innate defense. The lectin- and classical pathway are later additions to direct complement more specifically and with higher activity to foreign material. Indeed, proteins with thioester-bonds related to C3 are found even in coelenterates; they hence predate the development of bilateral symmetry about 1.6 billion years ago [42].

11.3.2 What Does Complement Do?

C4a, C3a, and C5a are potent activators of inflammation (**anaphylatoxins**). They induce smooth muscle contraction, increase vascular permeability and recruit granulocytes and macrophages to the site of infection. In addition, C5a and to a lesser extent C3a stimulate the migration, endothelial cell binding, phagocytic response, and oxydative burst of macrophages, neutrophils, and eosinophils by binding to 7 membrane-spanning domain type of G-protein coupled receptor (C5aR and C3aR, respectively). In basophils these receptors trigger the release of histamine. This is particularly important early during an infection, as there is no F_c-receptor for IgM. In T-cells activation of C5aR and C3aR strengthens interaction with antigen-presenting cells, increase proliferation, decrease apoptosis,

and increase interferon γ production. In antigen-presenting cells C5aR and C3aR activation stimulates the release of cytokines such as interleukin-12 and -13.

There is a second receptor for C5a, called C5L2. Binding of C5a to this receptor downregulates the response to C5aR; it may also serve to bind and neutralise circulating C5a. C5L2 does not work through G-proteins.

Some cells in our body have complement receptors that recognise pathogens opsonised with complement. Four such complement receptors (CR) have been described:

CR1 (CD35) occurs on **erythrocytes**. immune complexes left over after an infection triggers the complement system, covering them with C3b and C4b. This complex binds to the erythrocyte, in liver and the marginal zone of the spleen; the immune complexes are then abstracted from the erythrocytes by specialised macrophages without destroying the erythrocytes. By this process potentially dangerous immune complexes (which can cause rheumatic diseases, hence the name "**rheuma factor**") are removed from the bloodstream.

CR2 occurs on B-cells and binds C3d, the final product of C3 activation. B-cells bound to pathogens covered with C3d have a lower threshold for activation by their antigen receptor. B-cells also express CR1, binding of complement to which seems to inhibit B-cell activation.

CR3 occurs on myeloid cells. Binding of iC3b to CR3 induces phagocytosis. Some species of *Staphylococcus* can bind directly to CR3, without complement.

CR4 occurs on lymphoid and myeloid cells and on tissue macrophages and binds iC3b. This stimulates phagocytosis.

In the kidney, the function of CR1 on podocytes and CR3 and -4 on dendritic cells is unknown.

Autoimmune Disease in Patients with Deficiencies in Components of C1
CR1 on macrophages binds C1q on apoptotic cells; this marks those cells for unobtrusive removal. Homozygous mutations in either of the 3 subunits of C1q (called A-, B-, and C-chain) predisposes patients to SLE-like autoimmune diseases (OMIM #613652). Both nonsense-mutations and amino acid substitutions leading to nonfunctional protein have been described. Patients with deficiencies in C1r and C1s are similarly affected (OMIM %216950).

The membrane attack complex C5b-9 causes cell death. However, it appears that this does not occur only—or even mainly—by cell lysis as was once thought. The exact mechanism of cell killing by C5b-9 is still unclear.

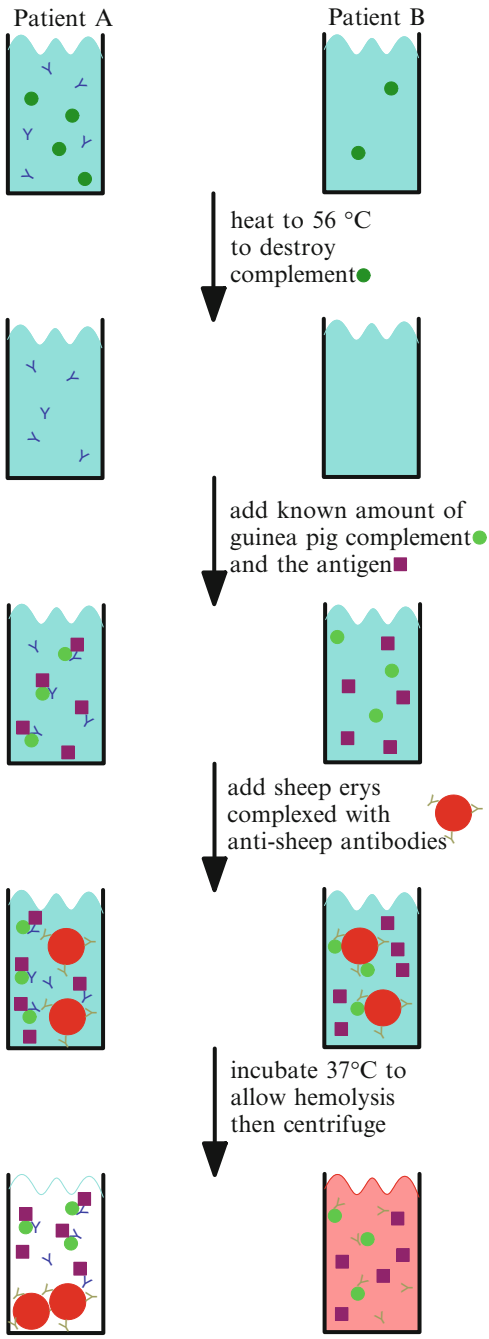


Fig. 11.16 Complement fixation assay for disease-related antibodies as originally developed by WASSERMANN and coworkers [55]. In samples that do not contain antibodies against the antigen the complement is available to lyse the erythrocytes, and the supernatant will be dark red. In samples that contain such antibodies, formation of a ternary antigen-antibody-complement complex reduces available complement and fewer, if any, erythrocytes will be lysed. The supernatant will be pale

Deficiencies of C5 (#609536), C6 (#612446), C7 (#610102), C8 (#613790) and C9 (#613825) do not lead to overt immunodeficiencies. However, these genotypes are enriched in patients suffering from some bacterial infections, for example, *Meningococcus*.

11.3.3 How Is Complement Inactivated?

Activated complement is unstable and degraded rapidly by carboxypeptidases in plasma that cleave off the C-terminal Arg. C5a*des*Arg has a 100-fold lower affinity for C5aR, whilst the affinity for C5L2 is unchanged. C3a*des*Arg no longer binds to C3aR.

There is a complement inhibitor circulating in the blood (C1 inhibitor, C1INH), which dissociates C1r and C1s from C1q. C1q remains bound to the immunoglobulins on the pathogens, but can no longer activate C4. C1INH also inhibits MASP-1 and -2 in the lectin pathway.

Inability to form C1INH leads to **hereditary angioneurotic œdema** (OMIM #106100). C1INH inactivates not only the classical and lectin pathway of complement inactivation, but also factor XIIa and kallikrein in the fibrinolytic pathway. Thus in C1INH deficiency excess bradykinin is produced, resulting in vascular permeability and hence œdema.

Factor I (CFI) (see Fig. 11.17) is a serine protease in plasma that degrades C3b first to iC3b and then C3dg and C3c, these fragments are inactive as C5-convertase. CFI does not act alone, but requires protein cofactors, one of them being **factor H (CFH)**, a glycoprotein which is recruited to our cells by binding to sialic acid, a component of glycoproteins and glycolipids that is absent from many pathogens. Another such cofactor for CFI is **membrane cofactor protein (MCP, CD46)**, a membrane protein that protects only those cells able to produce it. Complement receptor 1 (CR1, see above) also activates CFI.

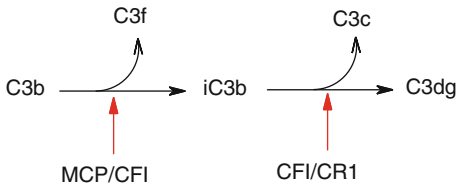


Fig. 11.17 Degradation of C3b. For details see text

Missing or nonfunctional CFI or its cofactors leads to excessive activation of the alternative pathway and hence secondary complement deficiency (OMIM #610984).

If C4b does not form a covalent bond with the pathogen's membrane, its thioester is hydrolysed by water and C4b is inactivated.

Decay-accelerating factor (DAF, CD55) is a type 6 membrane protein that accelerates the dissociation of the C3bBb complex and thereby down-regulates the alternative pathway. In the same manner it down-regulates classical and lectin pathways by dissociating C4bC2b, as does **C4-binding protein (C4BP)**.

Vitronectin (Protein S) prevents the polymerisation of C9 by binding to the C5b-7 complex; the resulting SC5b-9 complex terminates the complement cascade. Similarly, clusterin prevents membrane insertion of C5b-7 and inactivates C8 and C9.

Paroxysmal Nocturnal Hæmoglobinuria
(MARCHIAFAVA–MICHELI syndrome, OMIM #300818) CD59 is widely expressed on our cells and prevents C9 binding to C5b-8 and hence the formation of a membrane attack complex. This protects our cells from "stray bullets". Inability to synthesise the glycosyl phosphatidylinositol tail that binds CD59 (and DAF) to the cell membrane causes **paroxysmal nocturnal hæmoglobinuria**, a very serious disease that leads to the destruction of erythrocytes by complement, explaining the hæmoglobinuria. The monoclonal antibody **Eculizumab**, which blocks the conversion of C5 to C5a and C5b, is used to treat this disease. "Paroxysmal" means suddenly recurring or intensifying. As the OMIM-number indicates, the disease is X-linked.

Pathological Complement Activation
Overwhelming complement activation can occur in sepsis, reperfusion injury and in some chronic diseases leading to apoptosis or necrosis. In this situation, complement components, especially MAC, are deposited in the glomeruli; this leads to cytokine and collagen production and to fibrosis and apoptosis. In addition, there is immune cell infiltration. Clinically, this impresses as glomerulonephritis, membranous nephropathy, and tubulointerstitial nephropathy.

11.4 Cellular Immunity

Antigen-presenting cells present antigens to T-cells bound to specific cell-surface proteins, called MHC-I[2] and MHC-II (Major Histocompatibility Complex, also known under their German name *Hauptlymphozytenantigen*, HLA).

T-cells have a receptor for the MHC-antigen complex, the T-cell receptor. This name is somewhat unfortunate, as the receptor does not bind to, but is located on, T-cells.

11.4.1 The Major Histocompatibility Complex

11.4.1.1 Sampling Proteins Produced in Our Cells: MHC-1

Our cells constantly present samples of the proteins produced by them on their surface; this status is read by T-killer cells. Cancer cells (characterised by expression of embryonal antigens) and virus-infected cells are killed, hopefully before more damage occurs.

Ubiquitin and the Proteasome

Proteins no longer needed in our cells are marked by poly-ubiquitination (see Sect. 2.6.14 on page 57 for a discussion of ubiquitin).

The 26 S proteasome (see Fig. 11.18) consists of the 20 S subunit, a stack of 4 rings made of 7 proteins each. The proteasome is an ATP-dependent protease. Both ends of that stack are covered with a 19 S cap,[3] which recognises ubiquitinated proteins and acts as an antichaperone, unfolding proteins so that the amino-acid thread can pass through the pore of the 20 S-subunit. The entire proteasome has a mass of $\approx$ 2.5 MDa and is large enough to be found in 3D X-ray microscopic studies of yeast cells [50].

For cells protein production and destruction are a big number game. About 4×10^6 protein molecules are produced each minute on the 6×10^6 ribosomes of a mammalian cell, resulting in a total of 2.6×10^9 protein molecules per cell. Of the 4×10^6 protein molecules produced per minute, 2×10^6 are immediately destroyed

[2] According to standard nomenclature in human genetics all-caps names like MHC-I are reserved to genes, for the corresponding proteins Mhc-I should be used. However, in the literature the all caps spelling is found for many proteins, including MHC. I follow the common use here and use italics (*MHC-I*) to indicate genes. Maybe one day we will get a universally accepted, perhaps even species-independent, nomenclature.

[3] The SVEDBERG (S) is the unit of the sedimentation velocity of a particle in an ultracentrifuge. It depends not only on the size, but also on the shape of the particle. Thus S-values are not additive, as in the proteasome, where 20 S + 19 S + 19 S = 26 S.

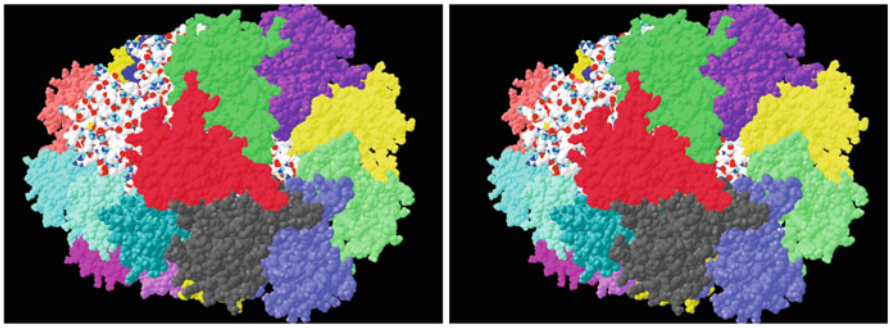


Fig. 11.18 Stereo view of the proteasome (PDB-code 1ryp) from archaebacteria. This complex consists of 4 rings with 7 proteins each. It has ATP-dependent protease activity. Presumably unfolded proteins are threaded through the hollow core, where they are degraded. There are 2 isoforms of the proteasome in mammals, a constitutively expressed and an interferon-inducible one

again by its 30.000 proteasomes, resulting in 50.000 peptides principally capable of binding to MHC-1 (see Fig. 11.19). Of those about 1000 actually bind and are displayed on the cell surface. About 10 foreign peptides on the surface of a cell are required to mount an immune response [47]. Digesting proteins before presenting them on MHC exposes antigens hidden inside the protein core.

Proteasome Inhibitors

One of the functions of the proteasome is the destruction of **cyclins**, proteins required for the control of cell division. For this reason inhibition of the proteasome can prevent cell division in cancer cells. The peptide analogue **Bortezomib** (see Fig. 11.20) was introduced in 2003 as a second-line drug to treat multiple myeloma. Because of the reactive boron group Bortezomib is rather unspecific and has considerable side effects. α',β'-epoxyketones are derivatives of the actinomycete antibiotic epoxomicin. They have fewer side effects than Bortezomib; **Carfilzomib** has been available since 2012. **ONX 0914** is an epoxyketone that is specific for the immuno-isoform of the proteasome and hence has no anticancer activity. However, it is in clinical trial as an immunosuppressant in auto-immune diseases.

Some Pathogens can Inhibit Antigen Presentation

Some pathogens have evolved mechanisms to inhibit antigen presentation to the cell surface. The following mechanisms have been identified so far:

Inhibition of proteasomal degradation of proteins is found, for Example, in EPSTEIN-BARR-virus, the causative agent of "kissing disease". The exact mechanism is unclear.

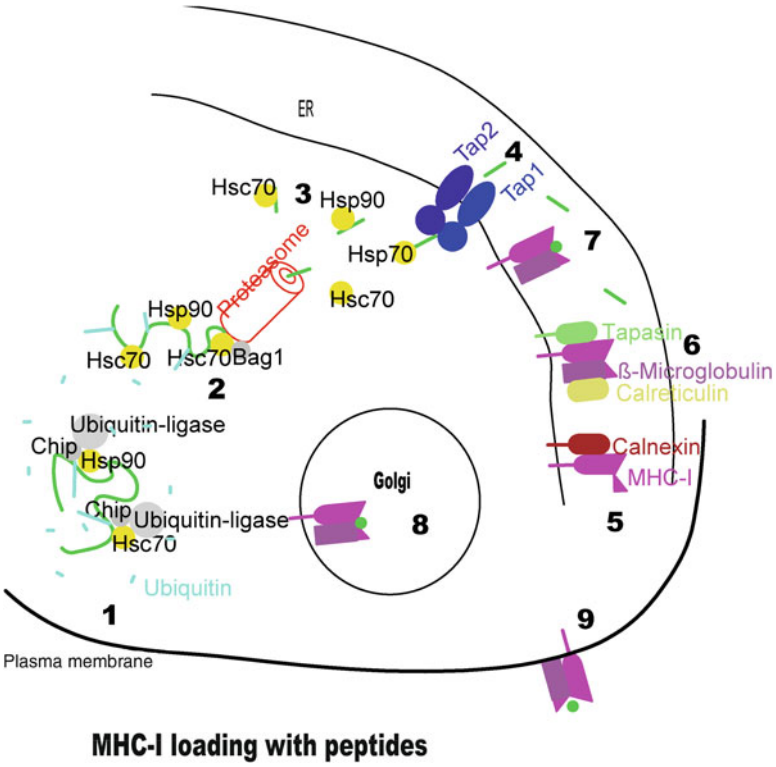


Fig. 11.19 1: Damaged proteins are recognised by molecular chaperones such as Hsp70, Hsc70 and Hsp90 (for a detailed discussion of chaperones see page 343ff), together with some cochaperones such as Chip. This leads to binding of ubiquitin-ligases, which transfer a destruction-marker peptide, ubiquitin, to the damaged protein (see also Fig. 2.36 on page 58). 2: Ubiquitinated proteins are recognised and digested by the proteasome, an ATP-dependent protease. 3: Peptides released from the proteasome are shuttled to the ER by chaperones and 4: transported into the ER by the Tap1/Tap2 system, an ABC-type transport ATPase (see Sect. 18.2.1.4 on page 444 for a fuller discussion of Tap). 5: Newly synthesised MHC-1 molecules are associated with the chaperone **calnexin** , until β_2-microglobulin has bound to them. 6: Then they interact with **calreticulin, tapasin** and Tap, until the peptide binding site has been filled. 7: This stabilises MHC-1 enough that from now on it can make do without chaperones. It is transported via the GOLGI-apparatus (8) to the cell surface (9), where the peptide can be read by $CD8^+$ T_k-cells, which destroy cancer cells and cells expressing foreign antigens (after infection with virus)

Prevention of peptide transport into the ER can occur by several mechanisms

Blocking of peptide binding to Tap1/Tap2 by herpes simplex virus 1 (HSV-1) ICP47, a protein which associates to the ER membrane on the cytosolic side and then binds to Tap1/Tap2 with a dissociation constant of 50 nM. ATP-binding to Tap is still possible afterwards, but not its hydrolysis.

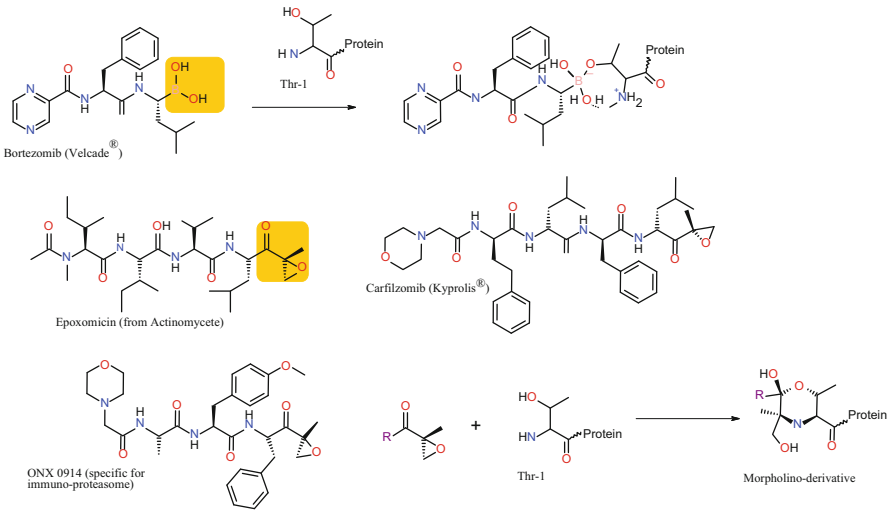


Fig. 11.20 Inhibitors of the proteasome. For details see text

Blocking of ATP binding by human cytomegalovirus US6, a type 1 membrane glycoprotein. It attacks Tap from the ER lumen, but prevents ATP binding and hydrolysis on the cytosolic side. Peptide binding is still possible, but transport is prevented.

Inhibition of transport without inhibition of peptide or ATP binding is caused by the luminal and transmembrane part of bovine herpes virus type 1 glycoprotein N.

Destruction of Tap1/Tap2 by the cytosolic domain of bovine herpes virus type 1 glycoprotein N occurs by stimulation of proteasomal digestion. Glycoprotein N will be digested in this process as well.

Destruction of the loaded MHC-I/peptide complex by stimulation of the quality control machine of the ER (see Sect. 16.5.3 on page 385). The complex is retrotransported into the cytosol, deglycosylated, ubiquitinated, and degraded by the proteasome. Examples include human cytomegalovirus US1 and US2 proteins.

Prevention of the transport of the MHC-I/peptide complex to the plasma membrane. Human cytomegalovirus US3 and US10 proteins lead to retention in the ER, murine cytomegalovirus m152 to retention in *cis*-GOLGI, and HSV-7 (cause of infectious mononucleosis) U21 protein rerouts them to the lysosome for degradation.

Degradation of MHC-I/peptide complex after arrival at the plasma membrane is, for example, caused by the HSV-8 (causes KARPOSI-sarcoma in AIDS-patients) K3 and K5 proteins. These type 1 membrane proteins reside in the ER and ubiquitinate the MHC-I protein. Once the latter arrives at the plasma

membrane, it becomes internalised and destroyed along the endosomal/lysosomal pathway. The Nef-protein (negative factor) of HIV is membrane associated via a myristoyl-tail and causes MHC-I to be incorporated into endocytic vesicles by interacting with adaptor protein (see Sect. 17 on page 393).

As seen in the section on complement (Sect. 11.3 on page 249), our immune system is capable of detecting such a situation and destroying cells that do not express MHC-I on their surface. Some virus have reacted to this by encoding proteins that look like MHC-I to our body's immune system (*e.g.* human cytomegalovirus UL18 or murine cytomegalovirus m144). This is a beautiful example of coevolution of parasite and host, even if sufferers from viral diseases might disagree.

11.4.1.2 Phagocytic Cells Present Foreign Material: MHC-II

Professional antigen presenting cells such as macrophages present on their surface peptides from the material they ingested to T-helper cells (see Fig. 11.21). These in turn can stimulate the production of reactive oxygen species and hydrolytic enzymes in the phagosomes of macrophages.

B-cells can bind foreign material via their surface-expressed antibodies; this material is phagocytosed, degraded, and presented on the surface. $CD4^+$-T_{FH}-helper cells then stimulate the production of soluble antibodies.

Dendritic cells, which stimulate T_k-cells, have a special adaptation that allows them to present materials endocytosed or pinocytosed on MHC-I instead of MHC-II. By an unclear mechanism such proteins are first transported into the ER, then retrotransported into the cytosol by the ER quality control machine, where they are ubiquitinated and then degraded by the proteasome. The resulting peptides are then treated like those resulting from the breakdown of indigenous material.

11.4.1.3 Structure of MHC-I and -II

MHC-I (see Fig. 11.22) is a heterodimer, consisting of a membrane spanning α-chain (43 kDa), and a smaller protein, β_2-microglobulin (12 kDa). The α-chain has 3 domains, called α_1, α_2 (which form the peptide binding site) and α_3, the membrane-spanning domain which is structurally similar to the β_2-microglobulin. Both α_3 and the β_2-microglobulin have an immunoglobulin fold.

MHC-II is a heterodimer consisting of two similar membrane-spanning glycoproteins, α and β. Each of these has two domains: α_2 and β_2 are membrane spanning and have immunoglobulin fold, α_1 and β_1 together form the peptide binding site.

Both MHC-I and MHC-II have a peptide binding cleft with a bottom formed by a β-sheet and side walls made from two α-helices. Peptides bind into the valley between the helices.

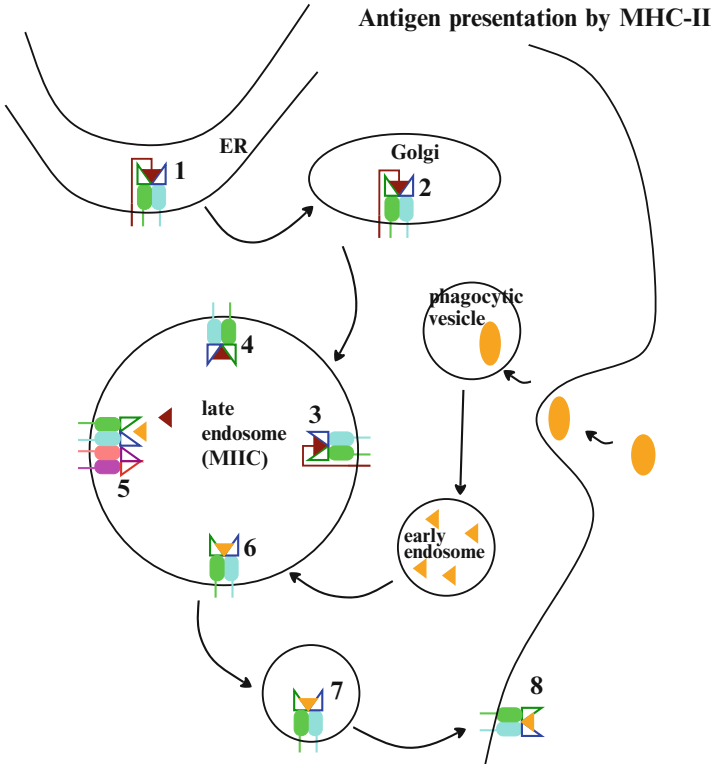


Fig. 11.21 1: Newly synthesised MHC-II is assembled in the ER; the peptide binding site is blocked by the MHC-II invariant chain (**Ii**) (*brown*), so that no self-peptides can bind. Ii also serves as a target signal for transport to the endosomes. 2,3: MHC-II is transported to the MHC-II compartment (**MIIC**) of the endosomes, where the Ii chain is degraded (5), leaving only the **CLIP**-fragment (Class II associated invariant chain peptide). 6: CLIP is exchanged for peptides from phagocytosed material. For this process a chaperone, **HLA-DM** (*pink*) is required. Once stable peptide binding has occurred, the MHC-II molecules are transferred to the cell membranes (7,8), where they are read by $CD4^+$-T_H-cells

The peptide binding sites of MHC-I and -II are similar, except that in MHC-I the peptide is tightly bound at both ends, whereas in MHC-II binding occurs more centrally; the ends of the peptide are free. Thus peptides bound to MHC-I tend to have the same length; those bound to MHC-II are more variable, and usually longer than those binding to MHC-I.

Peptides are bound to MHC at only a few amino acids, the **anchor residues**. Once binding of peptide to MHC has occurred, the complex tends to be stable for several days.

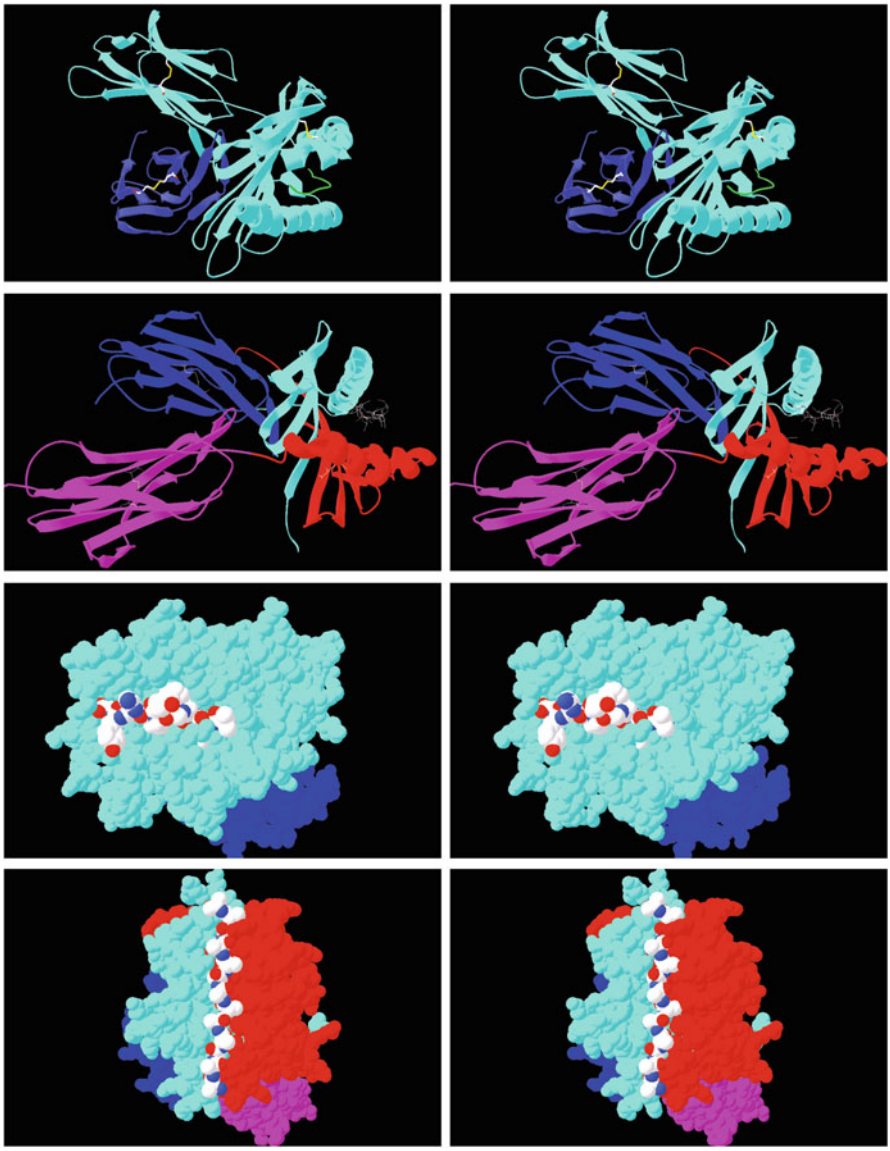


Fig. 11.22 *Top:* Ribbon-representation of the MHC-I molecule (PDB-code 1QSE). α cyan, β_2-microglobulin blue. *Second:* MHC-II (PDB-code 1SEB). α blue, β red. In both molecules disulphide bridges are marked in yellow; the antigenic peptide is also drawn. Note the immunoglobulin domains in both molecules. The peptide is bound in both molecules in a valley with a bottom formed by a β-sheet and two α-helices as side walls. In MHC-I peptide is bound only by the α-subunit, in MHC-II by both α and β. *Bottom*: Space filling models of the peptide-binding clefts of MHC-I and MHC-II, as seen by the T-cell receptor. The peptide is in an extended conformation. In MHC-I the peptide is short and both ends are surrounded by the binding side. In MHC-II the peptide is longer, and the ends stick out of the binding site. All diagrams are stereo views

11.4.1.4 MHC Variability

Both MHC-I and -II genes occur in several copies in the MHC-gene cluster (they are **polygenic**), which in addition are highly **polymorphic**, with several hundred alleles in some cases. Thus each person has a unique set of MHC-proteins, being heterozygous at most loci. A full match is usually only seen in identical twins. This variability is increased by recombination and gene conversion between misaligned chromosomes during meiosis, which constantly creates new variants of MHC.

The reason for the high variability is obvious: if all people had the same MHC-alleles, a pathogen could avoid detection by not expressing proteins that have the anchor residues to bind to MHC. This is indeed observed with small isolated populations, for example, on remote islands. Such populations tend to be very homogeneous with respect to MHC-alleles, and virus (e.g., Epstein-Barr virus) infecting such populations have adapted to this situation by avoiding epitopes that bind to such a MHC-molecule.

Several cases have been described in recent years where persons with a particular MHC-type are more resistant against a certain disease than other people. If a certain disease, say malaria, is endemic in an area, an MHC-type offering resistance to that disease will be selected for, explaining the uneven distribution of MHC-alleles in different populations (see, e.g., [48]).

Allelic variation tends to occur particularly often in those amino acids of the MHC-molecules that interact with the bound peptides.

The MHC gene complex contains more than 200 genes in humans, occupying about 4 centimorgans on chromosome 6 (except the β_2-microglobulin gene, which is on chromosome 15). The complex is divided into 3 areas:

I HLA-A, -B and -C (MHC-I genes) and a variety of so-called class-Ib genes
II HLA-DR, -DP and -DQ (MHC-II genes), an additional β-chain in the HLA-DR gene that can pair with any MHC-allele, Tap, LMP (interferon inducible proteasome subunit), DMα and -β, DOα and -β, tapasin
III C4, C2, and B (complement proteins), TNF-α and -β (cytokines), also 21-hydroxylase for steroid biosynthesis

Many of the MHC-II genes (except the negative regulator of peptide binding, HLA-DO) are induced by interferon-γ via the **MHC Class-II transactivator (CIITA)**. Patients lacking this transactivator suffer from an inherited form of severe combined immunodeficiency: **bare lymphocyte syndrome** (OMIM #209920).

After stimulation by interferon-γ, cells produce reactive oxygen species to kill invaders. This also leads to oxydative damage to the cell's proteins, especially newly translated, not yet folded ones. Such misfolded proteins are called **defect ribosomal protein products (DRiPs)**. These are degraded by the ubiquitin/proteasome system.

In order for this system to handle the increased flux of substrate, interferon-γ induces the expression of special proteasome subunits such as LMP2, LMP7, and MECL1, which can turn over substrate proteins 3–4 times faster than the conventional isoforms. In addition, the assembly of complete proteasomes from the subunits is accelerated threefold by expression of the **proteasome maturation protein (POMP)**. Thus, within 8–12 h, a sizeable fraction of standard proteasomes are replaced by i-proteasomes, and the production of peptides capable of binding to MHC-I is increased. Exchange of the normal against the interferon-induced isoforms requires disassembly of the proteasomes, leading to a transient drop in proteasome activity and the accumulation of **aggresome-like induced structures (ALIS)**, that can be detected in immunohistochemistry with anti-ubiquitin antibodies. Within 48 h, these are degraded by the increasing activity of proteasomes, even though the rate of production of DRiPs is still high. In mice that cannot express i-proteasomes, the accumulation of ALIS in neuronal tissue during inflammation leads to a multiple sclerosis-like disease. In addition, the animals have higher rates of apoptosis.

MHC-1b genes encode many proteins with immune-related function; their number is variable between species and even individuals of the same species. One example is **HLA-G**, which apparently prevents fetal placenta cells from being killed by the mother's immune system. **HLA-E** is encoded together with MHC-I, an imbalance between HLA-E and MHC-I protein on the surface informs T-killer cells that this cell has been infected with a virus that suppresses peptide presentation by MHC-1. **CD1** specialises in the presentation of bacterial cell wall components including mycolic acid.

Transplantation

As the name major histocompatibility complex indicates, the high variability creates a problem in organ transplantation, as foreign MHC-molecules are powerful antigens. Not only would a recipient of a nonmatched organ destroy it (**rejection**), but lymphocytes invariably transplanted with the organ would also attack the recipient (**graft-versus-host response**). About 1–10 % of our T-cells actually recognise foreign (**allogeneic**) MHC-molecules; these are called **alloreactive**.

One might ask now why there are so many T-cells against transplants. After all, we are not normally exposed to foreign MHC-molecules. It appears that T-cells recognise MHC in general. However, T-cells interacting with our own MHC-molecules die during their early development by apoptosis.

Each potential organ recipient is **typed** by MHC-variants; these data are stored in a central computer in a country or cooperating group of countries (such as **Eurotransplant** in Amsterdam for Europe and the **Organ Procurement and Transplantation Network** in Richmond, Virginia for the

(continued)

USA). If an organ becomes available, the MHC-type is compared with those stored in the computer. The organ is given to the recipient with the closest match. Medical urgency is also taken into account. This procedure ensures that valuable organs go to those recipients who benefit most.

11.4.2 The T-Cell Receptor

Peptides bound to MHC-molecules on the surface of our cells are recognised by a specific protein expressed on the cell membrane of T-cells. This protein is called the **T-cell receptor**. Each T-cell has about 30.000 receptor molecules on its surface. Note that contrary to normal receptor nomenclature, the T-cell receptor is located on, rather than recognises, T-cells.

11.4.2.1 Structure of T-Cell Receptor

The T-cell receptor resembles the F_{ab}-fragment of an antibody. It is a heterodimer consisting of an α- and a β-chain, each of which consists of two **immunoglobulin domains** (one variable and one constant) and an additional **tail domain** at the C-terminus. The tails are held together by a disulphide bond just outside the cell membrane. The tail can be divided into a **hinge-domain** (extracellular), a **transmembrane domain**, and a **cytoplasmic domain**.

Because the function of the T-cell receptor is determined by cell–cell interactions, it does not need a F_c-domain. There is an alternative receptor isoform with γ- and δ-chains instead of α- and β-, but its function is unknown.

Thus T-cell receptors are **monovalent**, unlike the immunoglobulins which have 2, 4, or 10 binding sites. Also, T-cell receptors are *never secreted*, they only exist membrane bound.

The α-chain is homologous to the immunoglobulin light chain and has V- and J-regions in its variable domain; the β-chain is homologous to the immunoglobulin heavy chain, with V-, D-, and J-segments.

However, there are some differences between T-cell receptor and immunoglobulin structure. The C_α-domain has only one β-sheet; where most immunoglobulins have the second β-sheet there are loosely packed strands and even an α-helix in C_α. α- and β-chains are held together in part by hydrogen bonds with carbohydrate site chains.

11.4.2.2 Origin of T-Cell Receptor Variability

The gene for the α-chain has 70–80 V-regions, 61 J-regions and one constant region, giving 4270 – 4880 possible combinations.

The gene for the β-chain has 52 V-, 2 D-, 13 J-, and 2 C-regions, resulting in 2704 possible combinations.

Thus, recombination can account for about 13×10^6 T-cell receptors. The mechanism of recombination is very similar to that seen in immunoglobulins.

T-cell receptors do not change by somatic hypermutation; this is an adaptive specialisation of B-, not T-cells. The reason might be that hypermutation would create receptors that either cannot recognise MHC-I or that become self-reactive.

11.4.2.3 Specificity of T-Cell Receptors

If a mouse with MHC-I of a particular type (let's call it MHC-I^a) is infected with a virus, proteins from that virus will be presented on the surface of infected cells and lead to the stimulation of killer cells, which destroy the infected cells.

If such T_k-cells are isolated and brought into contact with mouse cells of another MHC-I type (say, MHC-I^b), which are infected with the same virus, no killing will occur. In other words: the T_k-cells are able to recognise the viral peptide only if presented by a particular MHC-I. This is called **MHC-restriction** .

If those same T_k-cells are brought into contact with mouse cells that have MHC-I^a genotype, but are not infected with virus, no killing will occur. Also, if the cells were infected with a virus unrelated to the first one, no killing will occur.

Thus T-cells are specific not only for an antigenic peptide, but for the combination of peptide and presenting MHC-I.

11.4.2.4 T-Cell Receptor Interactions with MHC

The T-cell receptor, as we have seen, does not only interact with an antigenic peptide, but also with the presenting MHC-molecules. Indeed, the T-cell receptor binds diagonally across the peptide binding site of MHC-1, with CDR-3 making contact with the peptide (see Fig. 11.23). Not surprisingly, it is this region that contains most of the T-cell receptor variability. CDR-1 and -2 are much less variable, and interact with the MHC-surface.

V_α binds to the amino-terminus of the peptide, and V_β to the carboxy-terminus.

The MHC-II/TCR complex has a quite similar structure to that of the MHC-I/TCR complex. Contact between T_h-lymphocytes and antigen-presenting cells occurs in the peripheral lymphoid organs.

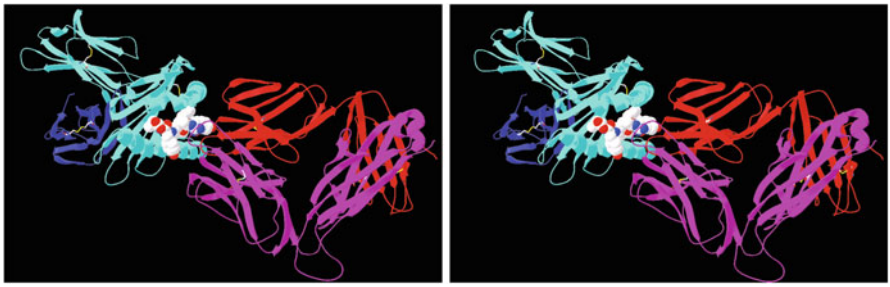


Fig. 11.23 Binding of the T-cell receptor to the MHC-1/peptide complex, stereo view (PDB-code 1QSE). The T-cell receptor straddles the MHC-I/peptide complex diagonally, with the CDR-3s of the α- and β-subunit interacting with N- and C-terminal end of the peptide, respectively

11.4.2.5 Superantigens

Some bacteria and virus produce proteins that bind not to the peptide grove of MHC-II, but to the outer surface of MHC-II and the TCR-β-subunit. Such a superantigen can stimulate T_h-cells with one or a few T-cell receptor types. Rather than stimulating a specific immune response this type of binding results in suppression of adaptive immunity and systemic toxicity, caused by a massive release of cytokines by CD4-T_h-cells.

Food poisoning by enterotoxins and toxic shock syndrome, both caused by *Staphylococcus ssp.*, are clinically important examples.

11.4.2.6 CD4 and CD8 Coreceptors

CD4 and CD8 proteins have been used as immunological markers for T_h- and T_k-cells, respectively, for some time. They were discovered because they reacted with certain monoclonal antibodies. It was also discovered that T-cells express either the one or the other, but not both, and that expression of either surface marker correlated with a particular function of the cell (stimulation of antibody production and cell killing, respectively).

Only recently it has become clear that they act as coreceptor by binding to both the T-cell receptor and MHC. Thus they stabilise MHC-TCR interactions and reduce the concentration of antigen required for T-cell stimulation by about a factor 100.

CD4 has gained some notoriety by being the binding site for the **Human immunodeficiency virus (HIV)**. It is a rod-like molecule made from 4 immunoglobulin domains and a peptide stalk in extended conformation, which goes though the cell

membrane of the T_h-cell. The intracellular part can interact with a tyrosine-kinase, **Lck**. How this contributes to T-cell activation is unclear at present.

CD8 looks quite different, being a heterodimer of an α- and a β-chain. These two proteins are quite similar: each consists of an extracellular immunoglobulin domain and a stalk, which passes the cell membrane. The stalks are cross-linked by a disulphide bridge just outside the membrane, and are heavily glycosylated. Just as CD4, however, CD8 interacts with the tyrosine kinase Lck.

In addition to the CD4 and CD8 coreceptors, cell adhesion molecules are involved in the contact of T-cells with the antigen-presenting cells. This unspecific contact precedes MHC-TCR interactions, which result in the paracrin release of mediators by the T-cell.

11.4.2.7 Function of T_h- and T_k-Cells

T_k-cells induce **apoptosis** (controlled cell death) in cells presenting foreign peptides on their MHC-I. These may come from a virus infection, or they may be embryonal proteins expressed in cancer cells. Lytic granules inside the T_k-cell contain cytotoxins (which kill the offending cell by inducing apoptosis or by forming pores in their membrane) and cytokines (which signal neighbouring tissue to activate defense mechanisms). The main cytokine is interferon-γ, which inhibits viral growth.

T_h1 cells activate macrophages. Macrophages can become infected by intracellular bacteria (e.g., *Mycobacterium tuberculosis* and *M. leprae*), which actually live inside the phagosomes. This is possible because mycobacteria have a particularly resistant cell wall, and because they prevent the release of lytic enzymes into the phagosome. Activation of the macrophage by T_h1-cells overcomes at least the latter problem. Major cytokines of T_h1-cells are IFN-γ and TNF-β. T_H2 produce cytokines that increase mucous production during parasitic infection. This strengthens the barrier function of our mucosal epithelia and is also part of the allergic response. T_H17 produce cytokines that direct neutrophiles to a site of infection. T_{FH} stimulate antibody production in B-cells.

TCR-MHC interactions lead to the clustering of TCR-molecules on the side of the T-cell closest to the antigen-presenting cell, and as a result to a reorientation of its cell skeleton. Thus the T-cell becomes **polarised** and can direct its mediators toward the antigen presenting cell.

11.5 Proteins Involved in Innate Immunity

11.5.1 PAMP-Receptors

Virus, bacteria, fungi and protozoans contain molecules that do not occur in animals (see Fig. 11.24). The innate immune system recognises those. Many of the proteins

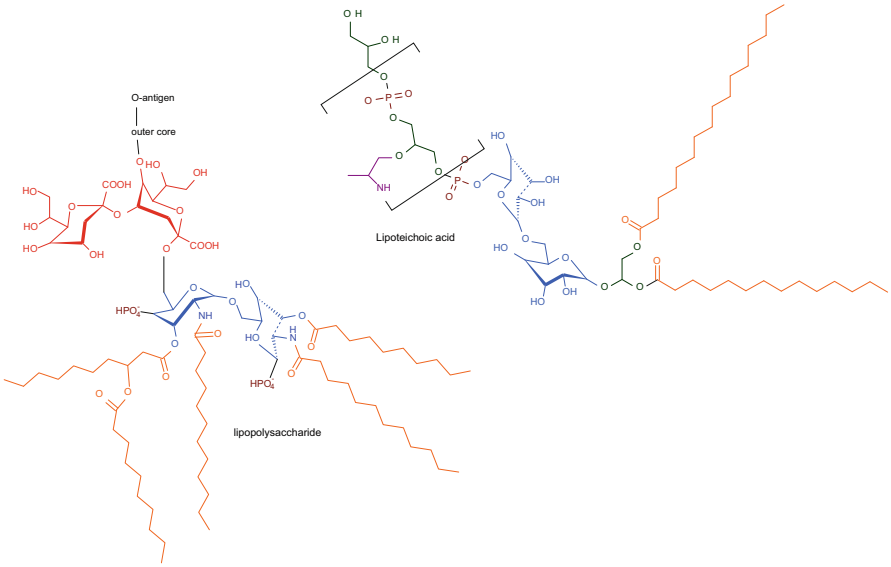


Fig. 11.24 Pathogen-associated molecular pattern (PAMP) of bacteria. *Left*: Lipopolysaccharide (LPS) are found in GRAM-negative bacteria. They consist of glucosamine (*blue*) and fatty acids (*orange*) that together form **lipid A**. The fatty acids are often β-hydroxylated; this group may carry further acyl-groups. The glucosamine may be phosphorylated (*brown*). Linked to the lipid A is the inner core saccharide (*red*), which often consists of Ketodeoxyoctulosonate (KDO). Linked to the inner core are polysaccharides that form the outer core and the O-antigen (not shown). The length of the O-antigen determines the appearance of bacterial colonies (smooth (long) versus rough (short)) and the susceptibility to hydrophobic antibiotics (higher if the O-antigen is short). *Right*: Lipoteichoic acids are found in GRAM-positive bacteria. In addition to fatty acids (*orange*), phosphate (*brown*), and sugars (*blue*) they also contain glycerol (*green*) and hydrophilic groups (*purple*)

in the innate immune system are evolutionarily very old; some even occur in both animals and plants. Some animals lacking specific immunity have instead increased the number of different PAMP-receptors, which then evolve rapidly.

Toll-like receptors [29, 44] (TLR1 through TLR13 in mammals, see Fig. 11.25) are found on the plasma membrane and on phagolysosomes of macrophages, dendritic cells, B-cells, granulocytes, and NK-cells. They act as PAMP-receptors, for example, for bacterial lipoproteins (TLR1, TLR6), peptidoglycans (TLR2), heat shock proteins (TLR2, TLR4), lipoteichoic acid (TLR2), lipopolysaccharide (TLR4), flagellin (TLR5), DNA with unmethylated CpG-islands (TLR9), dsRNA (TLR3), or bacterial rRNA (TLR13).

TLRs have an extracellular domain with a characteristic horseshoe shape made from Leu-rich repeats (LRR) that form β-strands. Substrate binding leads to dimerisation; this is transmitted through the transmembrane helix to the cytoplasmic domain that belongs to the same protein family as the interleukin-1 receptor (Toll-IL-1 receptor, TIL-1). Dimerisation of TLRs leads to activation of the NF-κ B,

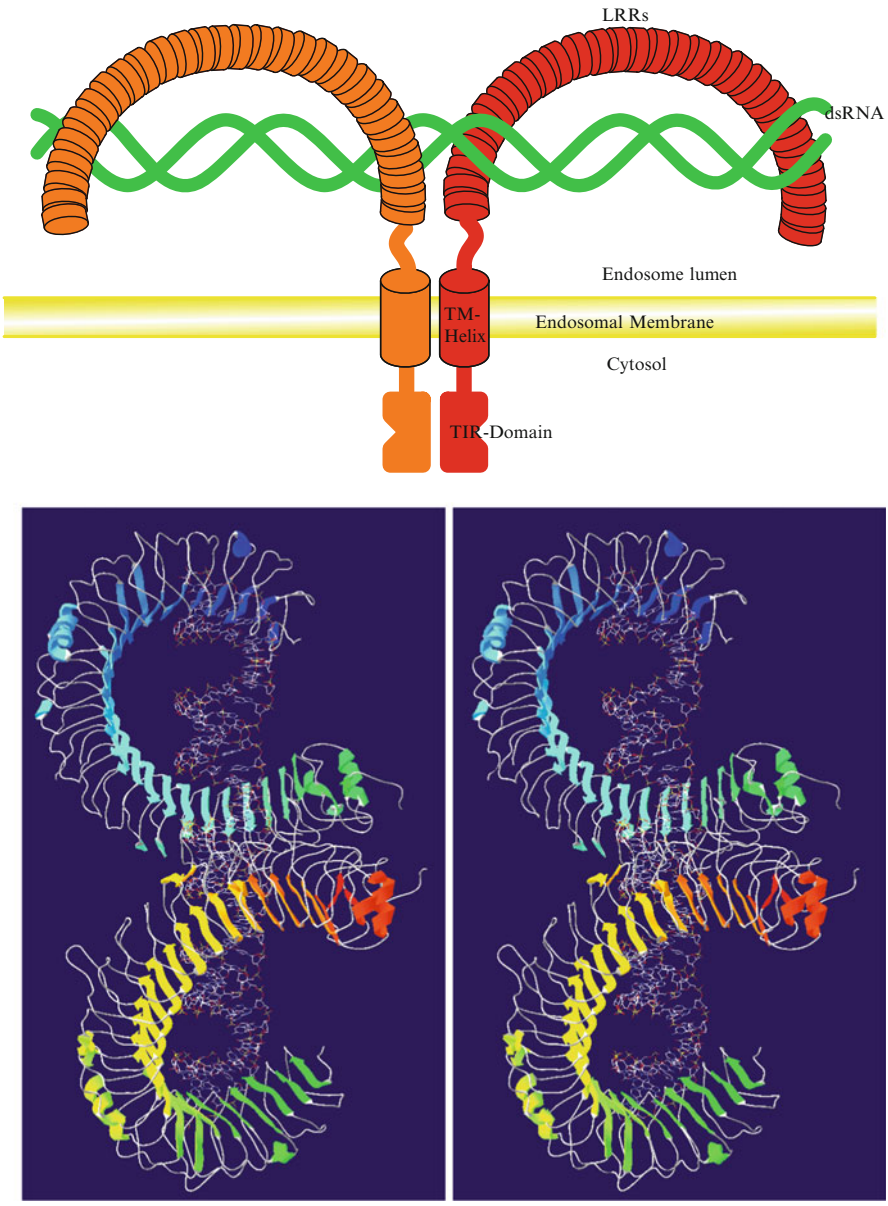


Fig. 11.25 *Top*: Schematic diagram of a Toll-like receptor, here mouse TLR3. TLRs are PAMP-receptors with a characteristic horseshoe extracellular domain made of Leu-rich repeats (LRRs), a single transmembrane helix, and a cytoplasmic interleukin-1 receptor domain (TIL-1). Ligand binding leads to dimerisation of the receptor and autophosphorylation of the TIL-1 domain. This starts the signaling cascade. Different TLRs use partly different signalling cascades, so that the innate immune system can start an appropriate response to the pathogen at hand. TLR3 is an endosomal receptor for dsRNA, a marker for viral infections. *Bottom*: X-ray crystallographic structure of the luminal domains of two molecules of mouse TLR3 (PDB-code 3ciy) bound to dsRNA. The transmembrane helix and the cytoplasmic TIL-1 domain are not part of the crystal

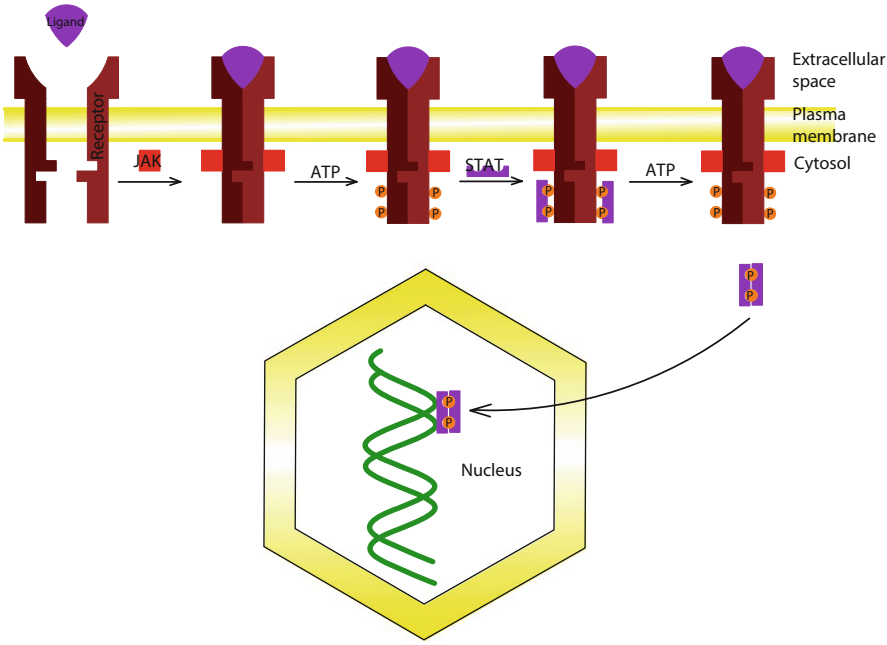


Fig. 11.26 JAK-STAT pathway. Ligand binding results in receptor dimerisation (or stabilisation of a pre-existing dimer) and binding of SH2 domains of Janus kinase (JAK) to a Pro-rich region in the cytosolic domain of the receptor. JAK is a Tyr-kinase that phosphorylates the receptors. These phosphate groups serve as binding sites for Signal transducer and activator of transcription (STAT), which is also phosphorylated by JAK. The phosphate groups on STAT allow its dimerisation (homo- or heterodimers); the dimer is transported into the nucleus where it binds to DNA, controlling gene expression. Nuclear phosphatases eventually switch off STAT. There are 4 types of JAK (JAK1–3 and TYK2) and 7 types of STAT (STAT1, STAT2, STAT3, STAT4, STAT5a, STAT5b, STAT6) in mammals

interferon regulatory factors (IRF, intermediates of the JAK/STAT pathway) (see Fig. 11.26) and MAP-pathways (see Fig. 8.1 on page 186) of gene activation, leading to the production of cytokines.

TLRs have the same structure as the *Drosophila* receptor toll (the German word for "amazing"), which is required both for establishing the dorsoventral axis during embryogenesis and for defence against fungal infections. Similar proteins have also been found in plants.

The 2011 NOBEL Prize in Medicine or Physiology was awarded to BRUCE A. BEUTLER and JULES A. HOFFMANN for their work on TLRs.

In the cytoplasm of epithelia, macrophages and dendritic cells occur **NOD-like receptors** – proteins that have a C-terminal LRR, a central Nucleotide-binding oligomerisation domain (NOD) and a N-terminal Caspase recruitment domain (CARD). CARD is a member of the **DEAD-domain** protein family, characterised by the conserved motif Asp-Glu-Ala-Asp. NLRs bind breakdown products of peptidoglycans from intracellular bacteria and activate cytokine expression via NF-κ B.

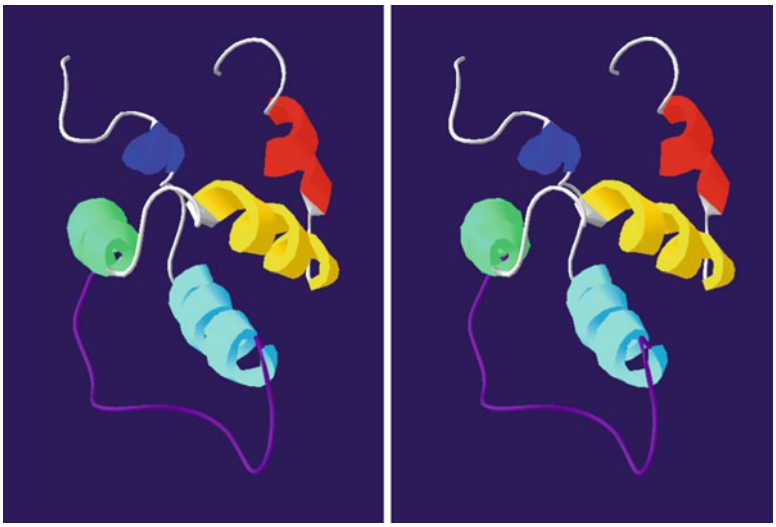


Fig. 11.27 Pyrin domain of NALP-1 (PDB-code 1pn5). The pyrin domain has a flexible, unordered loop (*purple*) where other death domains (DD, DED and CARD) have an α-helix

CROHN **Disease and** BLAU **Syndrome**
NOD2 is expressed specifically in the PANETH cells of the gut. If this protein is nonfunctional, production of antimicrobial peptides such as defensin is impaired resulting in chronic inflammation of the bowel (CROHN disease, OMIM #266600). On the other hand, gain of function mutations lead to familial granulomatous inflammatory arthritis, dermatitis, and uveitis (BLAU syndrome, OMIM #186580).

NALP are NLR with pyrin-domains (see Fig. 11.27) instead of CARD in their N-terminus. They detect stress and cellular damage (low cytosolic $[K^+]$), but also infection with virus or cytosolic bacteria and particles including asbestos or $Al(OH)_3$ (an important adjuvant!). Humans have 14 such proteins. NALP, PYCARD, caspase 1, and sometimes caspase 5 together form the inflammasome in myeloid cells. The inflammasome proteolytically activates cytokines such as interleukin 1β and IL 18. Activation also induces pyroptosis, a process similar to apoptosis.

NALPs in Disease
Gout is a very painful inflammation of the joints caused by precipitation of uric acid crystals. Uric acid is only sparingly soluble in water; if the solubility

(continued)

limit of plasma is exceeded, uric acid precipitates in the coldest part of the body, in toes and fingers. Because of proteoglycans the joints have an acidic environment that converts urate ions into the even less soluble acid. Excess uric acid formation is a consequence of a diet rich in meat and worsened by alcohol consumption. Uric acid crystals activate NALP-3 and hence the inflammasome. Treatment is by anti-inflammatory drugs such as colchicine, prevention of uric acid formation with Allopurinol, or by preventing urate reabsorption in the kidney with Probenecid.

Inappropriate activation of NALPs by gain-of-function mutations leads to spontaneous production of inflammatory cytokines and hence autoinflammatory diseases including **familial cold inflammatory syndrome** (OMIM #120100) and MUCKLE-WELLS **syndrome** (OMIM #191900).

Cellular mRNA is capped by addition of 7-methylguanosine to the 5'-triphosphate residue during synthesis in the nucleus. Viral RNA, however, is synthesised in the cytosol and hence not capped. Sensors with **RIG-like helicase** domain (RLH) detect such RNAs; their CARD-domains then stimulate signalling via adaptor-proteins. Picornavirus evade detection by RLHs by modifying the 5'-end of their RNA with a protein. Such modified RNAs are detected by **melanoma-differentiation associated protein 5 (MDA-5, helicard)** which is specific for dsRNA.

Diabetes Mellitus Type I

Loss-of-function mutations in RIG-1 or MDA-5 do not lead to clinically detectable immunodeficiencies, however, such mutations in MDA-5 seem to protect against type I diabetes, as do some mutations in various HLA-genes (others sensitise). This type of diabetes is caused by autoimmune destruction of the β-cells of the pancreas presumably following virus infection. Coxsackie or rubella virus is thought to be the culprit. It usually occurs in youth; patients require lifelong maintenance on injected insulin. The exact pathomechanism is unclear. Breast-feeding [4, 51] and calciol (vitamin D) supplementation in infancy [57] are reported to have a protective effect, nicotinamide (vitamin B_3) may reduce progression to diabetes after autoantibodies are detected [16].

11.5.2 Cytokines

Phagocytic cells activated by pathogen binding to PAMP receptors produce a number of protein molecules that act as autocrine (on the cell itself), paracrine (on nearby cells), or endocrine (on distant cells) hormones. Different PAMP-receptors

can produce different cytokines and thus steer the immune response into appropriate directions. The cytokines have a molecular mass of about 25 kDa and come in different families:

Interleukin 1 includes IL-1α, IL-1β, IL-18, and 8 less well-characterised proteins. They are produced as propeptides; the N-terminal end is cleaved off by caspases in inflammasomes upon stimulation of TLRs. IL-1 receptors have a cytosolic TIL-1 domain and act via NF-κ B as discussed for TLR. These interleukins increase the permeability of the vascular endothelium and thus increase lymphocyte access to the site of infection.

Hæmatopoietins include erythropoietin (growth factor for erythrocytes, [23]), growth hormone, IL-6, and GM-CSF. They have dimeric receptors [12] that belong to several subfamilies and signal through the JAK/STAT-pathway. One chain of the receptor determines the cytokine bound, and the other the STAT-protein through which the receptor acts.

In addition to the systemic role of kidney-produced erythropoietin in hæmatopoiesis, tissue-produced erythropoietin has a paracrine role, inhibiting macrophage activation and inflammatory cytokine release after immune stimulation and attracting stem cells for wound healing [7].

Tumor necrosis factor (TNF) are homo- or heterotrimers of membrane-bound proteins that can, however, be cleaved from the membrane by the protease ADAM17 and released into the bloodstream. There are two types of TNF-receptors: TNFR-I on most cells and TNFR-II expressed on lymphocytes. Both contain death-domains and signal via adaptors. TNF stimulates extravasation of white blood cells and blood clotting in small blood vessels to close off the area of infection.

Chemokines are a family of more than 50 related proteins [41]. They direct chemotaxis of granulocytes, fibroblasts, NK- and T-cells, and phagocytic cells out of the blood vessel toward the site of infection. They act through G-protein coupled receptors. CC-chemokines have two Cys-residues near the N-terminus; in CXC-chemokines these Cys are separated by another amino acid.

Type I interferons like IFN-α, interferon-β and interferon-like cytokines such as IL-4 also act through IFN-α/β receptors (IFNAR1 and IFNAR2) via the JAK/STAT-pathway. They are produced in response to viral infections. Most cells can produce interferons, but plasmacytoid dendritic cells are particularly busy. IFN-β induces many cell types to produce IFN-α, thereby amplifying the signal. IFN binding to an infected cell, or to cells nearby, induces several enzymes that reduce virus replication:

2',5'-oligoadenylate synthetase EC 2.7.7.84 produces oligomers in 2',5'-linkage (rather than 3',5') from ATP that activate RNase L, which in turn degrades dsRNA.

Protein kinase R (PKR) EC 2.7.11.1 has a N-terminal dsRNA-binding and a C-terminal Ser/Thr protein kinase domain. Binding of dsRNA to PKR activates the kinase, leading to the phosphorylation of eIF-2 and hence a stop of translation of RNA with AUG start codons. In addition, PKR

phosphorylates Iκ B, leading to activation of the NF-κ B pathway. By an unknown route, PKR can also stimulate apoptosis.

MX-proteins are nuclear GTPases with homology to dynamin. Their mode of action is unknown, but mice lacking these proteins are highly susceptible to virus infection.

Note that IFN-γ does not belong to the type I interferons. It is the only member of type II. Its production is restricted to T- and NK-cells in response to IL-12; it acts through the IFNGR receptor.

The mechanism of cytokine reception has been reviewed recently in [1]. Locally cytokines increase vascular permeability for lymphocytes, granulocytes, phagocytic cells, immunoglobulins, and complement. More fluid is drained into the lymphatic system, where pathogens can be cleared from it. They stimulate T- and B-cells and hence the development of the adaptive immune system. Systemically, they produce fever (hence the name **endogenous pyrogens** for TNF-α, IL-1β and IL-6) by stimulating the expression of cyclooxygenase and hence prostaglandin production. They also mobilise metabolites and induce the production of acute-phase proteins in the liver. Leucocytes are released from both the bone marrow and from binding sites in blood vessels. In infections of the blood (**sepsis**), systemic release of cytokines, most notably soluble TNF-α, can produce **septic shock** leading to death by multiple organ failure.

11.5.3 The Acute-Phase Response

The endogenous pyrogens TNF-α, IL-1β and IL-6 act on hepatocytes, which start producing acute-phase proteins. This happens within 24–48 h after infection.

C-reactive proteins is a pentamer that binds to the phosphocholine part of lipopolysaccharides (see Fig. 11.28). Then it binds C1q, starting the classical complement pathway.

Mannan binding lectin (MBL) In the blood of healthy individuals, MBL concentration is low. It increases during infection. As discussed above, MBL binding starts the lectin pathway of complement activation, it also opsonises mannose-containing bacteria and fungi for monocytes (which unlike macrophages do not have a mannose receptor).

Surfactant proteins -A and -D are produced not only by hepatocytes, but also epithelial cells. They stimulate macrophages.

Fibrinogen is a 340 kDa plasma hexameric glycoprotein (2 copies each of α-, β-, and γ-subunits) that is converted into fibrin during blood clotting. This conversion is achieved by the Ser-protease thrombin. Fibrin strands are cross-linked by factor VIII to form the clot.

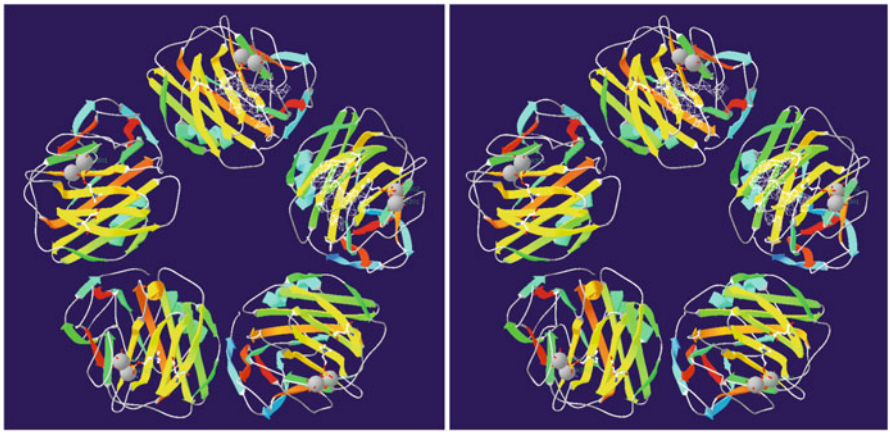


Fig. 11.28 Human C-reactive protein with bound Ca^{2+} and phosphocholine (PDB-code 1b09). The phosphocholine is held in place by bonds between its phosphate group and the two Ca^{2+}-ions and by a bond between its amino group and Glu-81. There is a hydrophobic patch (white mesh) near the phosphocholine binding site that presumably would bind to other components of lipopolysaccharide. C-reactive protein belongs to the pentraxins, a group of pentameric proteins

11.5.4 Antibacterial Proteins

Mammalian plasma and secretions such as tears, milk, or saliva contain a couple of proteins that inactivate pathogens as part of our innate defence:

Lysozyme (EC 3.2.1.17) hydrolyses the glycosidic bond between carbohydrates in the murein of the bacterial cell wall (see Fig. 11.29). It is most effective against GRAM-positive bacteria as those are not protected by an outer membrane.

Phospholipase A_2 (EC 3.1.1.4) cleaves the acyl-residue of the C2 of the glycerol in phospholipids. This results in the production of a lysolipid and the salt of a fatty acid (i.e., a soap). Both act as detergent and lyse membranes. The fatty acid on C2 is often arachidonic acid, which is converted to eicosanoids, powerful mediators of inflammation (Fig. 11.30).

Lectins are sugar-binding proteins. Mannan binding lectin (MBL) and the ficolins bind to sugar residues in the bacterial cell wall, lipopolysaccharide and lipoteichoic acids in bacterial cell membranes, or the high-mannose glycoproteins of yeasts. Lectin-binding to pathogens opsonises them for lysis by the complement system as discussed above (page 249).

Antimicrobial polypeptides are contained in lysosomes or secreted into bodily fluids. They destroy both GRAM-positive and -negative bacteria by mechanisms not fully understood. They also have immune-modulatory effects. Some are under development as antibiotics. We distinguish

Defensins (see Fig. 11.31) are small, cationic, hydrophobic proteins with several disulphide bridges. They bind to the anionic membranes of microor-

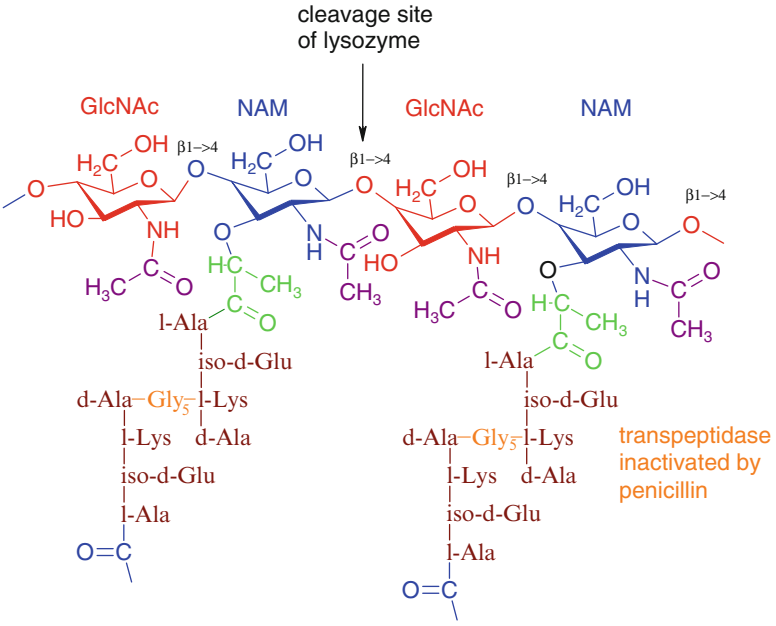


Fig. 11.29 Structure of murein in the bacterial cell wall. Murein consists of polysaccharide chains (N-acetylglucosamine (GlcNAc) and N-acetylmuraminic acid (NAM)) cross-linked by peptides. The peptides contain D-amino acids, which are otherwise very unusual in biological structures. Lysozyme cleaves the β1→4 glycosidic bond between the carbohydrates. Lactam-antibiotics such as penicillin prevent the formation of the peptide cross-links

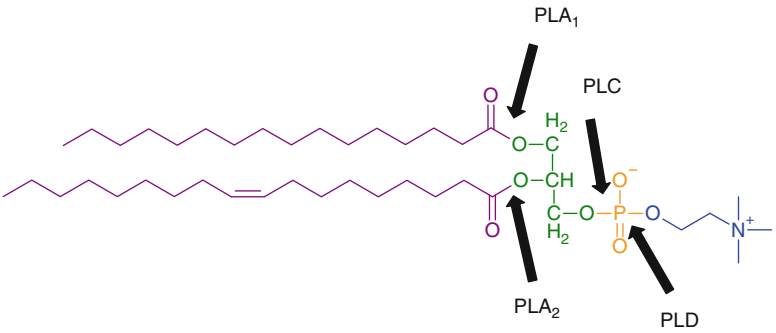


Fig. 11.30 Cleavage sites of the various phospholipases. An enzyme with both PLA_1 and PLA_2 activity is called phospholipase B

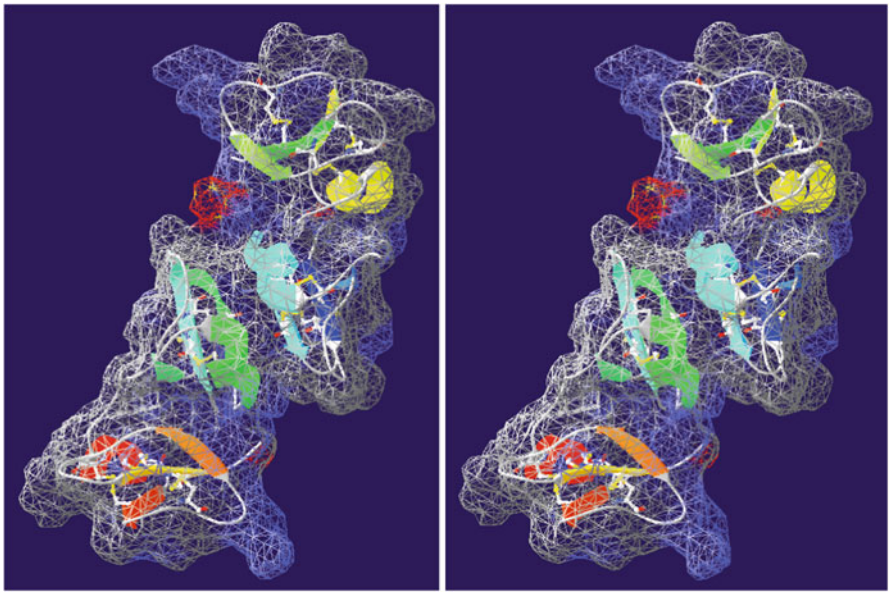


Fig. 11.31 Human β-defensin 2 (PDB-code 1fd3). Areas with positive charge on the surface are shaded blue

ganisms (the outer leaflet of the mammalian plasma membrane contains mostly zwitterionic lipids such as sphingomyelin and phosphatidylcholine), where they probably form pores. Defensins are categorised into three groups: α, β and θ, of which humans have 6, 30, and 6, respectively. However, the mRNA of the human θ-defensins contain premature stop-codons, hence they are not expressed. θ-defensins are cyclic peptides, the only ones known in mammals.

Cathelicidins are small proteins with amphipatic α-helices that occur in lysosomes of macrophages and polymorphonuclear leukocytes and some other cells. They are also excreted into body fluids. Expression is increased by calcitriol, the activated form of vitamin D. Humans have only one cathelicidin, hCAP-18, called LL-37 after proteolytic activation.

Histatins are antimicrobial peptides found in saliva.

Antimicrobial polypeptides also act as natural broad-spectrum antivirals [20], both against enveloped and nonenveloped virus. Of the currently 60 licensed antiviral drugs half are directed against HIV, the reminder against only six other virus species (hepatitis B and C, herpes simplex, varicella-zoster,

(continued)

influenza, and cytomegalovirus). The species-specificity of the drugs is a major obstacle to treatment:

- No drugs are available for many viral disease.
- Makes the drugs expensive as they work only in a relatively small number of patients.
- Requires laboratory identification of virus species that is time consuming and may not be feasible in developing countries.

Given the high morbidity and mortality from viral diseases especially in children, the polypeptides, or drugs developed from them, may come in very handy indeed!

11.6 Exercises

11.6.1 Problems

11.1. Critically discuss the physiological significance of the various immunoglobulin isoforms. What are the consequences for immunisation?

11.2. Bortezomib (Velcade®) is a proteasome inhibitor used to treat certain cancers.

The anticancer effect is most likely due to inhibition of proteasomal degradation of

A fetal antigens expressed by de-differentiated cancer cells.
B envelope proteins of cancer viruses.
C cyclins.
D damaged proteins.
E elongation factors.

11.3. In the 1990s several attempts were made to develop an AIDS vaccine. Apes were injected with antigens from SIV (simian immunodeficiency virus, an experimental model for HIV). They developed antibodies against those antigens and upon i.v. challenge with the virus showed protection. However, if the virus was smeared onto the vagina, no protection was observed.

Failure of immunization to this route of infection was caused by failure of the apes to produce virus-specific:

A IgA
B IgD
C IgE
D IgG
E IgM

11.4. A 35-year-old woman comes to her physician complaining, as every year in late spring, of sneezing, wheezing, running nose, and watery eyes. Secretions are clear and there are no signs of infection.

Which of the following immunoglobulin isoforms is most likely involved in the pathogenesis of her condition?

A IgA
B IgD
C IgE
D IgG
E IgM

11.5. You want to perform a bone marrow transplantation to cure a patient of X-linked severe combined immunodeficiency (X-SCID).

Which of the steps normally required for this procedure can be dispensed with in patients with SCID?

A Finding a histocompatible donor
B Destruction of T-cells in the graft with monoclonal antibodies
C Enrichment of stem cell by fluorescence-activated cell sorting
D Pretransplantation chemotherapy of the recipient
E Injection of colony-stimulating factors after transplantation

11.6.2 Solutions

11.1 See pages 234 and 241, and Table 11.3

11.2

A fetal antigen expressed by dedifferentiated cancer cells False. Would make it even harder for the patient's immune system to fight the cancer
B envelope proteins of cancer virus False. Few cancers in humans are related to virus.
C cyclins Correct. This would interfere with cell division, the primary problem in cancer.
D damaged proteins False. Would make cancer cells grow even more rapidly.
E elongation factors False. Destruction of elongation factors is not required for mitosis.

11.3 Mucous membranes are protected against infection by immunoglobulin A, which can be secreted.

11.4 Patient is suffering from hay fever, an allergy against pollen. This is caused by IgEs responding to an inappropriate target.

11.5 Because the patient doesn't have an immune system, it need not be destroyed to protect the graft.

References

1. M. Atanasova, A. Whitty, Understanding cytokine and growth factor receptor activation mechanisms. Crit. Rev. Biochem. Mol. Biol. **47**(6), 502–530 (2012). doi:10.3109/10409238.2012.729561
2. F. Barré-Sinoussi, J.C. Chermann, F. Rey, M.T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum, L. Montagnier, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**(4599), 868–871 (1983). doi:10.1126/science.6189183
3. E. von Behring, S. Kitasato, Über das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. Dtsch. Med. Wochenschr. **16**, 1145–1148 (1890). URL http://archiv.ub.uni-marburg.de/eb/2013/0164/view.html
4. K. Borch-Johnsen, T. Mandrup-Poulsen, B. Zachau-Christiansen, G. Joner, M. Christy, K. Kastrup, J. Nerup, Relation between breast-feeding and incidence rates of insulin-dependent diabetes mellitus. Lancet **324**(8411), 1083–1086 (1984). doi:10.1016/S0140-6736(84) 91517-4
5. J. Bordet, Sur le mode d'action des sérums préventifs. Ann. Inst. Pasteur **10**, 193–219 (1896). URL https://archive.org/stream/cbarchive_48033_surlemodedactiondesserumspreve1887/surlemodedactiondesserumspreve1887_djvu.txt
6. C. Brack, M. Hirama, R. Lenhard-Schuller, S. Tonegawa, A complete immunoglobulin gene is created by somatic recombination. Cell **15**(1), 1–14 (1978). doi:10.1016/0092-8674(78)90078-8
7. M. Brines, A. Cerami, The receptor that tames the innate immune response. Mol. Med. **18**, 486–496 (2012). doi:10.2119/molmed.2011.0041
8. Sir MacFarlane Burnet, *The Clonal Selection Theory of Acquired Immunity: The* ABRAHAM FLEXNER *lectures of Vanderbilt University, 1958* (Vanderbilt University Press, 1959) URL https://www.archive.org/stream/clonalselectiont00burn/clonalselectiont00burn_djvu.txt
9. E. Buxbaum, *Biophysical Chemistry of Proteins: An Introduction to Laboratory Methods* (Springer, New York, 2011). ISBN 978-1-4419-7250-7
10. J.D. Chalmers, B.J. McHugh, C. Doherty, M.P. Smith, J.R. Govan, D.C.Kilpatrick, A.T. Hill, Mannose-binding lectin deficiency and disease severity in non-cystic fibrosis bronchiectasis: A prospective study. Lancet Resp. Med. **1**(3), 224–232 (2013). doi:10.1016/S2213-2600(13)70001-8
11. N.E.C. Clough, P.J. Hauer, Using polyclonal and monoclonal antibodies in regulatory testing of biological products. ILAR J. **46**(3), 300–306 (2005). doi:10.1093/ilar.46.3.300
12. D. Cosman, The hematopoietin receptor superfamily. Cytokine **5**(2), 95–106 (1993). doi: http://dx.doi.org/10.1016/1043-4666(93)90047-9
13. Creighton, T.E. The physical and chemical basis of molecular biology. Helvetica Press (2010)
14. P. Ehrlich, Experimentelle Untersuchungen über Immunität. I. Ueber Ricin. Dtsch. Med. Wochenschr. **17**(32), 976–979 (1891a). URL http://www.pei.de/SharedDocs/Downloads/institut/veroeffentlichungen-von-paul-ehrlich/1886-1896/1891-experimentelle-untersuchungen-immunitaet-ricin.pdf
15. P. Ehrlich, Experimentelle Untersuchungen über Immunität. II. Ueber Abrin. Dtsch. Med. Wochenschr. **17**, 1218–1219 (1891b). URL http://www.pei.de/SharedDocs/Downloads/institut/veroeffentlichungen-von-paul-ehrlich/1886-1896/1891-experimentelle-untersuchungen-immunitaet-abrin.pdf
16. R.B. Elliott, C.C. Pilcher, D.M. Fergusson, A.W. Stewart, A population based strategy to prevent insulin-dependent diabetes using nicotinamide. J. Pediatric Endocrinol. Metabol. **9**(5), 501–509 (1996). doi:10.1515/JPEM.1996.9.5.501
17. E. Engvall, P. Perlman, Enzyme-linked immunosorbent assay of immunoglobulin G. Immunochemistry **8**, 871–874 (1971). doi:10.1016/0019-2791(71)90454-X
18. A. Fagreaus, Antibody production in relation to development of plasma cells. Acta Med. Scand. **130**(suppl. 204), 3–122 (1948)

19. D. Fera, A.G. Schmidt, B.F. Haynes, F. Gao, H.-X. Liao, T.B. Kepler, S.C. Harrison, Affinity maturation in an HIV broadly neutralizing B-cell lineage through reorientation of variable domains. Proc. Natl. Acad. Sci. USA **111**(28), 10275–10280 (2014). doi:10.1073/pnas.1409954111
20. E.G. Findlay, S.M. Currie, D.J. Davidson, Cationic host defence peptides: Potential as antiviral therapeutics. BioDrugs **27**, 479–493 (2013). doi:10.1007/s40259-013-0039-0
21. P.D. Frederiksen, S. Thiel, C.B. Larsen, J.C. Jensenius, M-ficolin, an innate immune defence molecule, binds patterns of acetyl groups and activates complement. Scand. J. Immunol. **62**(5), 462–473 (2005). doi:10.1111/j.1365-3083.2005.01685.x
22. P.A. Gorer, S. Lyman, G.D. Snell, Studies on the genetic and antigenic basis of tumour transplantation. linkage between a histocompatibility gene and 'fused' in mice. Proc. Roy. Soc. Lond. B **135**(881), 499–505 (1948). URL http://rspb.royalsocietypublishing.org/content/135/881/499.full.pdf+html
23. V.H. Haase, Regulation of erythropoiesis by hypoxia-inducible factors. Blood Rev. **27**(1), 41–43 (2013). doi:10.1016/j.blre.2012.12.003
24. J. Heine, *Beobachtungen über Lähmungszustände der unteren Extremitäten und deren Behandlung* (Köhler, Stuttgart, 1840)
25. F.G.J. Henle, *Pathologische Untersuchungen von den Miasmen und Kontagien und von den miasmatisch-kontagiösen Krankheiten* (Berlin, 1840)
26. C. Honoré, T. Hummelshoj, B.E. Hansen, H.O. Madsen, P. Eggleton, P. Garred, The innate immune component ficolin 3 (Hakata antigen) mediates the clearance of late apoptotic cells. Arthritis Rheum. **56**(5), 1598–1607 (2007). doi:10.1002/art.22564
27. N. Hozumi, S. Tonegawa, Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Natl. Acad. Sci. USA **73**(10), 3628–3632 (1976). URL http://www.pnas.org/content/73/10/3628.full.pdf
28. S. Inouye, A. Hasegawa, S. Matsuno, S. Katow, Changes in antibody avidity after virus infections: Detection by an immunosorbent assay in which a mild protein-denaturing agent is employed. J. Clin. Microbiol. **20**(3), 525–529 (1984). URL http://jcm.asm.org/content/20/3/525.full.pdf+html
29. C.A. Janeway, Jr., R. Medzhitov, Innate immune recognition. Ann. Rev. Immunol. **20**, 197–216 (2002). doi:10.1146/annurev.immunol.20.083001.084359
30. E. Jenner, *An Inquirey into the Causes and Effects of the Variolæ Vaccinæ, a Disease Discovered in some of the Western Countries of England, Particularly Glaucestershire, and known by the name of the Cow Pox* (Published by the author, London, 1798)
31. N.K. Jerne, The natural-selection theory of antibody formation. Proc. Natl. Acad. Sci. USA **41**(11), 849–857 (1955). URL http://www.pnas.org/content/41/11/849.full.pdf+html?sid=56d3763c-5641-46ab-9ab4-a0e2af2ea44a
32. R. Koch, Die Ætiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des *Bacillus anthracis*. Beiträge zur Biologie der Pflanzen **2**(2), 277–311 (1876)
33. G. Köhler, C. Milstein, Continuous cultures of fused cells secreting antibody of predefined specificity. Nature **256**(5517), 495–497 (1975). doi:10.1038/256495a0
34. A. Krarup, S. Thiel, A. Hansen, T. Fujita, J.C. Jensenius, L-ficolin is a pattern recognition molecule specific for acetyl groups. J. Biol. Chem. **279**(46), 47513–47519 (2004). doi:10.1074/jbc.M407161200
35. N.J. Lynch, S. Roscher, T. Hartung, S. Morath, M. Matsushita, D.N. Maennel, M. Kuraya, T. Fujita, W.J. Schwaeble, L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of gram-positive bacteria, and activates the lectin pathway of complement. J. Immunol. **172**(2), 1198–1202 (2004). doi:10.4049/jimmunol.172.2.1198
36. D.D. McGregor, M.B. Gowans, The antibody response of rats depleted of lymphocytes by chronic drainage from the thoracic duct. J. Exp. Med. **117**(2), 303–320 (1963). URL http://jem.rupress.org/content/117/2/303.full.pdf+html
37. P.B. Medawar, Immunological tolerance. Science **133**(3449), 303–306 (1961). URL http://nfs.unipv.it/nfs/minf/dispense/immunology/lectures/files/references/medawar_1960.pdf

38. E. Metschnikoff, Untersuchung über die mesodermalen Phagocyten einiger Wirbeltiere. Biologisches Centralblatt **3**(III), 560–565 (1883–1884). URL http://www.biodiversitylibrary.org/item/27783#page/572/mode/1up
39. J.F.A.P. Miller, Immunological function of the thymus. Lancet **278**(7205), 748–749 (1961). doi:10.1016/S0140-6736(61)90693-6
40. K. Murphy, *Janeway's Immunobiol.*, 8th edn. (Taylor & Francis, New York, NY, 2011). ISBN 978-0-8153-4243-4
41. H. Nomiyama, N. Osada, O. Yoshie, The evolution of mammalian chemokine genes. Cytokine Growth Factor Rev. **21**(4), 253–262 (2010). doi:10.1016/j.cytogfr.2010.03.004
42. M. Nonaka, A. Kimura, Genomic view of the evolution of the complement system. Immunogenetics **58**(9), 701–713 (2006). doi:10.1007/s00251-006-0142-1
43. M. Noris, G. Remuzzi, Overview of complement activation and regulation. Sem. Nephrol. **33**(6), 479–492 (2013). doi:10.1016/j.semnephrol.2013.08.001
44. B.S. Park, J.-O. Lee, Recognition of lipopolysaccharide pattern by TLR4 complexes. Nature Exp. Mol. Med. **45**, e66 (2013). doi:10.1038/emm.2013.9
45. M. Pasteur, An address on vaccination in relation to chicken cholera and splenic fever. Br. Med. J. **2**, 283–284 (1881). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2264103/pdf/brmedj05026-0035.pdf
46. H.P. Pöhn, G. Rasch, *Statistik meldepflichtiger, übertragbarer Krankheiten vom Beginn der Aufzeichnungen bis heute (Stand 31. Dezember 1989)* (MMV Medizin Verlag, München, 1994). URL http://edoc.rki.de/documents/rki_ab/reYwfdwOXfVLs/PDF/220lwYF098W2I.pdf
47. M.F. Princiotta, D. Finzi, S.-B. Qian, J. Gibbs, S. Schuchmann, F. Buttgereit, J.R. Bennink, J.W. Yewdellemail, Quantitating protein synthesis, degradation, and endogenous antigen processing. Immunity **18**, 343–354 (2003). doi:10.1016/S1074-7613(03)00051-7
48. D.J. Roberts, A.G. Craig, A.R. Berendt, R. Pinches, G. Nash, K. Marsh, C.I. Newbold, Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature **357**, 689–692 (1992). doi:10.1038/357689a0
49. R. Rokyta, J. Fricová, Ontogeny of the pain. Physiol Res. **61**(Suppl. 1), S109–S122 (2012). URL http://www.biomed.cas.cz/physiolres/pdf/61%20Suppl%201/61_S109.pdf
50. G. Schneider, P. Guttmann, S. Heim, S. Rehbein, F. Mueller, K. Nagashima, J.B. Heymann, W.G. Müller, J.G. McNally, Three-dimensional cellular ultrastructure resolved by X-ray microscopy. Nature Meth. **7**(12), 985–987 (2010). doi:10.1038/nmeth.1533
51. N. Shehadeh, R. Shamir, M. Berant, A. Etzioni, Insulin in human milk and the prevention of type 1 diabetes. Pediatr. Diabetes **2**(4), 175–177 (2001). doi:10.1034/j.1399-5448.2001.20406.x
52. B. Suligoi, C. Galli, M. Massi, F. Di Sora, M. Sciandra, P. Pezzotti, O. Recchia, F. Montella, A. Sinicco, G. Rezza, Precision and accuracy of a procedure for detecting recent human immunodeficiency virus infections by calculating the antibody avidity index by an automated immunoassay-based method. J. Clin. Microbiol. **40**(11), 4015–4020 (2002). doi:10.1128/JCM.40.11.4015-4020.2002
53. Y. Tang, J. Lou, R.K. Alpaugh, M.K. Robinson, J.D. Marks, L.M. Weiner, Regulation of antibody-dependent cellular cytotoxicity by IgG intrinsic and apparent affinity for target antigen. J. Immunol. **179**, 2815–2823 (2007). doi:10.4049/jimmunol.179.5.2815
54. M. Tsujimura, T. Miyazaki, E. Kojima, Y. Sagara, H. Shiraki, K. Okochi, Y. Maeda, Serum concentration of Hakata antigen, a member of the ficolins, is linked with inhibition of *Aerococcus viridans* growth. Clin. Chim. Acta **325**(1–2), 139–146 (2002). doi:10.1016/S0009-8981(02)00274-7
55. A. Wassermann, A. Neisser, C. Bruck, Eine serodiagnostische Reaktion bei Syphilis. Dtsche Med. Wochenschr. **32**(19), 745–746 (1906). doi:10.1055/s-0028-1142018
56. P. Wentworth, Jr. et al., Evidence for antibody-catalyzed ozone formation in bacterial killing and inflammation. Science **298**, 2195–2199 (2002). doi:10.1126/science.1077642
57. C.S. Zipitis, A.K. Akobeng, Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. Arch. Dis. Childhood **93**(6), 512–517 (2008). doi:10.1136/adc.2007.128579

Chapter 12
Cell Skeleton

Abstract The shape of a cell is stabilised by a protein scaffold, known as the **cell skeleton**. We distinguish

microfilament (7 – 9 nm diameter) consists of actin and is responsible for
- cytokinesis, the separation of daughter cells after mitosis
- muscle contraction
- cell shape

intermediate filament (10 nm diameter), consists of keratins (in epithelial cells) or vimentin and is responsible for
- stability of the cell against mechanical forces
- growth of nerve axons

microtubules (24 nm diameter), consists of tubulin and is responsible for
- orientation of chromosomes in the equatorial plane during metaphase and their separation into the daughter cells during anaphase of mitosis
- intracellular traffic
- cell motility either by cilia or by crawling
- organisation of the cell wall in plants and yeasts

These principal components of the cell skeleton are linked to each other and to other cell structures by hundreds of accessory proteins. Together, cell skeleton proteins form contacts to other cells and define cell polarity.

It had long been assumed that the cell skeleton is a specific development of **eukaryotes** and not present in prokaryotes. In recent years **eubacterial** homologues of the proteins involved in the formation of all three types of cell skeleton proteins have been found. They seem to serve similar functions as in eukaryotes: stabilisation of cell shape, motility, and cell division. However, our knowledge about the bacterial cell skeleton is limited; its presence in **archaea** is still controversial [3–5]. Here we limit the discussion to eukaryotes.

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_12

12.1 The Microfilament

12.1.1 Basic Actin Structure

The microfilament is formed by **actin** molecules [2]. Actin is a very abundant protein inside cells. About 1–10 % of all cellular protein is actin, leading to a concentration of several hundred µM. Actin is a clover-leaf shaped protein of approximately 40 kDa (see Fig. 12.1). Humans have six actin-isoforms: four α-actins are expressed in muscle, and β- and γ-actin in nonmuscle cells. They are highly conserved. Muscle contraction requires α-actin (see section 13.1.2 on page 306) and amœboid movement of cells β-actin. Little is known about functional differences of actin isoforms.

Free actin molecules (**G-actin**) tend to have MgATP bound to them. Actin is able to polymerise to long, helical filaments, called **F-actin** (see Fig. 12.2). F-actin will have MgADP bound most of the time:

$$\text{G-Actin·ATP} \overset{\text{fast}}{\rightleftharpoons} \text{F-Actin·ATP} \xrightarrow{\text{moderate}} \text{F-Actin·ADP·P}_i \xrightarrow{\text{slow}} \text{F-Actin·ADP} + \text{P}_i$$

Note, however, that ATP-hydrolysis only facilitates actin polymerisation; it is not strictly required and both G- and F-actin can occur with either ATP or ADP bound. Removal of nucleotide leads to rapid denaturation of the protein.

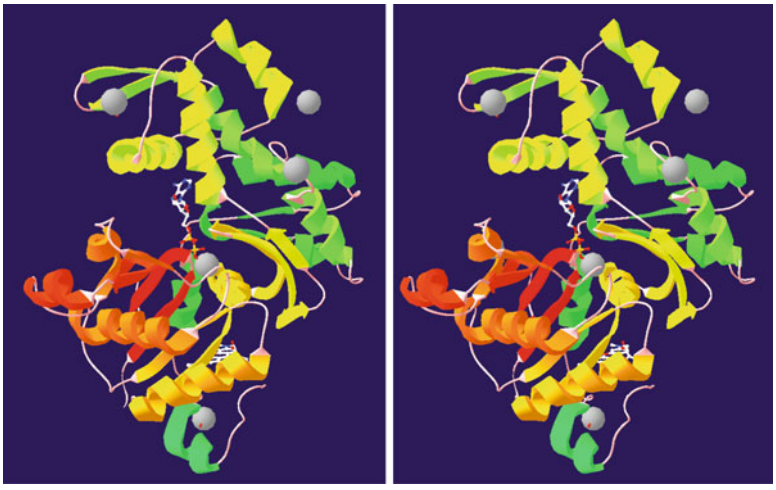


Fig. 12.1 Complex between α-actin, Ca^{2+} and ADP (PDB-code 1j6z). G-Actin aggregates in the absence of other proteins; here crystallisation was made possible by binding tetramethylrhodamine to the C-terminal Cys. In addition, the critical concentration of actin·ADP is much higher than that of actin·ATP. Atomic resolution structures of F-actin are not yet available. The – end of the actin molecule is near the nucleotide binding site

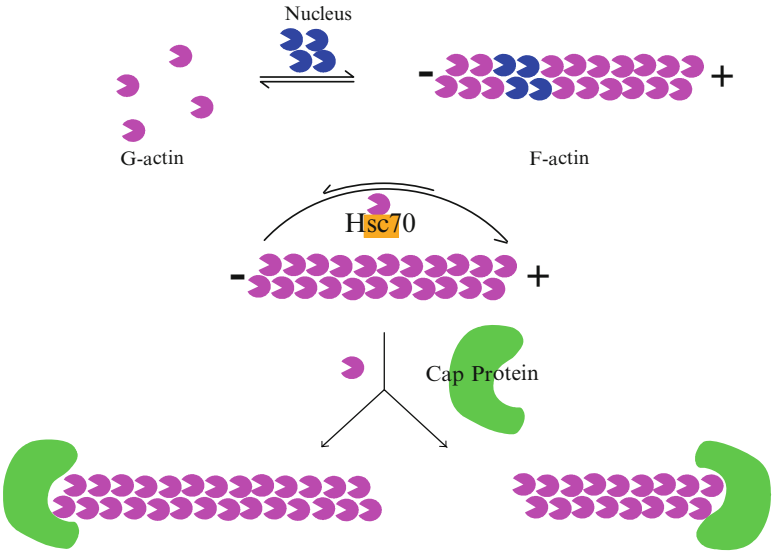


Fig. 12.2 *Top*: Soluble actin monomeres (G-actin) can *in vitro* polymerise to form filaments of F-actin. This process is aided by the presence of nucleï of polymerised actin (but these are not absolutely required). The two ends of the actin filament are not identical; polymerisation occurs preferentially at the positive end. As a result the actin filament "moves" from − to +. *Bottom*: *In vivo* G-actin is chaperoned by Hsc70, the cognate form of the 70 kDa heat shock protein (see page 345). Cap proteins can block either end of the actin filament. If the positive end is blocked, net dissociation from the negative end will lead to shortening of the filament. If the negative end is blocked, net association at the positive end results in filament elongation

Polymerisation of G-actin to F-actin is spontaneous and occurs whenever the actin concentration is above a critical value (which depends on ion concentration, temperature, and other factors). In this respect the polymerisation resembles crystallisation. As in crystal formation it is aided by the presence of nucleation sites, that is, small aggregates of actin already assembled. Once three to four actin molecules have come together, they form a stable nucleus, on which further molecules can assemble. The kinetics of such protein–protein associations has been discussed for amyloid in Fig. 10.5 on page 208. The thermodynamics of filament polymerisation is discussed in Fig. 12.3. *In vivo* the rate of net actin assembly is estimated at about 200 monomers s^{-1}.

Actin filaments have two different ends, designated plus (+) and minus (–). In the older literature, the terms "barbed" and "pointed" ends may be found instead; they refer to how actin-bound myosin looks under the electron microscope. Assembly is preferentially to the **plus-end**, and dissociation preferentially from the **minus-end**. At the minus-end the ATP-binding site of actin is exposed (Fig. 12.1).

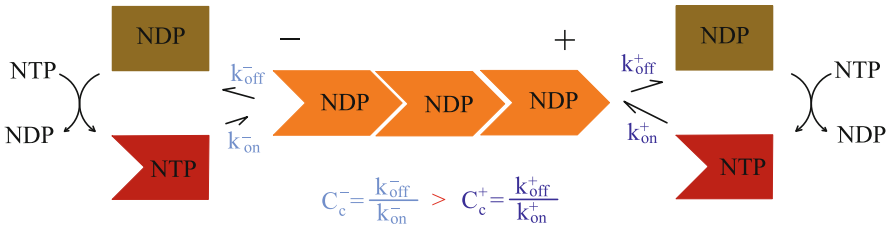


Fig. 12.3 Thermodynamics for the polymerisation of actin and tubulin. Both proteins usually contain nucleotide triphosphates as monomers, but diphosphates in filaments. In the case of actin, the nucleotide is ATP, whereas in tubulin it's GTP. Both proteins are polar, with a + and a – end; the – end corresponds to the nucleotide binding pocket in Fig. 12.1. Steric fit of the monomer is better at the + end, but in the absence of nucleotide hydrolysis this would affect both association and dissociation and the critical concentration of the monomer c_c—at which filaments neither grow nor shrink—would be identical at both ends. Nucleotide hydrolysis and -exchange change that; it stores free energy in the protein lattice. ΔG_{diss} is more negative for the NDP- than the NTP-bound protein. The protein monomers are almost exclusively in the NTP-bound form, because in the cytosol [ATP] $\approx 10 \times$ [ADP]. The association of the NDP-bound monomer, and the dissociation of the NTP-bound monomer are so slow that they can be ignored; they are not shown in the diagram. Then $c_c^- = \frac{k_{off}^-}{k_{on}^-} \quad \wedge \quad c_c^+ = \frac{k_{off}^+}{k_{on}^+}$ Because the rate constants k_{off} and k_{on} refer to different reactants there is no need for $c_c^- = c_c^+$ and usually $c_c^- \approx 15 \times c_c^+$. Therefore, filaments tend to grow on the + end and shrink from the – end

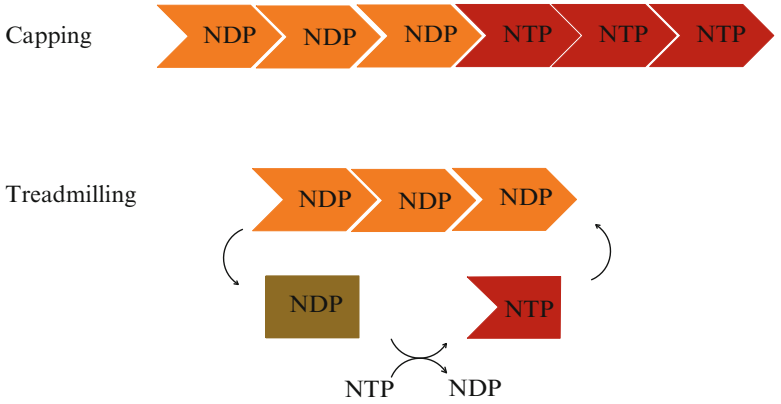


Fig. 12.4 *Top*: Association of monomers at the + end of actin or tubulin filaments can be faster than nucleotide hydrolysis, resulting in a NTP **cap**. Such caps reduce the dissociation of monomers from the + end, and hence stabilise filaments. *Bottom*: If the concentration of monomer is intermediate between c_c^+ and c_c^- then net dissociation at the – end and growth at the + end will occur. The length of the filament will stay constant, but the filament will "move" toward the + end. This is called **treadmilling** ($\approx 2\,\mu m\,h^{-1}$ in the absence of other proteins) and requires energy in form of NTPs

12.1.2 Actin-Binding Proteins

Actin in the cells is about 50 % each soluble and in filaments. [G-Actin] in the cytosol is about 50–200 µM; this is possible despite a c_c of ≈ 1 µM because G-actin molecules are chaperoned by other proteins that serve several functions [2]. **Thymosin** binds to G-actin and prevents it from binding to either end of F-actin. This results in a large storage pool of soluble actin (Fig. 12.5).

Profilin binds to the + end of G-actin and prevents it from binding to the – end of the F-actin. Once the G-actin/profilin complex has bound to the + end of F-actin profilin is released. Profilin and thymosin compete for actin. The net effect of profilin is a regulated transfer of G-actin from inactive, thymosin-bound storage pools to the + end of the growing microfilament. Profilin also acts as an ADP/ATP exchange factor and is in turn regulated by the second messenger **Phosphatidylinositol 4,5-bisphosphate (PIP_2)** and by phosphorylation.

Gelsolin breaks actin filaments into smaller units by binding to an actin molecule in the chain and waiting until there is a gap between this and a neighbouring actin molecule (during random release and rebinding). Gelsolin then enters this gap and prevents re-formation of the actin–actin bond. Gelsolin activity too is regulated by PIP_2 and Ca^{2+}. Dynamin competes with gelsolin [8].

Branching is possible when the **Arp2/3**-complex binds to an actin filament and serves as an anchor for the minus-end of an additional filament, which goes off at a 70° angle to the original one. This branching is tightly regulated and essential for the formation of filopodia. Arp2 and -3 are structurally related to actin and probably originated by duplication of the actin gene.

> WISKOTT-ALDRICH-syndrom protein (WASP) is a regulator of the Arp2/3-binding to actin filaments, encoded on the X-chromosome. Failure to make the functional WASP results in an inherited immunodeficiency and blood clotting disorder (OMIM #301000).

WASP itself is activated by a small G-protein, Cdc42. In plants the Arp2/3 pathway is responsible for correct cell morphology (root hair formation and interlocking of epidermal cells; see Fig. 12.6), but not essential.

Formins nucleate the formation of new microfilaments from soluble actin. Formins act as dimers, each of the monomers can bind one actin molecule. During growth, the formins stay at the + end of the filament, adding one actin monomer after the other on alternating chains. There are 15 formins encoded in the mammalian genome which in addition to their conserved actin-binding sites also have binding sites for various cellular structures. Hence they can anchor actin filaments, for example, to the plasma membrane.

Actin depolymerising factor (ADF) and **cofilin** accelerate treadmilling of actin filaments and recycle G-actin, for example, after cytokinesis. In vertebrates, these

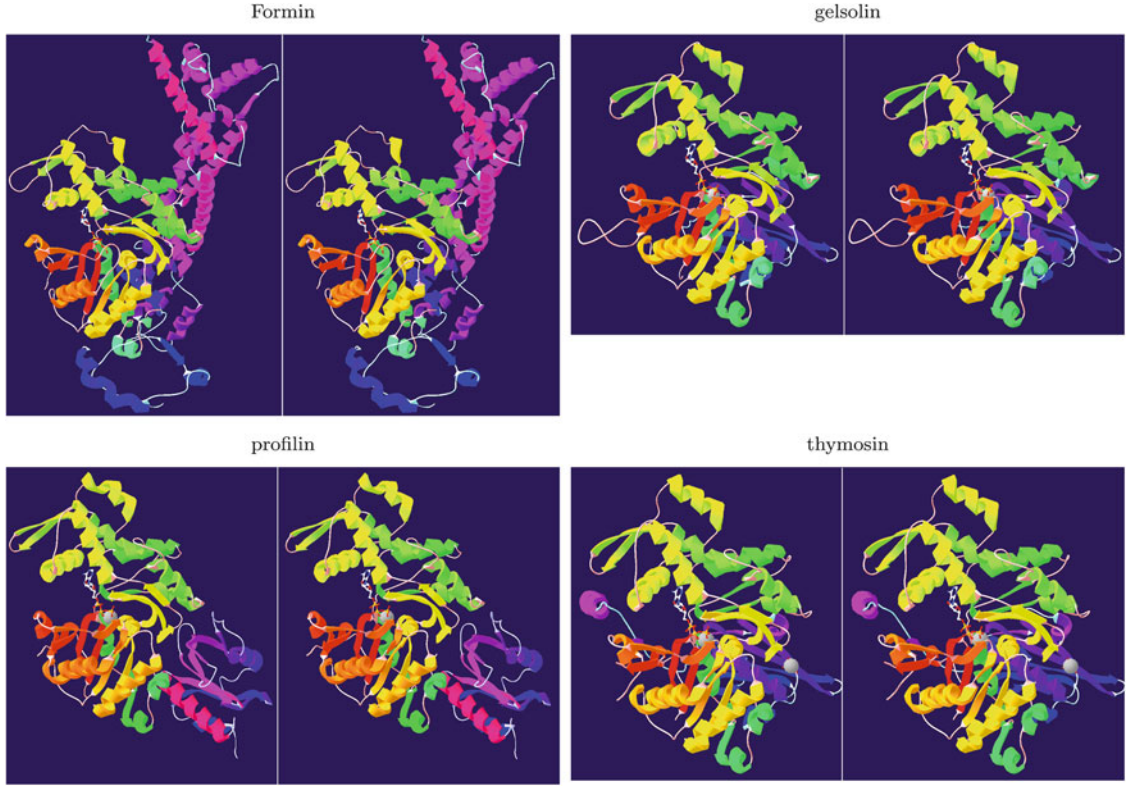


Fig. 12.5 Complexes of actin with formin (PDB-code 4eah), gelsolin (PDB-code 1yag), profilin (PDB-code 2pdb) and thymosin (PDB-code 1t44). Actin molecules (*red to cyan*) are in the same orientation as in Fig. 12.1, to make comparison easier. The other molecules are blue to magenta. Binding sites for the various proteins on actin are different; their binding to actin leads to subtle structural changes

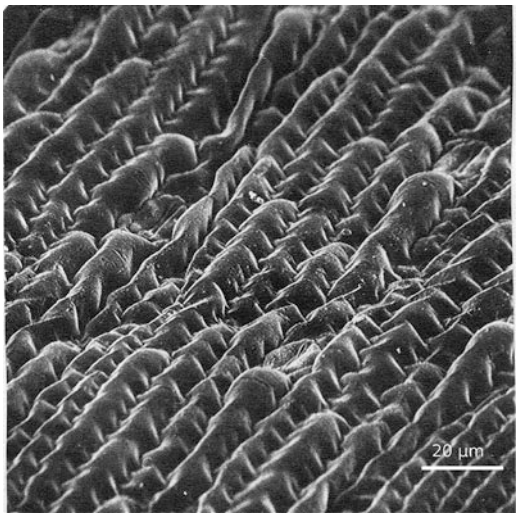


Fig. 12.6 Scanning electron microscopic image of the epidermis of a leaf from maize (*Zea mais* L. 1753). The cells interdigitate, giving strong connections to each other. The cell shape is determined by the actin network in the cell's cortex

are the only members of a large family of similar proteins found in other animals, fungi, and plants. Their affinity for actin is downregulated by phosphorylation on Ser-3. Phosphorylated cofilin stimulates ADP/ATP exchange in actin. PIP2 binding has similar effects. Cofilin also disassembles actin filaments by binding to stretches of ADP-actin and twisting the molecules so that the filament breaks. Because ATP hydrolysis is slower than actin polymerisation, this affects mostly older filaments that had enough time to hydrolyse their bound ATP.

Tropomyosin stabilises the filament by binding at its side, **CapZ** and **tropomodulin** cap the + and – ends of filaments, respectively. This prevents both association and dissociation of actin molecules. They are found mostly in the sarcomere (see page 306).

Toxins That Interfere with Actin Polymerisation

Several toxins interfere with the actin polymerisation/depolymerisation equilibrium (Fig. 12.7). **Cytochalasin D** (see Fig. 12.7, from leaf blight fungi of genus *Helminthosporium*) caps the plus-end of F-actin, preventing further polymerisation. It is used in the laboratory to block cell division at the G1-S transition. **Swinholide** (from the marine sponge *Theonella swinhoei* GRAY 1868) cuts actin filaments, **latrunculin** (from the Red Sea toxic finger-sponge *Latrunculia magnifica* KELLER 1889) binds to G-actin, preventing its polymerisation.

Phalloidin, a poison from the death cap mushroom (*Amanita phalloides* (VAILL. EX FR.) LINK 1833), binds to the minus end of F-actin and prevents its depolymerisation. The nasty bit about poisoning with *A. phalloides*—

(continued)

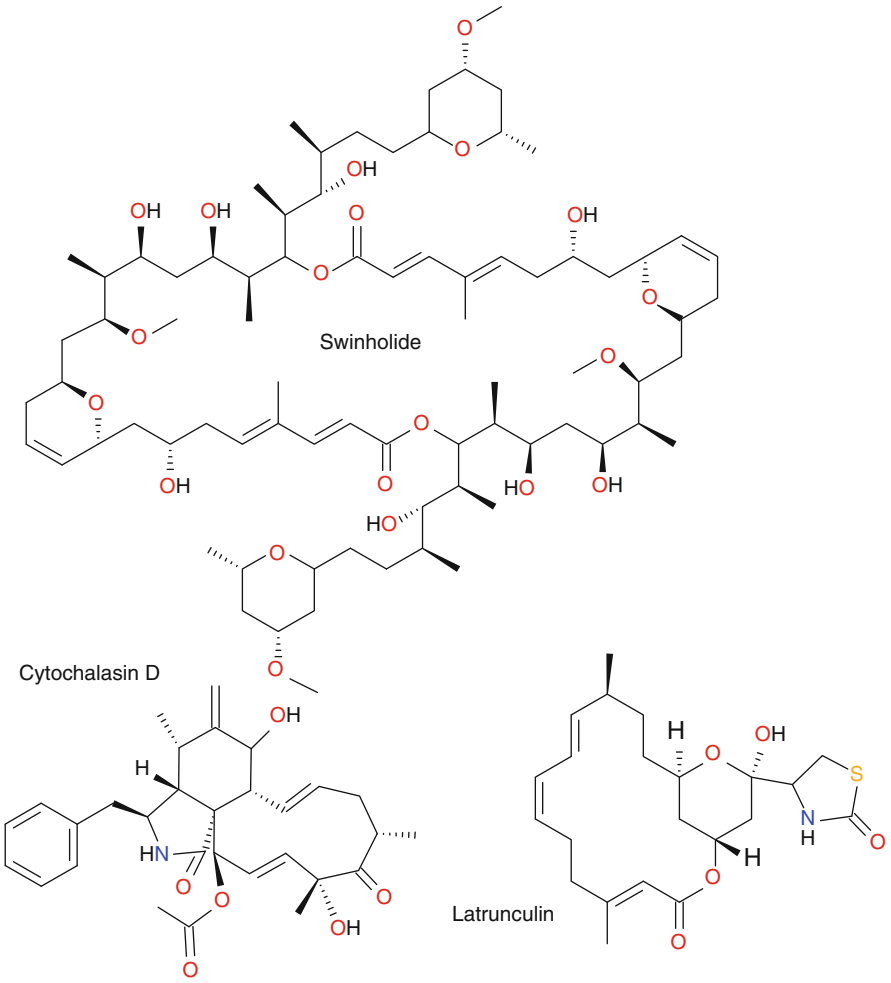


Fig. 12.7 Some compounds that interfere with actin polymerisation. For details see text

due mainly to RNA-polymerase inhibition by amanitins—is that symptoms appear only about three days after ingestion of the mushrooms, when it is too late to empty the patient's GI-tract to minimise absorbtion. On the bright side, fluorescently labelled phalloidin is invaluable as a specific stain for actin filaments in microscopy (see Fig. 12.8).

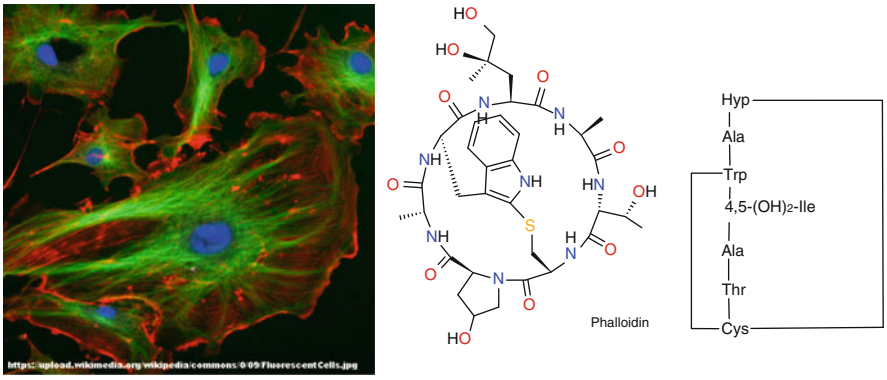


Fig. 12.8 *Left*: Epithelial cells stained for actin (*red*, TRITC-labelled phalloidin), microtubules (*green*, FITC-labelled antibody), and DNA (*blue*, DAPI). Image by JAN R., taken from https://en.wikipedia.org/wiki/Phalloidin#mediaviewer/File:FluorescentCells.jpg. *Right*: Phalloidin is a bicyclic peptide isolated from the death cap mushroom *Amanita phalloides*

12.1.3 Functions of Actin

Cells move by extending their β-actin filaments in one direction; the cytoplasm follows, forming **lamellipodia** (leading edges) and **filopodia** (leading spikes). These will constantly form **focal adhesions**, points of attachment to the substrate. The trailing end of the cell then simply follows, after its attachments to the substrate have been dissolved.

Several intracellular parasites (*Listeria monocytogenes*, Vaccinia virus) can stimulate actin polymerisation, thus pushing themselves through the cell.

One important function of actin polymerisation is the **acrosome reaction** during fertilisation. Once a sperm has made contact with an egg, actin bundles form in the head of the sperm (**acrosomal process**), which pierces the jelly-layer of the egg and allows the sperm nucleus to move into the egg.

12.1.4 Actin-Networks

Actin filaments in the cell are cross-linked into a two-dimensional network below the cell membrane and a three-dimensional network throughout the cytoplasm. Cross-linking is performed by specialised proteins with two actin binding sites separated by coiled-coils or immunoglobulin domains. The actin binding sites

usually have homology to the muscle protein **calponin** and are called CH-domains (calponin homology). Many of these proteins also have calmodulin-like Ca binding sites.

Filamin, ankyrin and **dystrophin** associate with transmembrane proteins like Na^+/K^+-ATPase and anchor actin filaments with respect to the plasma membrane. Actin filaments form a two-dimensional network below the membrane, in the **cortex** of the cell (see Fig. 12.8). This may be further supported by a **spectrin** network.

Some of the transmembrane proteins transmit signals into the cell, for example, **Gp1b-IX** in platelets. This protein binds to blood clots, and signals such binding to the actin filaments via **filamin**.

As a result, the actin network (and therefore the platelet) contracts, closing the wound.

In muscle cells the actin network is connected to transmembrane glycoproteins via **dystrophin**. The glycoproteins in turn link the muscle cell to the extracellular matrix.

Mutations in the dystrophin gene lead to DUCHENNE **muscular dystrophy** (OMIM #310200), an X-linked genetic disease with onset around age 4; patients are wheelchair-bound by age 10 and die at a median age of 17, usually from respiratory failure. The disease is usually caused by frameshift mutations. A milder form is BECKER **muscular dystrophy** (OMIM #300376). In this disease onset occurs at around 12; patients frequently survive to advanced age. X-inactivation in carrier females has been observed in both diseases, often as mosaic.

Actin bundles, cross-linked to the membrane, support **microvilli** and **filopodia**. Many cross-links between the actin filaments in the bundles make them stiff.

12.2 Microtubules

Microtubules are involved in:

- Beating of **cilia** and **flagella** (see Sect. 13.3 on page 313).
- **vesicle transport** in the cell (see Chap. 17 on page 393) and phagocytosis.
- Formation of the **mitotic spindle** (see Sect. 13.4 on page 317).
- Stabilisation of **cell shape**, elasticity of erythrocytes and the stability of the long axons in nerve cells.

- Keeping of **cell organelles** such as GOLGI-apparatus, ER, and mitochondria in their proper place. If microtubules depolymerise (during mitosis or with colchicine), ER and GOLGI-apparatus fragment into many vesicles.

12.2.1 Microtubule Structure

12.2.1.1 α- and β-Tubulin

Microtubules are made by the ordered polymerisation of **tubulin**. They have a diameter of 24 nm; their length can reach several hundred µm.

Microtubules are made from two isoforms of tubulin, called α and β, which form a heterodimer (see Fig. 12.9). α-Tubulin contains **GTP** bound in a pocket which is covered by the β-tubulin and therefore cannot be exchanged. β-Tubulin has an open binding pocket, which binds GTP. After polymerisation, GTP is hydrolysed to GDP; exchange of this GDP for GTP occurs during depolymerisation. The GTP in the α-tubulin is not hydrolysed at all.

These heterodimers polymerise into long chains, called **protofilaments** . Thirteen such protofilaments form a microtubule (see Fig. 12.10). The large number of both longitudinal and lateral protein–protein bonds makes microtubules stiff; any bending or elongation would break these bonds.

Because all tubulin-dimers in a microtubule have the same orientation, one end (called **minus-end**) is capped with α-, and the other (called **plus-end**) with β-tubulin.

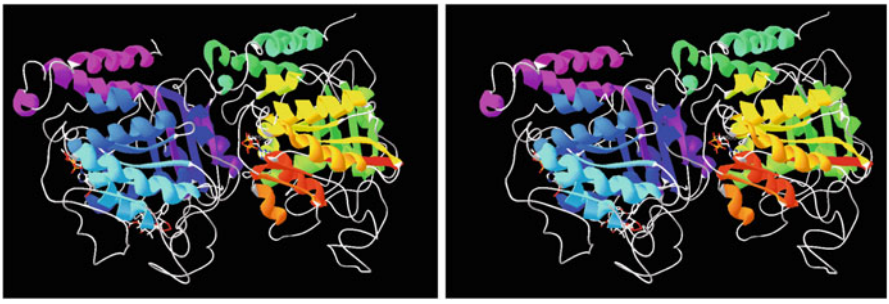


Fig. 12.9 Stereo view of the α- (*red-green*) and β-tubulin (*blue-purple*) dimer (PDB-code 1TUB). α-Tubulin has GTP bound to it; the binding site is occluded after dimerisation. β-Tubulin contains bound GDP. In addition this subunit carries one molecule of paclitaxel, an anticancer drug, to stabilise subunit interactions

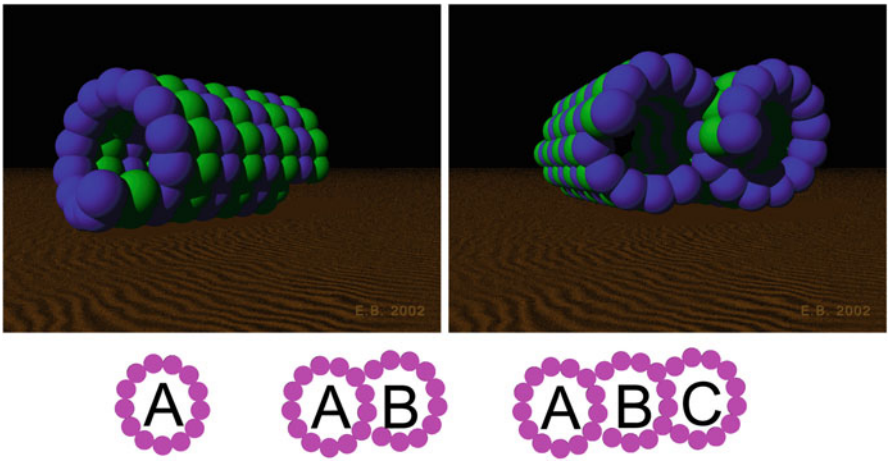


Fig. 12.10 Microtubules consist of α- and β-tubulin. The α, β-tubulin-heterodimers polymerise into tubes, which consist of 13 **protofilaments** (long rows of dimers). This is called an A-tubule. A further 10 protofilaments can be added to form a doublet, found in cilia and flagella. The second tubule is called a B-tubule. A further 10 protofilaments make a triplet, found in basal bodies and centrioles. This third tubule is called a C-tubule, and distinguished from a B-tubule by the fact that it is connected to a B- rather than an A-tubule

12.2.1.2 Mechanism of Tubulin Assembly and Disassembly

The minus ends of all microtubules in the cell start from a central hub, the **Microtubule organising centre (MTOC)**. These may be visible in electron microscopy as **basal body** or **centrosome**. Most cells have one MTOC, the centrosome which also contains the centrioles. Epithelial cells (and plant cells) have many MTOCs, which organise a network of tubules in the cell cortex. Cell polarity in these cells is linked to tubule orientation. Located on the surface of the MTOC are the 25 S-ring complexes, which consist mainly of γ-tubulin (γ Tu-RC). From each of these ring complexes a tubule can start. The mechanism of microtubule organisation is not yet known, however.

The microtubules radiate from the MTOC in all directions. When they meet obstacles, such as the cell membrane, the growing microtubule pushes the MTOC away from it. This force acting on the MTOC from all directions locates it in the centre of the cell. Moving from the – ends of the tubules in the centre of the cell toward the + ends in the periphery establishes a coordinate system for organelle localisation and intracellular transport.

During depolymerisation of tubulin the interactions between the protofilaments break first, so that under the electron microscope the ends of a shrinking microtubule look frayed, brush-like (see Fig. 12.11). Growth on the other hand must occur by the even addition of small units to the ends, because they look smooth in the EM. Also, disassembly is a much faster process than assembly (7 vs 1 μ m/s).

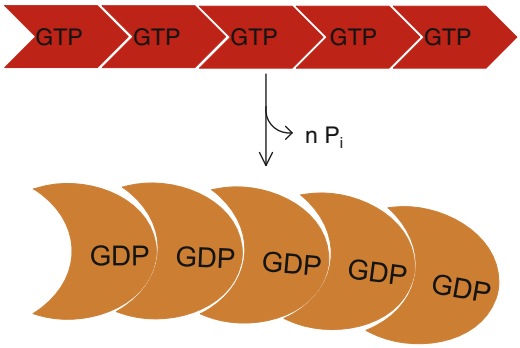


Fig. 12.11 Dynamic instability of microtubules. Tubules grow by association of tubulin dimers, which have GTP bound, to the tubule. The resulting GTP-cap prevents tubule dissociation. Hydrolysis of bound GTP to GDP changes the conformation of tubulin heterodimers. This has two consequences:

- Strength of protein–protein interactions within a protofilament is decreased; dissociation becomes easier.
- A curvature is introduced into the protofilament; the protofilaments can no longer associate into a tubule. The ends of the tubule become frayed; this too increases the rate of dissociation.

Thus, once the GTP-cap has been lost, the tubule will shrink rapidly until the cap is regained and net growth begins again. **Catastrophe factors** and **microtubule associated proteins** control the frequency of catastrophic GTP breakdown

12.2.1.3 Other Proteins Associated with Microtubules

A number of proteins have been found to copurify with microtubules in cell fractionation studies; fluorescent antibodies against these proteins are found on microtubules in fluorescence microscopy. These are called **microtubule associated proteins (MAP)**. Based on sequence similarity, these are grouped into type-I and type-II MAPs. These proteins can change the rate of tubulin polymerisation and depolymerisation, or cross-link microtubules to intermediate filament, membranes, or with each other. MAPs have recently been reviewed in [7] (during mitosis) and [1] (during interphase).

12.3 Intermediate Filament

Intermediate filaments IF have a structural role; no transport functions seem to be associated with them, nor are they involved in cell motility. Because they are bound to transmembrane proteins, which link cells together, they are responsible for the stability of tissues. However, some animals with exoskeleton make do entirely without IF, and even some of our own cells, including the SCHWANN cells that make up the myelin sheets around our nerve axons, do not express IF proteins.

Intermediate filaments are quite stable and survive extraction of the cells with detergents and high salt concentrations, which remove the membrane, microfilament, and microtubules. They can be depolymerised with urea, purified, and repolymerised by urea removal. Intermediate filament proteins polymerise into helical rods (see Fig. 12.13). There are a large number of lateral hydrophobic interactions between subunits in these filaments (see Fig. 12.12), but relatively few in the longitudinal direction. As a result, intermediate filaments have rope-like properties; they can bend but don't stretch much. Assembly does not require ATP.

12.3.1 IF-Proteins Are Cell-Type Specific

Six classes of IF-proteins can be distinguished:

Type I Acidic keratins are found in epithelial cells.

Type II Basic keratins are also found in epithelia. They form heterodimers with the acidic keratins, which then polymerise into fibres (see Fig. 12.13). Many isoforms of both types of keratin have been described; we distinguish hard keratins (hair, horn, claws) and cytokeratins that are found in internal epithelia.

Type III Vimentin (leucocytes, blood vessel endothelia, fibroblasts), Desmin (stabilises sarcomeres in muscle cells), peripherin (found in peripheral neurons, function unknown), and glial fibrilary acidic protein (glial cells and astrocytes) can form homo- and heterodimers.

Type IV Neurofilaments NF-L, -H, -M, and internectin are found in the nervous system. They are responsible for radial growth of the axon and determine its diameter.

Nonstandard type IV Philensin and phakinin are found in the lens of the eye.

Type V Lamin[1] -A, -B, and -C are found in the nuclear lamina. Lamin-A and -B are produced from the same gene by different splicing; lamin-C is encoded by a separate gene. They maintain the nucleus and have an essential role during cell division (see, e.g., [6]).

[1] Do not mix up lamin and laminin, a protein in the extracellular matrix (see Sect. 14.1 on page 323).

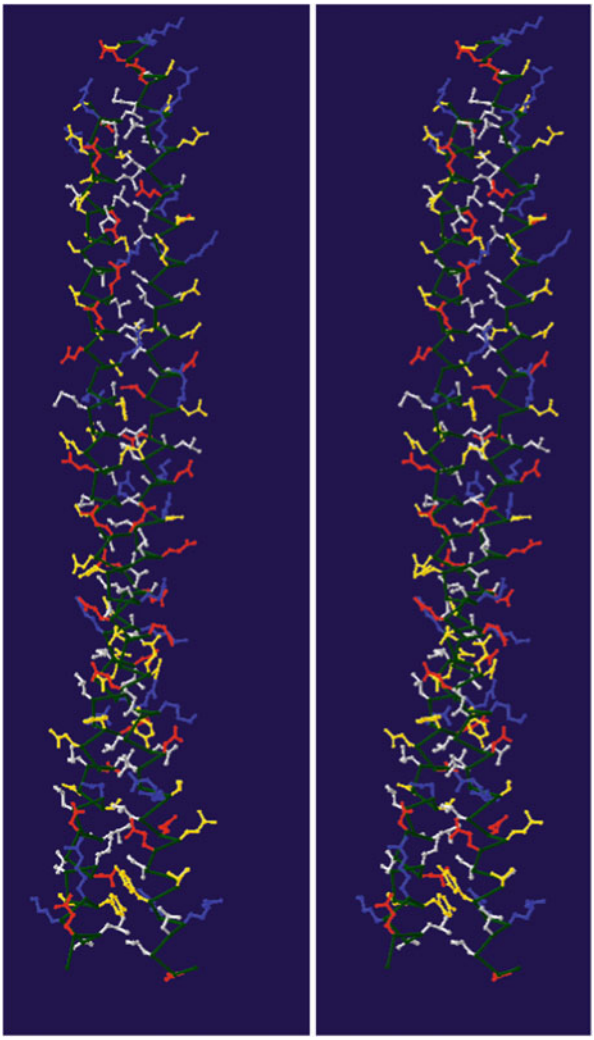


Fig. 12.12 Structure of a coiled-coil segment of human keratins KRT14 and KRT5 (PDB-code 3tnu). No structures for entire keratin molecules are available; as with all fibrous proteins they are difficult to crystallise. In this diagram the carbon backbone is shown in dark green, the side-chains of hydrophobic amino acids in grey, polar in yellow, negative in red, and positive in blue. Note the many hydrophobic residues at the interface between the proteins, which form hydrophobic interactions

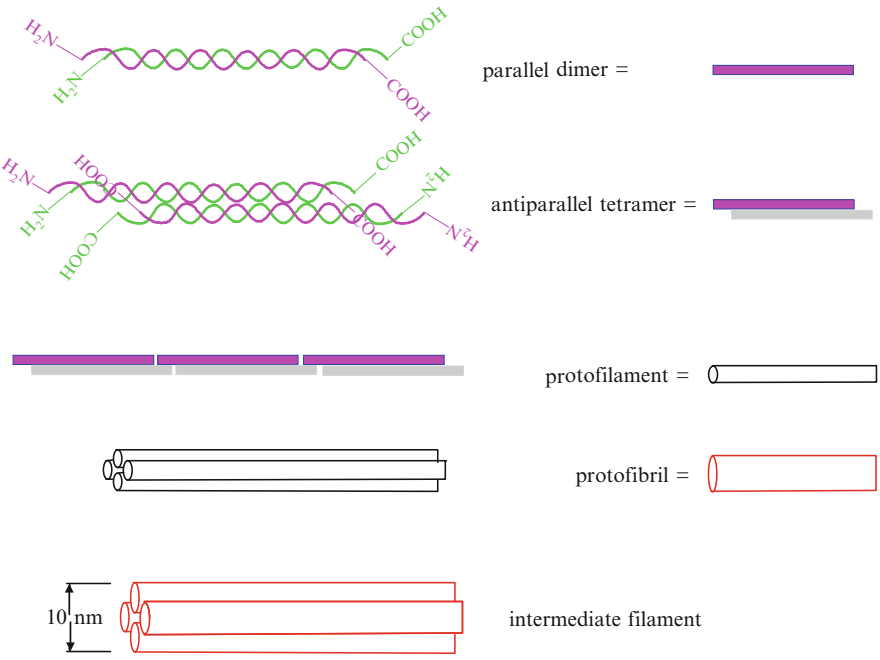


Fig. 12.13 IF-proteins contain long helical segments, with more unordered heads and tails. Two IF-proteins dimerise in parallel with their helical segments winding around each other (coiled-coil, held together mostly by hydrophobic interactions). Two such dimers then polymerise into an antiparallel tetramer. These form protofilaments; four of those form protofibrils and four of those intermediary filaments. The head-parts of the protein stick out from the filaments like the fibres from a lamp-brush, which can be seen in EM-pictures of the IF. Note that in IF tetramers both ends are identical; there are no + and – ends as in tubulin or actin

IF-Proteins as Tumour Markers
Because IF-protein are cell-type specific, they can be used to identify the origin of tumor cells in biopsies. For example, epithelial cells, and tumours derived from them, express keratins whereas mesenchymal cells express vimentin. Antibodies against these proteins can be used to distinguish tumours of these two origins. This is important for the selection of the right treatment for gastrointestinal or breast cancers.

12.3.2 Structure of Intermediate Filaments

It appears that the tetrameres of IF-proteins form a **soluble pool**, which is in dynamic equilibrium with the filaments, similar to what you have learned about

actin and tubulin. Tetrameres can be isolated from cytosol. If labelled IF-proteins like keratin are microinjected into a cell they become part of the IF after a couple of hours.

12.3.3 Intermediate Filaments and Cell Cycle

IF are fairly stable structures, much more so than microfilaments and microtubules. However, during mitosis the IF must be reorganised. IF protein head domains can be phosphorylated by **Cdk2** (cell cycle dependent protein kinase) on serine residues, which leads to their depolymerisation.

12.3.4 Other Proteins Associated with Intermediate Filaments

The IF are cross-linked with each other, with microtubules, and with the cell membrane by specialised proteins, collectively called **Intermediate filament associated proteins (IFAP)**. The function of the known IFAPs is only to increase the stability of the cell skeleton; they do not influence the polymerisation/depolymerisation equilibrium nor do they act as motor proteins.

Treatment of cells with colchicine (see Fig. 13.12 on p. 320) causes the microtubules to depolymerise. IF remain intact, but lump together in an unordered pile near the nucleus. Thus a microtubule structure is required to organise IF.

Plectin and **ankyrin** are examples of IFAPs. Plectin cross-links IF to microtubules and to the lamin network of the nucleus. Ankyrin connects them to plasma membrane proteins such as Na^+/K^+-ATPase, and to microtubules. They also bind to cell/cell- and cell/matrix-junctions (see Sect. 14.3 on page 337).

In muscle, where the forces created in a sarcomere are distributed to other sarcomeres, the cell membranes of neighbouring cells are connected by a **desmin** network.

Mutated IF proteins such as keratin do not kill a cell, but weaken its shape and its interactions with its neighbours. Thus forces can no longer be distributed between cells and tissues are weakened. An example for this is **epidermolysis bullosa simplex** (OMIM #131900), where skin keratin is mutated. As a result dermis and epidermis easily separate, forming blisters.

Mutations in plectin not only affect the intermediate filament, but also muscle organisation, resulting in **epidermolysis bullosa simplex with muscular dystrophy** (EBS-MD, #226670). Onset is early; progression is also marked by growth retardation and anæmia.

12.4 Exercises

12.4.1 Problems

12.1. A 65-year-old man is to be treated for gastrointestinal cancer. In order to select the right cytostatic drug, a biopsy is taken and immunostained for which of the following proteins?

A Vimentin and keratin
B Actin and tubulin
C Philensin and phakinin
D Lamin-A, -B, and -C

12.4.2 Solutions

12.1 Vimentin and keratin occur in tumors of mesenchymal and epidermal origin, respectively. Actin and tubulin occur in all cells. Philensin and phakinin occur in the eye lens. Lamins occur in the nuclear lamina of all cells.

References

1. A. Akhmanova, M.O. Steinmetz, Tracking the ends: a dynamic protein network controls the fate of microtubule tips. Nat. Rev. Mol. Cell. Biol. **9**(4), 309–322 (2008). doi:10.1038/nrm2369
2. C.G. Dos Remedios, D. Chhabra, M. Kekic, I.V. Dedova, M. Tsubakihara, D.A. Berry, N.J. Nosworthy, Actin binding proteins: Regulation of cytoskeletal microfilaments. Physiol. Rev. **83**(2), 433–473 (2003). doi:10.1152/physrev.00026.2002
3. D.M. Faguy, W.F. Doolittle, Cytoskeletal proteins: The evolution of cell division. Cur. Biol. **8**(10), R338–R341 (1998). doi:10.1016/S0960-9822(98)70216-7
4. P. Findeisen, S. Mühlhausen, S. Dempewolf, J. Hertzog, A. Zietlow, T. Carlomagno, M. Kollmar, Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. Genome Biol. Evol. evu187 (2014). doi:10.1093/gbe/evu187
5. E.V. Koonin, N. Yutin, The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. Cold Spring Harb. Perspect. Biol. **6**(4), (2014). doi:10.1101/cshperspect.a016188
6. H. Lodish et al., *Molecular Cell Biology*, 7th edn. (W.H. Freeman and Company, New York, 2012). ISBN 978-1-4292-3413-9
7. N. Tamura, V.M. Draviam, Microtubule plus-ends within a mitotic cell are 'moving platforms' with anchoring, signalling and force-coupling roles. Open Biol. **2**, 120132 (2012). doi:10.1098/rsob.120132
8. S. Sever, J. Chang, C. Gu, Dynamin rings: Not just for fission. Traffic **14**(12), 1194–1199 (2013). doi:10.1111/tra.12116

Chapter 13
Motor Proteins and Movement

Abstract The cell skeleton forms a scaffold, along which motor proteins can move. These proteins convert the chemical energy of ATP-hydrolysis into mechanical energy. Movement is unidirectional, either from minus to plus or *vice versa*. The most important systems are microfilament/myosin and microtubule/kinesin and -dynamin.

13.1 Myosin Moves Along Actin Filaments

Humans have 13 myosin genes, but only the functions of myosin-I, -II, and -V are known. Myosin-II is required for muscle contraction and cytokinesis, myosin-V for the transport of intracellular vesicles along the microfilament, and myosin-I on the plasma membrane interacts with actin-bundles in microvilli.

13.1.1 Myosin Structure

All myosin isoforms consist of an N-terminal head and a C-terminal tail of different length (see Fig. 13.1). In myosin-II and -V α-helical sections in the tails interact to form dimers; myosin-I molecules do not have this section and stay single. The tails of myosin-I and -V bind to **membranes**. Aggregation of tails in myosin-II forms a large supermolecular structure, the **thick filaments**.

Several regulatory **light chain** molecules bind to the neck regions of myosins; they have regulatory function and bind Ca^{2+}. In myosin-I and -V these include **calmodulin**.

The head is the business-end of myosin; it contains the actin- and ATP-binding sites. The ATPase-activity of myosin is actin-activated (Fig. 13.2).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_13

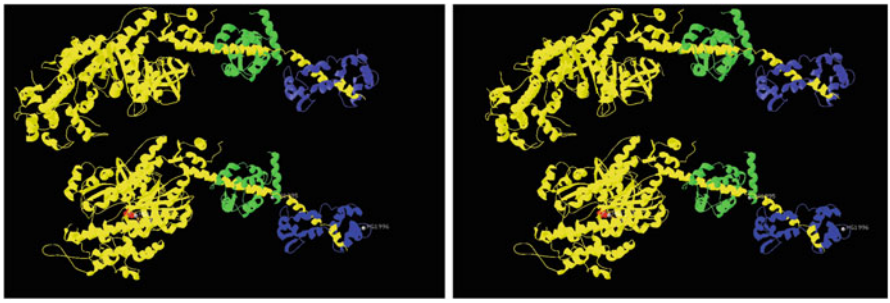


Fig. 13.1 Stereo view of X-ray crystallographic structures of the head and neck regions of scallop muscle **myosin-II**, without (*top*, PDB-code 1DFK) and with (*bottom*, PDB-code 1DFL) bound ADP+Vanadate (ATP-analogue). The tail region was removed by protease treatment to allow crystallisation. The molecules have about the same orientation. Note the extensive conformational changes upon ATP-binding

13.1.2 Myosin-II

The power/weight ratio of muscle is about 1.5 times lower than that of a car engine and about tenfold lower than that of an jet engine. We have three classes of muscle: **skeletal, smooth,** and **heart** muscle. Skeletal muscle can contract rapidly but also tires quickly. Smooth muscle contractions tend to be slower, but more prolonged. Heart muscle is specialised for the rhythmical contraction of this organ. Myosin-I in intestinal brush border membranes covers about 40 nm/s; myosin-II in skeletal muscle about 4.5 µm/s. Each individual cycle results in movement of 5–10 nm (the diameter of actin is about 5 nm), with a force of 1–5 pN (this is equivalent to the force of gravity acting on an *E. coli* cell).

Muscle activity may be **isotonical** (contraction at constant force) or **isometrical** (force is increased while length stays the same).

13.1.2.1 Skeletal Muscle

Actin and myosin form a highly organised actomyosin complex (Fig. 13.3), which results in a microscopically visible **stria** (see Fig. 13.4).

Regulation of Muscle Contraction

Muscle contraction is stimulated by Ca^{2+}-release from the **sarcoplasmic reticulum (SR)** (a specialised ER), where it is stored. Depolarisation of the cell after a nerve impulse leads to the opening of Ca^{2+}-channels in the SR. Once the muscle has been repolarised, Ca^{2+} is moved into the SR by the **Ca^{2+}-ATPase**, a P-type membrane transporter (see Sect. 18.2.1.1 on page 426). Around the actin-filaments

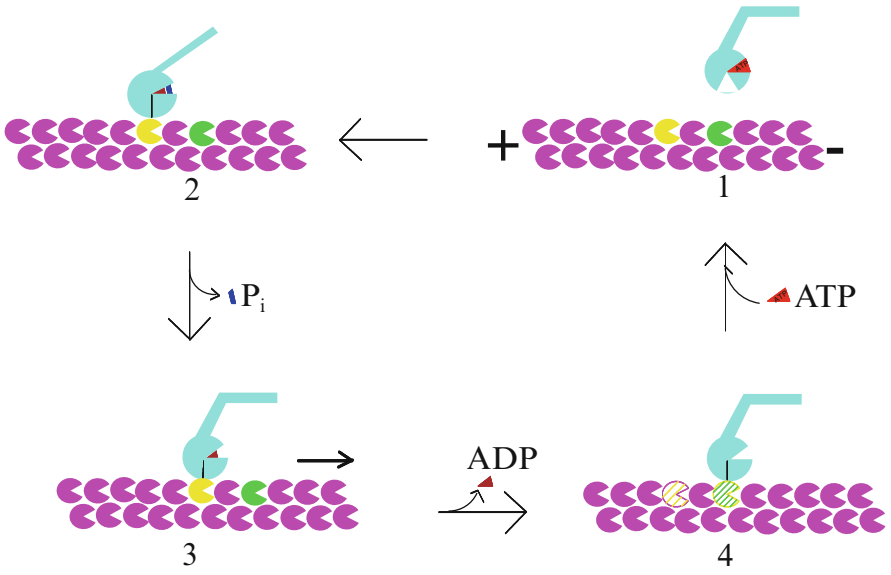


Fig. 13.2 Actin-myosin interactions. (1) ATP-binding to myosin results in opening of its actin-binding site. Actin is released. (2) The bound ATP is hydrolysed to ADP and P_i; the joint between stem and head of myosin extends, and a new actin molecule is bound. (3) P_i is released, the myosin joint flexes. As a result, the actin chain is pushed toward the minus end. This is the actual power stroke. (4) ADP is released, closing the cycle. Such interleaving of mechanical and chemical reactions ensures **tight coupling** between both, preventing unnecessary hydrolysis of ATP (a myosin molecule alone hydrolyses about 1 molecule of ATP per minute, in the presence of actin-filament that rate increases to 1000/min). This drawing shows monomeric myosin-I; myosin-II and -V are dimers, but work in a similar way

is wound a complex of **troponin** (a multisubunit protein) and **tropomyosin**. Tropomyosin molecules form a long continuous chain, which winds around the actin microfilament in the helical grove and holds the troponin in place. A rise in Ca^{2+}-concentration changes the conformation of this complex and exposes myosin-binding sites on actin. Additionally, Ca^{2+} binds to the regulatory light chain of myosin. This changes the conformation in the neck of myosin and allows myosin-actin interactions.

Titin (OMIM *188840) is also involved in regulation. It is a very big protein, 3.0–3.9 MDa depending on isoform. All isoforms are encoded by the same gene and result from alternative splicing. Splicing affects the length of the elastic elements of titin, which are located in the I-band. The longer and the more elastic titin molecules are the "softer;" a muscle cell responds to passive stretching. This is particularly important in the heart muscle, which must allow the passive filling of the heart with blood during the diastole. The titin isoform composition in a muscle cell is regulated in a tissue- and developmental stage-dependent manner; it changes during

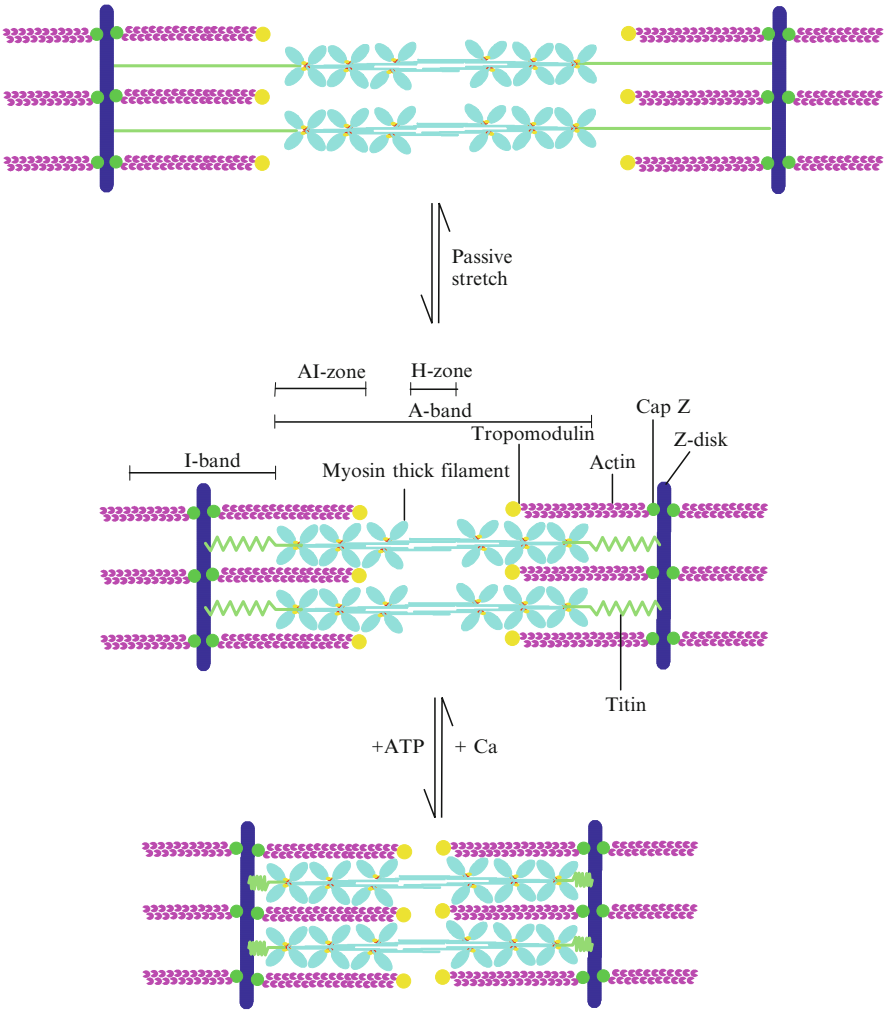


Fig. 13.3 Sliding filament model of muscle action: The **sarcomere** is the smallest repetitive unit in skeletal muscle. Many sarcomeres in series make a **myofibril**, many myofibrils in parallel fill a muscle cell (**myocyte**), and many myocytes form a muscle. Repeated binding of myosin heads to the actin filaments leads to muscle contraction; under the influence of ATP actin is released from myosin and the muscle relaxes. Muscle strength is the result of the summed binding strength of millions of myosin heads to actin microtubules. Passive stretching of muscle is possible, but elastically resisted by **titin** molecules that link the thick filaments to the Z-disks. Actin microfilaments are stabilised by capping proteins, **tropomodulin** at the minus and **Cap Z** at the plus end. **Nebulin** (not shown) binds along the entire length of the microfilaments and stabilises them. The association of actin and myosin in muscle results in an **actomyosin complex**. The Z-disk is composed mainly of α-actinin

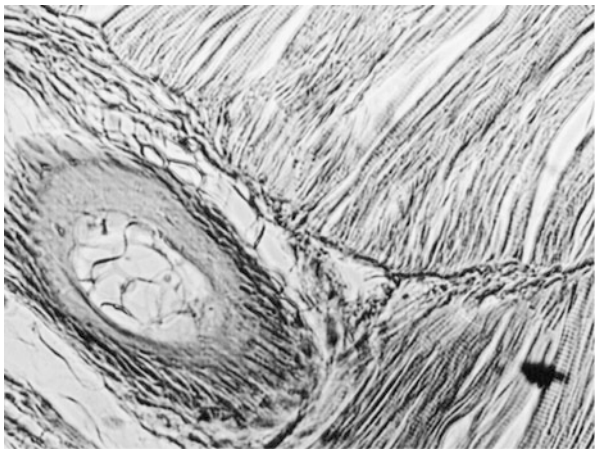


Fig. 13.4 As a result of its structure, skeletal muscle appears striated in histological preparations (Zebrafish, HE staining)

pathophysiological processes such as terminal heart insufficiency. We know little about the regulatory mechanisms; they are controlled in part by thyroxin and insulin, whose level controls the activity of the splicing factor RBM20 (OMIM #613172). In addition, titin stiffness is regulated by hormone-dependent phosphorylation in the N2Bus-domain (decreases stiffness in response to catecholamines, NO, natriuretic peptide), and the PEVK-domain (increases stiffness in response to insulin and angiotensin).

> Problems with titin may result in muscular dystrophies or cardiac myopathies. About 30 % of diastolic heart failure patients have changes in titin isoform composition, mutations in the titin gene, or failure to respond to hormonal regulation of titin stiffness [7].

Titin also binds other regulatory proteins in a mechanosensitive manner, including some calpains and some Zn-finger proteins. We do not currently understand the function of these interactions.

13.1.2.2 Smooth Muscle

In smooth muscle the actomyosin complexes do not form ordered arrays as in skeletal muscle. Instead, they end in **dense bodies** and **attachment plaques**, the latter linking them to the cell membrane. Attachment plaques contain both α-actinin (same as in Z-disks) and **vinculin**. Smooth muscle contains **caldesmon** instead of troponin. Ca^{2+}-dependent phosphorylation of caldesmon and of myosin light chains is required for smooth muscle contraction. Thus smooth muscles contract

much more slowly than skeletal muscle. Ca^{2+}-concentration in smooth muscle is regulated not only by nerve impulses, but also by hormones and growth factors.

Rigor Mortis
Muscle relaxation occurs when ATP-binding to myosin leads to release of bound actin. After death the ATP-stores of the cell are rapidly depleted, but actin-myosin interactions are still possible. This results in ***rigor mortis***, the stiffness of the dead. Only after lysosomal proteases have partially digested the muscle does it become soft again. Understanding of these processes is important for determination of the time of death in forensic medicine.

There is also an everyday application of this knowledge: meat prepared immediately after slaughter is tough and hard; "hanging" it in a cool place for a couple of days makes it softer. Alternatively, the meat may be marinated in protease-containing fruit juices (pineapple, papaya). "Meat tenderisers" contain such proteases in powdered form.

13.1.2.3 Cytokinesis

Myosin-II and actin are found also in the cleavage furrow of mitotic cells. They form a **contractile ring** of parallel molecules around the cell. These molecules slide past each other and thereby reduce the diameter of the contractile ring, until the daughter cells are separated. If the expression of myosin-II is prevented, cells will still undergo mitosis, but because cytokinesis is prevented, multinuclear cells (**syncytia**) are formed.

13.1.3 Myosin-I

Myosin-I stabilises microvilli by linking their plasma membrane to the actin bundles in their core. Myosin-I copurifies with certain GOLGI-derived **vesicles** and is apparently responsible for their movement along actin microfilaments in the cell. In single-cell organisms such as amœba and cilliata it is required for contraction of the **contractile vacuole**, and therefore control of osmolarity.

Cell movement requires sliding of myosin-I on growing actin-filaments. **Profilin** on the leading-edge membrane of the cell stimulates the extension of microfilaments in the direction of cell movement. At the same time **cofilin** stimulates depolymerisation of actin at the minus end. **Actin cross-linking protein** causes the formation of bundles from actin-filaments. These form the highways along which myosin-I slides, moving intracellular membrane systems and therefore also the cytoplasm. An extended lamellipodium will form a new focal adhesion to the substrate, at the

same time the trailing edge detaches from old adhesions, leaving them behind. The retraction of the trailing edge requires myosin-II, which can be localised there by immunofluorescence microscopy, whereas myosin-I is found at the leading edge.

13.1.4 Myosin-V

Myosin-V is required for the transport of exocytotic vesicles to the plasma membrane.

13.2 Kinesin and Dynein Move Along Microtubules

If one observes living cells under the microscope, one can see organelles and vesicles moving over long distances on straight paths. This movement is clearly different from random, BROWNIAN motion. Such directed transport can cover large distances, for example, in nerve cells, where proteins and membranes—synthesised on ribosomes and ER in the cell body—need to be moved along the axon into the synapse to replace material lost during exocytosis. An axon can be several meters long (whale, elephant). Thus if one incubates nerve cell bodies of the sciatic nerve (found in a dorsal root ganglion) with radioactive amino acids and then measures the concentration of radioactivity in different parts along the nerve at various time points, one can measure the speed of transport. This is called **anterograde transport**, from the centre of the cell toward the periphery, from the – end of microtubules to the + end. This transport requires kinesin. Transport from the synapse to the lysosomes in the cell centre (**retrograde transport**) also exists; it follows the microtubules from + to – and requires dynein as the motor protein [8]. Both anterograde and retrograde movement are ATP-dependent. Interestingly, transport speed varies. Vesicles move with 1–3 μm/s, cytoskeletal components with about 1/100 of this speed. Organelles such as mitochondria move at an intermediate velocity.

13.2.1 Kinesin Is Responsible for Anterograde (Minus to Plus) Transport

There are 14 isoforms of kinesin in humans, with different structure and function. Structurally, kinesin is related to myosin. A typical kinesin consists of 2 heavy and 2 light chains, with a total molecular mass of 380 kDa. Each of the heavy chains carries an N-terminal head, which has an ATPase-domain and a tubulin binding site. The C-terminal ends wrap around each other in a coiled-coil structure (see Fig. 13.5), these and the light chains together interact with the "cargo". Different kinesin isoforms are responsible for transporting different cargos.

Incubation with the nonhydrolysable ATP analogue AMP-PNP leads to tight binding of kinesin to microtubules, the protein is released only after the AMP-PNP has been replaced with ATP (Fig. 13.6).

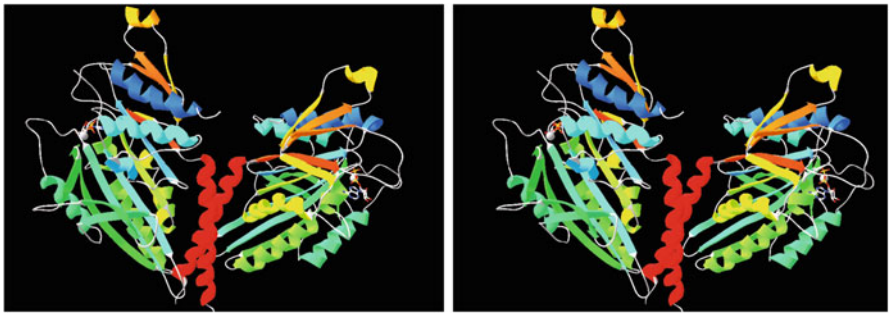


Fig. 13.5 Stereo view of two kinesin heavy chain fragments from the fruit fly *Drosophila melanogaster*, with ADP bound to them (PDB-code 1CZ7). The light chains were removed to ease crystallisation of the protein; they would be located somewhere at the end of the coiled coil

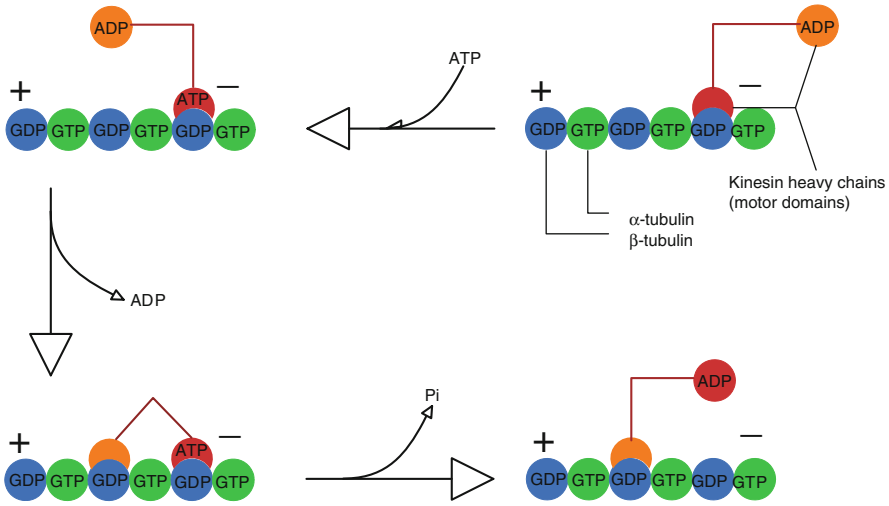


Fig. 13.6 Coupling of kinesin movement with its ATPase cycle by "hand-over-hand" movement. Kinesin is a heterotetramer with two heavy and two light chains. In solution it shows twofold symmetry, but once bound to tubulin the two protomers become different [11]. The actual motor is formed by the heavy chains. ATP-binding to one of the motor domains (*brown*) leads to rotation; that brings the second motor domain (*orange*) over the next β-tubulin (*green*). Release of ADP from the second motor domain leads to tubulin binding. P_i-release from the first motor domain leads to its release from β-tubulin. The cycle is then repeated, with the motor domains switching roles. Only one of the tubulin protofilaments of a tubule is drawn

13.2.2 Dynein Is Responsible for Retrograde (Plus to Minus) Movement

Dynein complements kinesin in that it moves along the microtubules **from the plus-to the minus-end**, that is, in the **retrograde** direction. There are several dynein isoforms, each consisting of four different subunits, with a total mass of 1.5 MDa: two heavy chains (520 kDa each) containing the ATPase activity, two intermediate chains (74 kDa each) containing the cargo binding sites, four light intermediate chains (54–59 kDa each), and several light chains. We distinguish **cytosolic** dyneins (responsible for vesicle and organelle transport) from the **axonemal** dyneins found in cilia and flagella. Limited data on dynein structure and mechanism are available; what is known has been reviewed recently [3, 11, 12]. The heavy chain is a ring of six **ATPases associated with diverse cellular activities (AAA+) domains** (see Fig. 13.7), of which the first seems to do most of the ATP hydrolysis. AAA+-domains 1–4 contain the typical WALKER-motifs (see Fig. 13.8) of nucleotide-utilising enzymes. The heavy chain also binds the intermediate chain, which in turn binds the light chains. It is assumed that changes in AAA+-domain 1 after ATP-hydrolysis are transmitted through the linker region to the other AAA+-domains and through the stalk to the rest of the dynein molecule.

Cargos of dynein carry a number of ill-characterised so-called **microtubule binding proteins (MBPs)**, which actually bind dynein, which in turn interacts with the microtubules. For a summary of motor proteins see Table 13.1.

13.3 Cilia and Flagella

13.3.1 Generic Structure of Cilia and Flagella

Many protozoans, but also **sperm cells**, are propelled by cilia (from the Latin word for eye lash) and flagella. Additionally, cilia on the **epithelial cells** of our airways and in the ventricles of our brain remove contaminants. Cilia and flagella have the same basic structure, but cilia are shorter and tend to occur in greater number on a cell than flagella.

One **nonmotile (primary) cilium** is found on nearly every vertebrate cell. Once thought to be unimportant, they are now known to carry receptors (chemical, mechanical, and thermal) for cellular environment and signalling pathways; they control differentiation, cell division, and movement.

Hair cells in the inner ear carry one kinocilium that controls the development of the bundle of sensory stereocilia (with a different ultrastructure). The kinocilium may be lost once the hair bundle has formed.

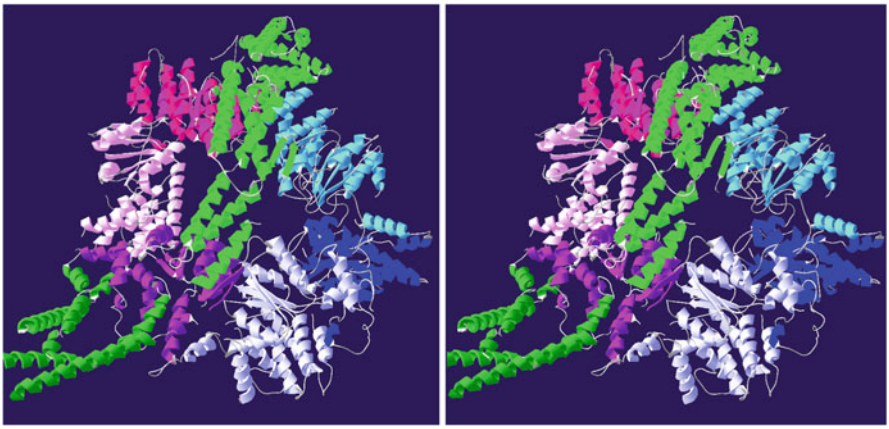


Fig. 13.7 Motor domain of the cytosolic dynein of baker's yeast (*Saccharomyces cerevisiae*, PDB-code 3qmz). The protein was crystallised as fragment (aa 1364–4092); the resolution is 6 Å. Many loops are not resolved. The six AAA+-domains (*various shades of blue and purple*), the linker (*light green*), and buttress and stalk (*dark green*) are visible

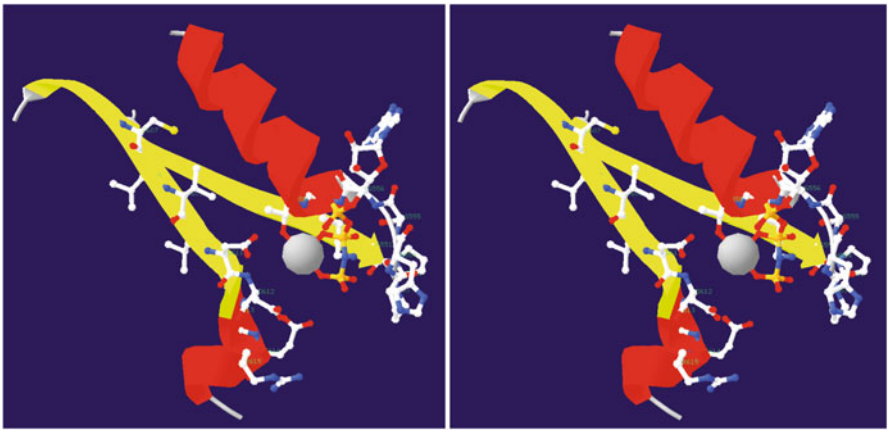


Fig. 13.8 Magnified view of NSF (PDB-code 1d2n) showing the WALKER A- and B-motives [15] of nucleotide-utilising enzymes. The WALKER A motif GxxxxGK(T/S) serves as the binding site for the polyphosphate chain; the Lys together with backbone N-atoms bind the β-phosphate. The WALKER B motif (R/K)xxxGxDxLhhhD(E/D) is much less conserved, the terminal acidic residue activates water for attack on the γ-phosphate, and the penultimate Asp coordinates the Mg^{2+}). In a subset of enzymes there is an aromatic residue located 25 amino acids upstream of the WALKER A motif that binds the adenine ring by stacking (not present in NSF)

Rods and cones in the retina have an outer segment where light is perceived. It is connected to the cell's body by the connecting cilium.

Olfactory neurons have 10 cilia per dendritic knob; these carry the G-protein coupled chemoreceptors.

Table 13.1 Summary of motor proteins

Motor protein	Substrate	Moves toward
Myosin	Actin	Plus
Kinesin	Tubulin	Plus (most)
Dynein	Tubulin	Minus

Transport of proteins between cilia and cytoplasm has been reviewed in [10], and diseases associated with their development (ciliopathies) in [6, 16].

Research Tool
Ciliary components, their structure, location, and function have been assembled into a database [14] (http://www.syscilia.org/goldstandard.shtml).

The travelling speeds of protozoans can be up to 1 mm/s, about 10–100 body lengths. Flagella from some insect sperm cells can reach a length of 2 mm.

The movement of cilia and flagella is complicated and can be studied with a high-speed video camera mounted onto a microscope. These patterns ensure maximum resistance against the medium during the forward and minimum resistance during the backward stroke.

The basic structure of most flagella and cilia is a "**9+2**" arrangement of microtubules: two central tubulin singlets are surrounded by nine peripheral tubulin doublets (see Fig. 13.9).

Inside the cell each flagellum connects to a basal body, a structure resembling a centriole. They contain 9 triplet microtubules. In a triplet microtubule we have A-, B-, and C-tubules (see Fig. 12.10). A- and B-tubules continue into the flagellum, but the C-tubule terminates in a transition zone between basal body and flagellum.

13.3.2 Mechanism of Movement

There are atypical flagella in some organisms with 3+0, 6+0, or 9+0 organisation, all of which are motile. At the very least this proves that the central pair of microtubules is not necessary for movement. Experimentally it is possible to remove the **plasma membrane** with detergent; the remaining structure is still motile, so the plasma membrane is not required for movement either. If such preparations are treated with proteases to remove the structural linkages between microtubules, bending is no longer possible, but upon addition of ATP the microtubules slide past

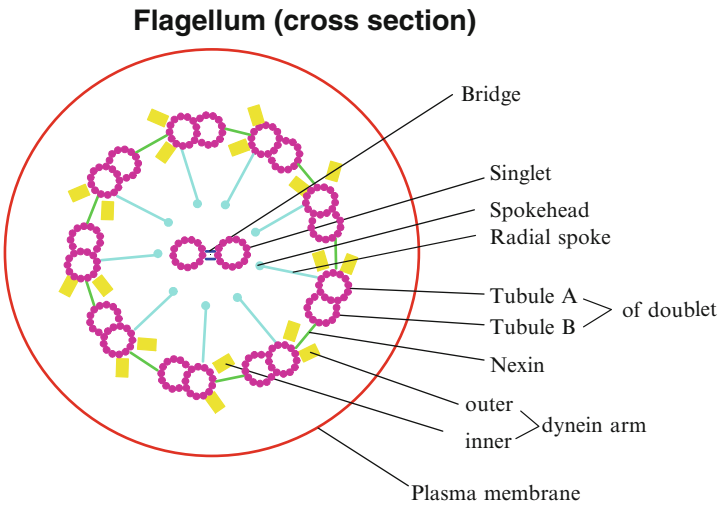


Fig. 13.9 Structure of flagella and cilia. There are two singlet and nine doublet microtubules (purple) which form the core of the structure (hence "9+2"-pattern). Each doublet is composed of an A-tubule with 13 and a B-tubule with 10 protofilaments. They are linked by nexin-bridges (*green*) and carry a radial spoke (*cyan*) and dynein arms (*yellow*). The two singlet tubules in the centre are connected by a bridge (*blue*). The entire flagellum is about 250 nm in diameter in all species, but the length can vary between 1 µm in ciliates to 2 mm in insect sperm. The plus-end of the microtubules points to the distal end of the flagellum

each other instead. Removal of **dynein** by washing with high salt prevents any movement. Dynein can be added back to such stripped preparations; this restores motility. Thus the bending of flagella is caused by the dynein molecules on one A-tubule walking toward the minus-end (proximal) of the neighbouring B-tubule, at the same time protein cross-links keep the structure intact.

No crystal structures are available for dynein, but upon negative staining dynein molecules appear under the EM as "blossoms" with two (inner dynein) or three (outer dynein in some species) heads, which are connected by smaller stalks. These heads probably represent the ATP- and tubulin binding unit, analogous to myosin.

Mutated *Chlamydomonas reinhardtii* algae lacking the **outer dyneins** are still motile, but the flagella beat slower. If the **inner dyneins** are mutated, no movement is possible [2]. Thus it is the inner dyneins that are most important for flagellum beating. Mutagenesis in *C. reinhardtii* is a particularly useful method to study flagellar proteins as regaining of motility of the dikaryon by complementation during mating is easily observed [5].

13.3.3 Cilia and Flagella Start Growing at the Basal Body

If the flagella of *C. reinhardtii* are removed (e.g., by treatment with mild acid) and then allowed to regrow in the presence of radioactive tubulin, it can be shown by auto-radiography that tubulin is only incorporated at the distal end of the growing flagellum [9, 17]. Also, no radioactive tubulin is incorporated into a flagellar microtubule when moving cells are incubated with radioactive tubulin. Thus, these microtubules are fairly stable structures.

13.4 The Mitotic Spindle

The mitotic apparatus, which is responsible for chromosome separation during cell division, is also made from microtubules (see Fig. 13.10). It consists of

- two centrosomes at opposing poles of the cell.
- a pair of "asters", tufts of astral microtubules which anchor the apparatus at the cell cortex.
- the central mitotic spindle, two symmetric bundles of microtubules shaped like a rugby ball. It consists of

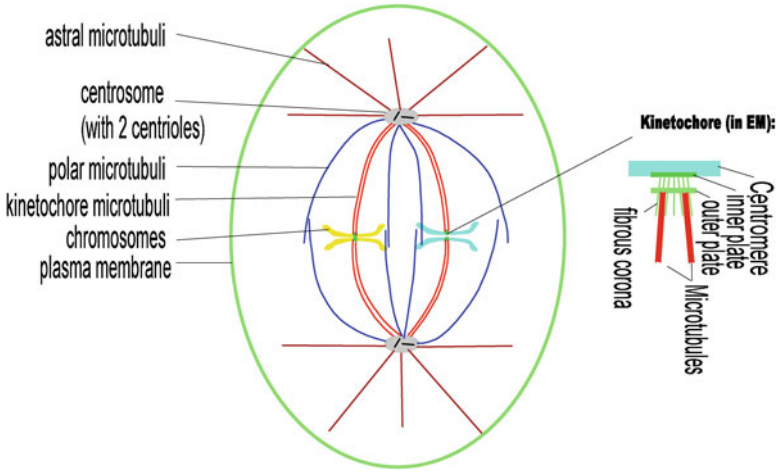


Fig. 13.10 The mitotic spindle in animals

- kinetochore microtubules, which bind to the chromosome near their centromere[1] at a specialised structure, the kinetochore. The centromere is visible under the microscope as a constriction in the chromosome; it is also the point where the two sister chromatids are held together.
- polar microtubules, which do not bind to chromosomes, but interdigitate at the equatorial plate with polar microtubules coming from the opposite centrosome.

As usual, all minus-ends of the microtubules originate from the centrosomes.

A proteomics screen has identified 795 proteins involved in the spindle apparatus [13]; this number is enormous if you consider that from the 23.000 proteins encoded by the human genome only about 2000 are expressed in a particular cell at a particular time.

Many proteins required for spindle function are stored in the nucleus and become cytosolic only once the nuclear membrane breaks down. In yeasts, on the other hand, the spindle is located inside the nucleus (spindle pole body, **closed mitosis** as opposed to the **open mitosis** of higher animals) and spindle proteins are specifically imported into the nucleus during mitosis. The switch from closed to open mitosis probably became necessary because of the greater chromosome number and mass in higher animals, which required an increase in the microtubuli number by some two orders of magnitude. Thus the spindle apparatus no longer fit inside the nucleus.

Microtubules attach to the chromosomes at specialised structures, called kinetochores (see insert in Fig. 13.10). In ultra-thin sections, plates can be seen, which are connected by a fibrous corona. This structure binds to specific DNA-sequences, which are identified by their ability to turn plasmids into artificial chromosomes. The precise structure of the kinetochore has yet to be elucidated [4].

In prophase kinesin-like motor proteins bound to the microtubules of one centrosome will move toward the plus-end of microtubules from the other centrosome. As a result, the centrosomes are pushed to opposing sides of the cell. At the same time minus-directed kinesins are responsible for parallel orientation of the microtubules. Dynein molecules in the cortex of the cell orient the astral microtubules and help to pull the centrosomes apart. Dynein also holds the microtubules in the centrosome.

Once the nuclear membrane has broken down, the plus-end of kinetochore microtubules rapidly extend and shorten, until they have found a centromere to which they can attach. If a centromere hits the side of a microtubule, it can slide toward the plus-end, using kinesin-like motors. Once the kinetochore has capped the plus-end of the microtubule, the latter is stabilised against depolymerisation.

When microtubules are attached to kinetochores, they extend or shorten to align chromosomes in the equatorial plate. The exact mechanism is still a matter of research.

[1] Do not confuse centromere, centrosome, and centriole! The centrosome (or central body) contains the centriole(s) and is a MTOC. The centromere is that part of the chromosome where the two sister chromatids are attached to each other and where the microtubules bind during mitosis.

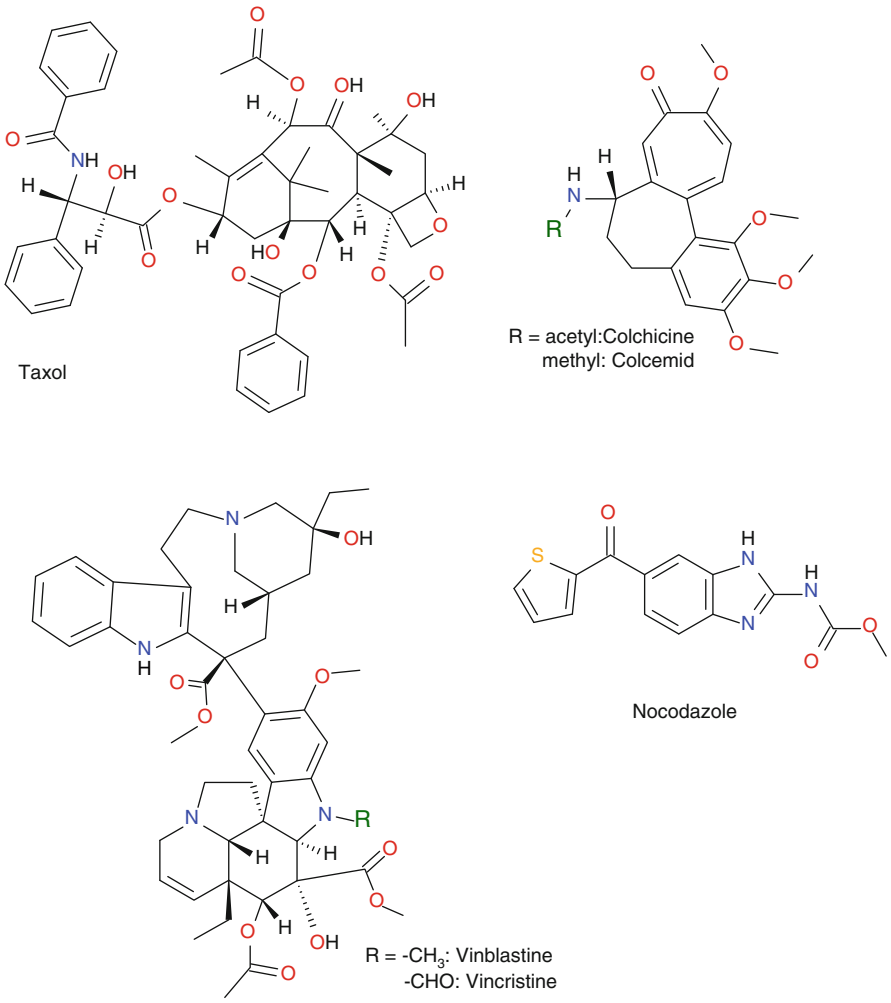


Fig. 13.11 Compounds that interfere with microtubule formation. For details see text

Once all chromosomes are bound and correctly aligned, the last mitotic check point is passed and cells enter anaphase. During **anaphase A** (or early anaphase) the kinetochore microtubules shorten by depolymerisation at the plus-end, pulling the chromosomes toward the centrosomes. This process is not energy dependent, thus it is not powered by motor proteins.

During **anaphase B** (late anaphase) the centrosomes are pushed further apart by elongation of polar microtubules and pulled toward the cell cortex by astral microtubules. Motor proteins are involved in this process as it can be inhibited by ATP-removal. Sliding forces between polar microtubules are created by plus-directed

kinesin, and pulling of astral microtubules by minus-directed dynein. Antibodies raised against these molecules can inhibit anaphase B, but not anaphase A.

Cytostatica
Because microtubules are required for mitosis, substances interfering with microtubule formation and dissociation (Fig. 13.11) are important anticancer drugs and may be beneficial in other diseases [1]. **Vinblastine** (from the Madagascan periwinkle *Catharanthus roseus* (L. 1759) G. DON 1837) binds to tubulin, prevents its polymerisation, and leads to cell cycle arrest in metaphase. **Paclitaxel**, a compound found in the bark of the Pacific yew tree (*Taxus brevifolia* NUTTALL 1849), on the other hand, leads to uncontrolled tubulin polymerisation, so that not enough monomers are available to form the mitotic spindle.

Cytogenetics
Colchicine (from the autumn crocus *Colchicium autumnale* L., an alpine flower) is used to treat acute gout, but can also arrest cells in metaphase by destruction of microtubules. In lower concentration it inhibits polymerisation

(continued)

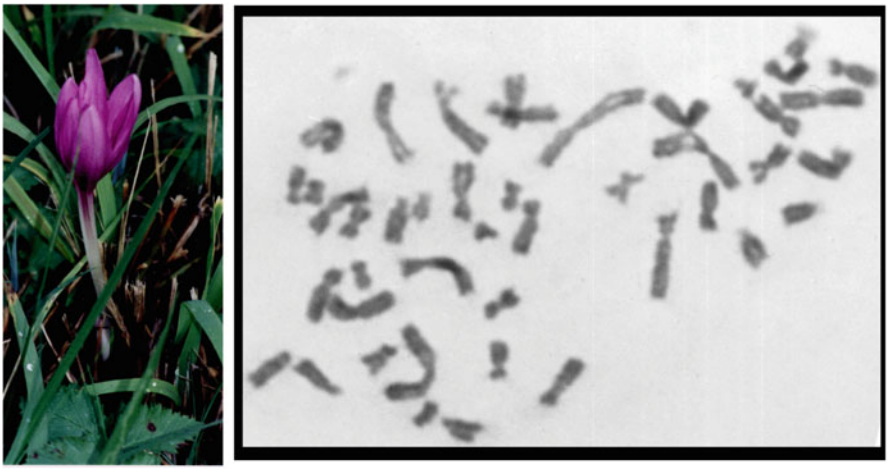


Fig. 13.12 The autumn crocus (*Colchicium autumnale* L., *top left*) produces colchicine (*bottom*), a substance that can depolymerise microtubules. This leads to cell cycle arrest in metaphase, when the chromosomes are condensed, an effect used in cytogenetics. *Right*: Metaphase spread prepared from human white blood cells, which were cultured *in vitro* after stimulation with phytohæmagglutinin. Colcemid (nitrogen methylated rather than acetylated) is also found in autumn crocus; it is less toxic than colchicine and used as cytostatic drug

at the + end; in higher concentrations it causes microtubules to detach from the organising centre, exposing their – end to depolymerisation. Metaphase-cells have their chromosomes fully condensed and aligned at the equatorial plate of the cell. Thus white blood cells or amniotic cells, stimulated into mitosis by **phytohæmagglutinin** and treated with colchicine can be used to examine chromosomes under the microscope for anomalies (missing, additional, or broken chromosomes, exchange of DNA between chromosomes). This is called **cytogenetics** and valuable for the diagnosis of some inherited diseases and some forms of cancer. An example is presented in Fig. 13.12. **Colcemid** acts in a similar way, but is considerably less toxic than colchicine. It can be used to prevent cell divisions in cancer therapy, in the same way as *Vinca*-alkaloids. **Nocodazole** acts in the same way as colchicine; it is produced by chemical synthesis rather than from plants.

Colchicine in Plant Breeding
The destruction of microtubules by colchicine is used by plant geneticists to create tetraploid cells. These cells are then grown into whole plants which form the basis for new, high-yield plant varieties.

Scleroderma
Auto-antibodies against kinetochore proteins are frequently found in patients suffering from scleroderma (thickening of the skin due to collagen overproduction, accompanied by atrophy of sebaceous glands); it can also affect inner organs (systemic scleroderma). If untreated the disease is progressive; treatment should occur in specialised centres. The pathogenesis of scleroderma is unknown, however, there are familial cases (OMIM %181750).

References

1. P.W. Baas, F.J. Ahmad, Beyond taxol: Microtubule-based treatment of disease and injury of the nervous system. Brain **136**(10), 2937–2951 (2013). doi:10.1093/brain/awt153
2. C.J. Brokaw, R. Kamiya, Bending patterns of Chlamydomonas flagella: IV. Mutants with defects in inner and outer dynein arms indicate differences in dynein arm function. Cell Motil. Cytoskeleton **8**(1), 68–75 (1987). doi:10.1002/cm.970080110
3. C. Cho, R.D. Vale, The mechanism of dynein motility: Insight from crystal structures of the motor domain. Biochim. Biophys. Acta **1823**(1), 182–191 (2012). doi:10.1016/j.bbamcr.2011.10.009
4. J.G. DeLuca, A. Musacchio, Structural organization of the kinetochore-microtubule interface. Cur. Opin. Cell Biol. **24**(1), 48–56 (2012). doi:10.1016/j.ceb.2011.11.003
5. S.K. Dutcher, The awesome power of dikaryons for studying flagella and basal bodies in *Chlamydomonas reinhardtii*. Cytoskeleton **71**(2), 79–94 (2014). doi:10.1002/cm.21157
6. A.M. Fry, M.J. Leapera, R. Baylissa, The primary cilium: Guardian of organ development and homeostasis. Organogenesis **10**(1), 62–68 (2014). doi:10.4161/org.28910
7. M. Krüger, S. Kötter, Titin – sensibler Riese der Muskelzellen. BioSpektrum **19**(4), 360–362 (2013). doi:10.1007/s12268-013-0319-8
8. F.J. Kull, S.A. Endow, Force generation by kinesin and myosin cytoskeletal motor proteins. J. Cell Sci. **126**(1), 9–19 (2013). doi:10.1242/jcs.103911
9. P.A. Lefebvre, S.A. Nordstrom, J.E. Moulder, J.L. Rosenbaum, Flagellar elongation and shortening in Chlamydomonas. IV. Effects of flagellar detachment, regeneration, and resorption on the induction of flagellar protein synthesis. J. Cell Biol. **78**(1), 8–27 (1978). doi:10.1083/jcb.78.1.8
10. J. Malickia, T. Avidor-Reissb, From the cytoplasm into the cilium: Bon voyage. Organogenesis **10**(1), 138–157 (2014). doi:10.4161/org.29055
11. K.C. Rank, I. Rayment, Functional asymmetry in kinesin and dynein dimers. Biol. Cell **105**(1), 1–13 (2013). doi:10.1111/boc.201200044
12. A.J. Roberts, T. Kon, P.J. Knight, K. Sutoh, S.A. Burgess, Functions and mechanics of dynein motor proteins. Nat. Rev. Mol. Cell. Biol. **14**(11), 713–726 (2013). doi:10.1038/nrm3667
13. G. Sauer, R. Körner, A. Hanisch, A. Ries, E.A. Nigg, H.H.W. Silljé, Proteome analysis of the human mitotic spindle. Mol. Cell. Proteomic. **4**(1), 35–43 (2005). doi:10.1074/mcp.M400158-MCP200
14. T.J.P. van Dam, G. Wheway, G.G. Slaats, M.A. Huynen, R.H. Giles, The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia **2**(7), (2013). doi:10.1186/2046-2530-2-7
15. J.E. Walker, M. Saraste, M.J. Runswick, N.J. Gay, Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**(8), 945–951 (1982). URL http://europepmc.org/articles/PMC553140
16. A.M. Waters, P.L. Beales, Ciliopathies: an expanding disease spectrum. Pediatr. Nephrol. **26**(7), 1039–1056 (2011). doi:10.1007/s00467-010-1731-7
17. G.B. Witman, The site of in vivo assembly of flagellar microtubules. Ann. N.Y. Acad. Sci. **253**(1), 178–191 (1975). doi:10.1111/j.1749-6632.1975.tb19199.x

Chapter 14
Cell–Cell Interactions

Abstract Cells interact not only with each other, but also with the extracellular matrix, which consists of proteoglycan, collagen, elastin, and multiadhesive proteins such as laminin. Both cell/cell and cell/matrix interactions require cell adhesion molecules, some of which occur in cell junctions. Disintegrins and matrix specific metalloprotease (MMP) dissolve these interactions; the latter are regulated by tissue inhibitor of metalloprotease (TIMP).

Metazoa consist of many different cell types, whose functions are closely interlinked. Only together can they perform all the functions that characterise a living organism: attaining **shape, movement, perception, metabolism, and reproduction**. We now show how all these cells are knit into a single organism. A nice summary description may also be found in [2, 4].

Before we start, first a few definitions: if several cells of the same origin come together, they form a **tissue**, of which there are four basic types: epithelial, connective, muscular, and neural. Several different tissues together can form **organs**, which are functionally and anatomically distinct parts of our body.

14.1 Extracellular Matrix

Cells secrete a network of proteins and carbohydrates, which is called **extracellular matrix (ECM)**. The mixture of ECM and associated cells is called the **stroma**. The ECM has four major components:

Proteoglycan form a viscous matrix of carbohydrate chains with many negative charges bound to a protein core. The carbohydrate binds a lot of water, which gives the ECM elasticity and the ability to absorb pressure. **Hyaluronic acid**, a polysaccharide, also serves this function.

Collagen is a fibrous protein, which provides resilience against pulling forces. Thus it acts like glass fibers in a polyester composite material.

Elastin forms elastic fibers, which allow tissue to elastically give to pulling forces, and to regain their original shape once the force stops. Arteries are a

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_14

typical example: they expand when the heart contracts and blood pressure rises; when blood pressure drops, they shrink again.

Multiadhesive proteins link collagen, proteoglycans, and cells together. The most prevalent of these proteins is laminin,[1] a heterotrimeric protein of 820 kDa.

These components are mixed in different proportions, depending on what the purpose of a tissue is: **Tendons** have a high collagen content to transmit forces between muscle and bone. **Cartilages** on the other hand have a high proteoglycan content, thus they can cushion the shocks caused in movement.

Differentiation of cells is controlled in part by their ECM environment. Cell receptors interacting with ECM components stimulate the expression of appropriate genes. In addition, the ECM can bind and slowly release hormones.

14.1.1 Collagen

Collagen is the most abundant protein in animals. There are 16 known isoforms, but almost 90 % of all collagen in our bodies belongs to the isoforms I, II, and III. These form fibrils. Type IV forms 2D-networks, for example, in the basal lamina. The remaining isoforms cross-link collagen fibrils with each other or their environment.

Fibrous collagens have a distinct structure (see Fig. 14.1). Collagen is a triple-helical coiled-coil of three protein chains, each exactly 1050 amino acids long. The core consists of repeating Gly-X-Pro/Hy-Pro motives. X can be any amino acid, except Trp. These amino acids have specific functions:

Glycine is the smallest amino acid, with just a hydrogen as side-chain. Only Gly allows the required tight packing of the collagen triple helix.

> Mutation of just a single Gly residue in the protein leads to a nonfunctional collagen, resulting in serious inherited defects such as *Osteogenesis imperfecta* (OMIM #166200, defect in bone formation) or EHLERS-DANLOS-syndrome (OMIM #130000, weak joints). These diseases are potentially fatal. Collagen I, the major collagen found in bone, tendons, skin, and interstitial tissue is formed from two α1(I) and one α2(I) chains. Mutations in α1(I) tend to be much more serious than those in α2(I). This is easy to understand: in a heterozygous individual 50 % of all collagen molecules will contain a correct copy of α2(I), if the mutation is in this protein. If the mutation is in α1(I), only 25 % of all collagen molecules produced will be functional, 50 % will have one, and the remaining 25 % two nonfunctional copies. This form of inheritance is called **dominant negative**.

[1] Do not confuse laminin in the extracellular matrix with lamin in the nucleus.

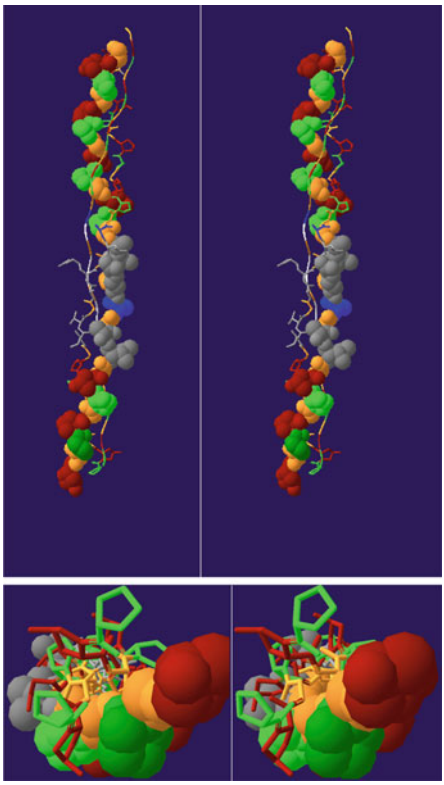


Fig. 14.1 Stereo view of the collagen triple-helix (PDB-code 1bkv). One chain is drawn as a ball-and-stick model, one shows the VAN DER WAALS radius, and the third a ribbon diagram. Gly is shown orange, Pro green, and Hyp brown. The top image shows a side view; the bottom image looks down the core of the molecule, showing that it's made of Gly residues

Proline The fixed sharp angle of the C-N bond in the ring of proline stabilises the triple helix. Note that Pro destabilises most other ordered secondary structures, such as β-strands. The tight wrapping of the chains around each other causes the high tensile strength of collagen, which is actually higher than that of steel (by weight and even by cross-section).

The triple helical segments have a 21 amino acids repeat length with hydrophobic amino acids in between; triple helix formation is assumed to start by hydrophobic bond formation, and then the tertiary structure is formed.

Collagen fibrils appear striped under the electron microscope; this pattern is caused by the gaps between adjacent triple helices (see Fig. 14.2). Dark stripes appear where many gaps come together; these are filled with uranyl acetate during preparation. Parts with few gaps exclude uranyl acetate and appear translucent. The gaps *in vivo* are places of collagen cross-linking (see Fig. 14.3), of interaction with other protein, of Ca^{2+}-binding during bone formation, and they also are primary targets for collagenase during collagen remodelling.

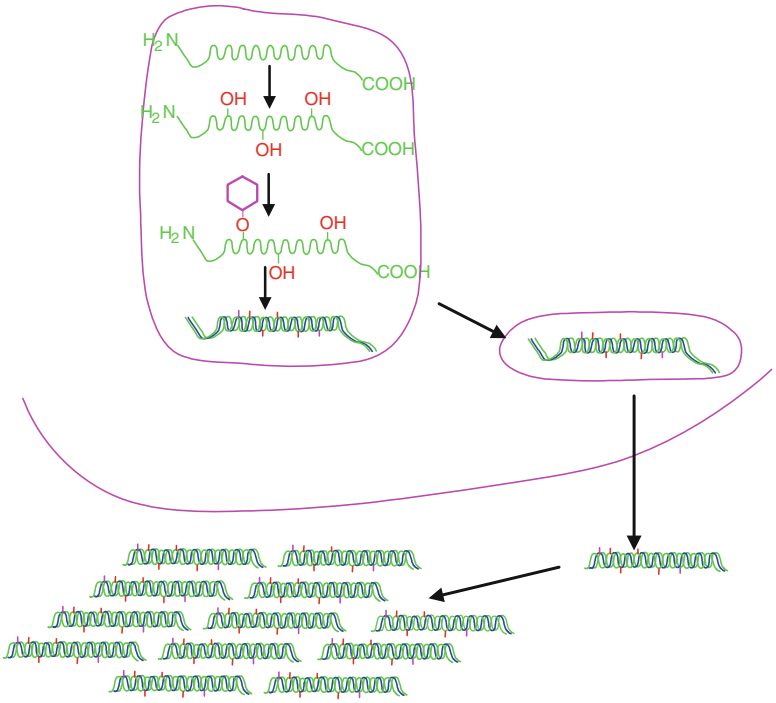


Fig. 14.2 Biosynthesis of collagen. Pre-pro-collagen is transported into the ER; some Lys and Pro residues are hydroxylated; some of the Hyl residues are glycosylated. The single collagen chains form triple helices. This is aided by disulphide bonds which form in the C-terminal pre-peptide, that is, formation of a triple helix starts at the C-terminus. For this reason collagen mutations are the more serious the closer to the C-terminus they occur. The pre-pro-collagen is clipped to pro-collagen and exported across the plasma membrane. In the extracellular space the N- and C-terminal pro-sequences are cut off, allowing the tropocollagen molecules to associate into fibrils, 67 nm thick cables. In a final step, cross-links are introduced between Lys and Hyl residues (not shown), both between the three chains of a collagen molecule and between molecules. Note the stripes in the developing collagen molecule

Markers of Bone Remodelling
Lysino-norleucine and pyridoline are released upon collagen breakdown; their concentration in urine is used as a clinical marker for bone pathology.

Correct triple helix formation (see Fig. 14.2) depends critically on the posttranslational hydroxylation of proline to hydroxyproline by a prolyl hydroxylase in the ER-membrane (see Fig. 14.4). If this reaction is not carried out, triple helix dissociation temperature is about 10 °C, *that is*, the helices are unstable at body temperature. In mammals about 50 % of Pro-residues are hydroxylated; in poikilothermic animals this figure can be lower depending on preferred environmental

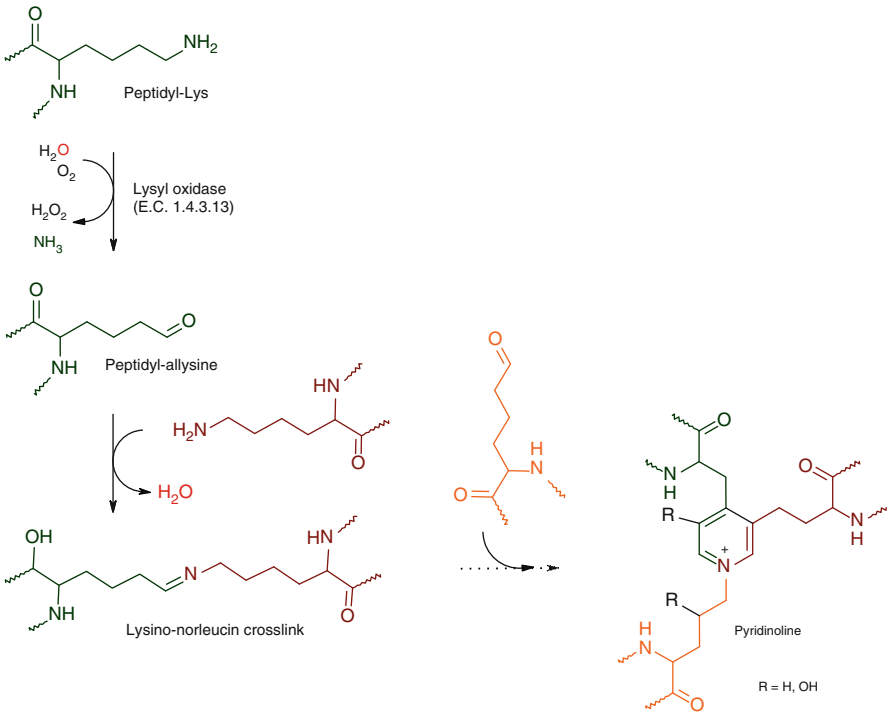


Fig. 14.3 Cross-linking of collagen. Lys oxydase (1.4.3.13) deaminates peptidyl-Lys to the corresponding aldehyde, called allysine. This spontaneously reacts with another Lys on a different chain of the same triple helix, forming an aldimine. The aldimine can slowly react with another allysine residue on a different triple helix. Depending on whether the Lys residues undergoing these reactions were hydroxylated we get **pyridoline** (R = OH) or desoxypyridoline (R = H) cross-links. No such cross-links are present at birth; they are formed slowly during life, and detectable by increased tissue autofluorescence and toughness. The wiggled lines represent the peptide bonds to the rest of the protein

temperature. Hydroxylation of Pro stabilises the *trans*-form of the amino acid and hence the tight coiling of collagen. In the lab, −OH may be replaced by −F, which is even more effective.

Scurvy
Pro-hydroxylase contains a Fe^{2+}-centre which can be oxydised to inactive Fe^{3+} by a side reaction. Hyp formation therefore requires the presence of **vitamin C (ascorbic acid)** to reduce Fe^{3+} to Fe^{2+} and regenerate the enzyme. Otherwise weakening of connective tissue by lack of properly formed collagen will lead to **scurvy**.

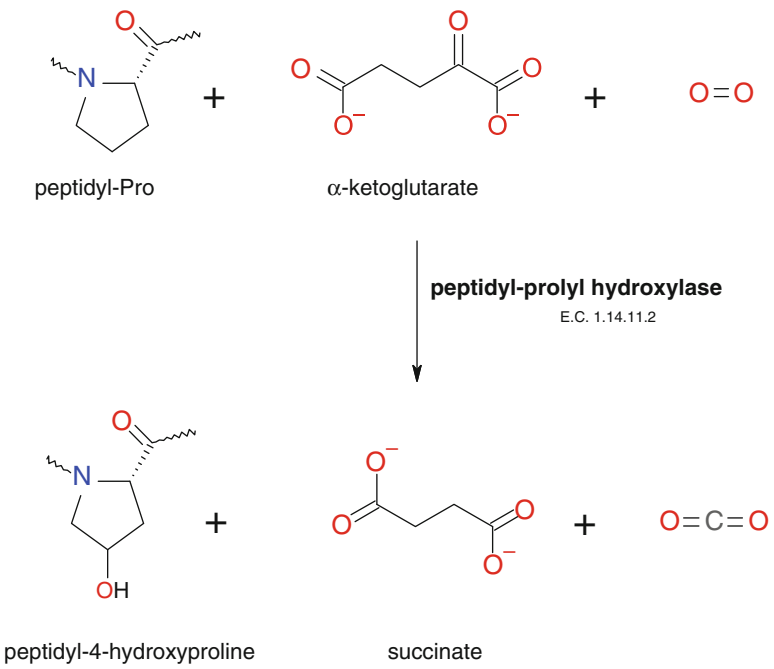


Fig. 14.4 Reaction of peptidyl-prolyl hydroxylase. The enzyme is an oxydoreductase (EC 1.) that reduces molecular oxygen (i.e., it is a dioxygenase) and uses two different electron donors (EC 1.14.). The second donor is α-ketoglutarate (EC 1.14.11.). Similar to the hydroxylation of Pro by EC 1.14.11.2 Lys may be hydroxylated by EC 1.14.11.4

Advanced Glucation End Products (AGE)
Glucation of collagen—and consequently AGE formation—in elderly and diabetic patients contributes to increased tissue stiffness [10]. It also interferes with tissue remodelling, as binding sites for tissue metalloproteases are blocked. AGE have a pro-inflammatory effect via the receptor for AGE (RAGE) [9, 13]. Polyphenols, contained in many plants, have a protective effect [12].

14.1.1.1 Nonfibrillar Collagen

Fibrillar collagens as described above are formed by five members of the collagen family (I, II, III, V, and XI). In nonfibrillar collagens the Gly-X-Pro repeats are shorter, and both the N- and C-terminus carry additional domains. They form beaded

ends on the triple helix. Two triple helices come together to form a supercoiled antiparallel dimer; two of these together form a tetrameric unit. Both dimers and tetramers are held together by many intra- and intermolecular disulphide bonds in the head regions. The tetramers form long chains with a repeat length of 100 nm.

Such nonfibrillar collagens strengthen fibrillar collagen in mechanically stressed places such as dermis or around blood vessels by binding into the gap regions. They also form **basal laminas**.

14.1.1.2 Cell-Collagen Interactions

Cells can change the morphology of the collagen they secrete by exerting pulling forces on the molecules; this leads to a tighter packing. If two explantates of embryonal fibroblasts are placed apart on a random collagen matrix, they will organise this matrix into a tight band between them. Along this band cells will move between the explantates. This is a model experiment demonstrating the origin of connective tissue during embryogenesis.

14.1.1.3 Matrix Metalloprotease

During pregnancy, development and growth collagen is degraded by MMPs; the reaction mechanism is shown in Fig. 14.5. These enzymes are also active after trauma or during infection. There are 28 different MMPs which degrade different proteins in the extracellular matrix. **Collagenases** cut collagen in the gaps; this exposes Leu-Pro bonds in the fragments that are further cleaved by **gelatinases**. **Stromelysins** break down stroma proteins other than collagen. Their activity is regulated by four **tissue inhibitors of metalloprotease (TIMPs)** (see Fig. 14.6). MMPs are secreted as zymogens and activate each other by cleavage of a propeptide.

Doxycyclin in subantibiotic dose is used as an inhibitor of MMPs in periodontal disease, corneal erosions, and metastasis of certain cancers. Other MMP-inhibitors have been designed against diseases such as arthritis; but were either too toxic or ineffective during clinical trials.

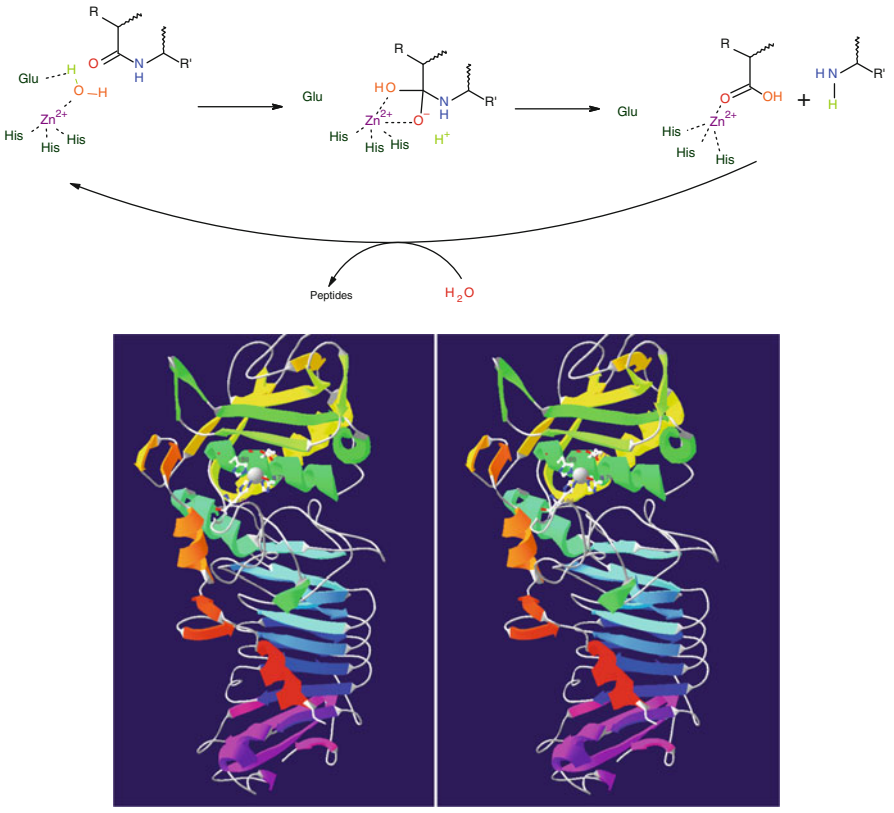


Fig. 14.5 *Top*: Reaction mechanism of Zn^{2+}-proteases. The catalytic site has the conserved sequence His-Glu-X-X-His-X-X-Gly-X-X-His. The ion is coordinated to the three His-residues of the protein; it also coordinates the O of the hydrolytic water molecule. One of the hydrogens of this water molecule is bound by the Glu carboxy-group. *Bottom*: Crystal structure of serralysin (PDB-code 4i35) as an example of a Zn^{2+}-protease. Compare this with the mechanism of Ser-proteases on page 134

14.1.1.4 Mineralisation

In bone and teeth Ca^{2+} and HPO_4^{2-}[2] precipitate to an amorphous $CaHPO_4$, which converts into crystalline hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$. The crystals may contain other ions as well, and therefore act as a storage site for minerals such as Mg^{2+} or Zn^{2+}. Replacement of OH^- by F^- makes apatite crystals mechanically harder and more resistant to dissolution. This is of particular importance in our teeth.

[2] HPO_4^{2-} makes about 81 % of total plasma P_i under physiological conditions; the rest is $H_2PO_4^-$. The activity factors α for Ca^{2+} and HPO_4^{2-} are 0.36 and 0.23, respectively, due, for example, to complex formation with other ions and binding to proteins.

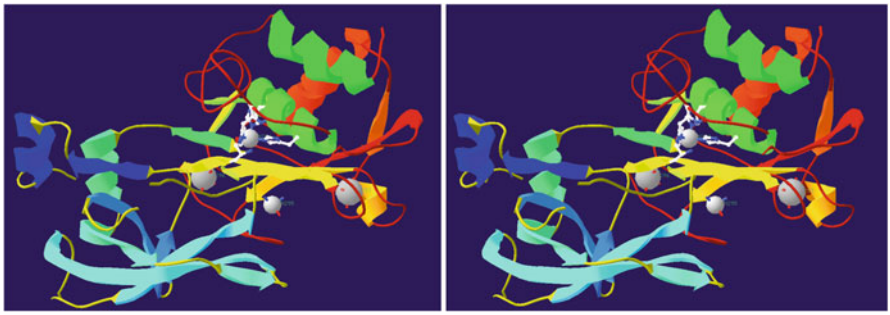


Fig. 14.6 Binding of Timp-1 (*cyan to blue*) to Mmp-14 (*red to green*), the PDB-code is 3ma2. The His-residues that coordinate the catalytic Zn^{2+}-ion are shown, as is Cys-1 of Timp-1, whose carboxy-group also coordinates with—and thereby inactivates—that zinc. The crystal was produced with a genetically engineered version of Timp-1, T98L. This mutation stabilises Timp-1 in a bound-like conformation by changing the H-bond network, decreasing the K_d to Mmp-14 by two orders of magnitude [3]. This experimental result is easily explained by the conformational selection model of binding (see Fig. 7.4 on page 170)

The solubility product K_{sp} for $CaHPO_4$ is $2.3 \times 10^{-7}\,M^2$ at 37 °C and *p*H 7.4 in physiological salt solution [8]; for apatite it's $10^{-50}\,M^2$. The product of the activities in plasma $[Ca^{2+}] \times [HPO_4^{2-}] = 1 \times 10^{-7}\,M^2$, is smaller than the former but much larger than the latter. Thus the precipitation of $CaHPO_4$ is easily reversible, but apatite is essentially insoluble under physiological concentrations. In plasma, precipitation is prevented by the presence of pyrophosphate (PP_i), which forms a soluble complex with Ca^{2+}. During bone formation local $[PP_i]$ is reduced by pyrophosphatase. Osteoblasts take up Ca^{2+} via Na/Ca-exchanger on the apical side and transport it out at the basal side by Ca^{2+}-ATPase 1b. Phosphate is taken up by Na/P_i-cotransport; it is unknown how it is secreted at the basal side. In matrix vesicles the gaps in collagen I serve as nucleation sites, especially after Ser-phosphorylation. The growing $CaHPO_4$ then ruptures the vesicle.

Noncollagen proteins in bone help with postmineralisation modelling of bone. Osteocalcin is probably the most important such protein. It contains Gla-residues that form Ca^{2+}-binding sites (see Fig. 14.7). Its synthesis is regulated by retinoic acid (hormone derived from vitamin A) and calcitriol (active form of vitamin D).

14.1.2 Elastin

Elastin is a protein of 750 amino acids. It forms elastic fibers which are about five times as extensible as a rubber tape of equal cross-section. Hydrophilic regions rich in Lys, Ala, and Pro alternate with hydrophobic regions containing VPGVG and VGGVG repeats; these form fluctuating β-structures where only the amide groups interact with water. Stretching of elastin fibres exposes hydrophobic R-groups to

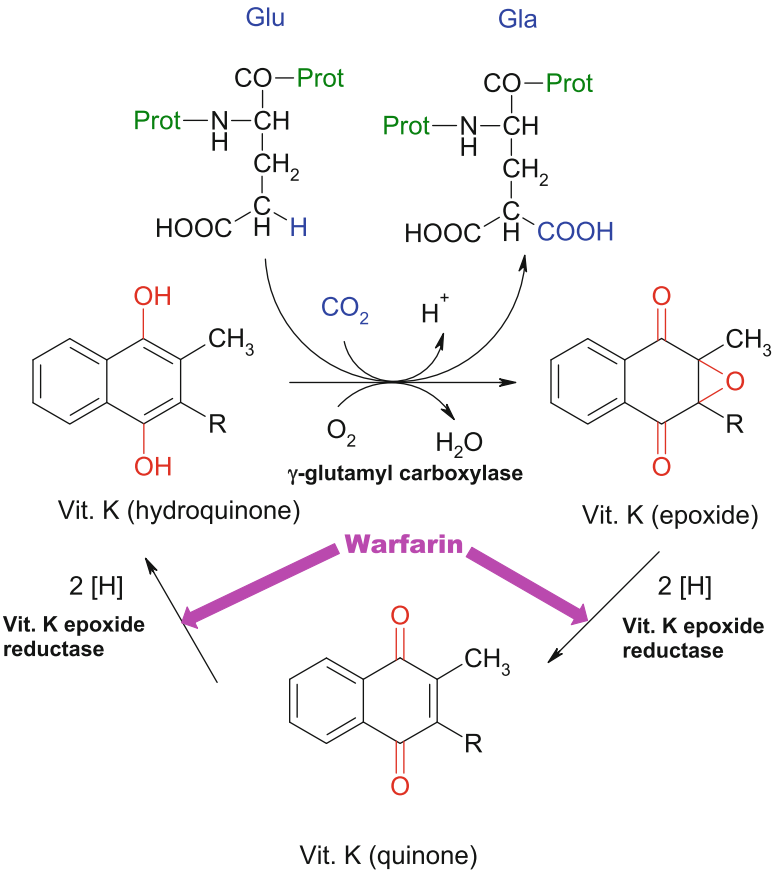


Fig. 14.7 Formation of carboxyglutamate (Gla) residues in proteins by γ-glutamyl carboxylase (EC 4.1.1.90) using vitamin K as cofactor. The two carboxy groups serve as chelation sites for Ca^{2+}. R denotes a hydrophobic chain (phytyl- in plants and isoprenyl- in animals). Both steps of the regeneration of vitamin K are performed by the same enzyme, vitamin K epoxyde reductase (EC 1.1.4.1), which is inhibited by warfarin. Warfarin analogues are used to prevent synthesis of clotting factors II, VII, IX, and X both in "blood thinners" and rat poisons

water, which is thermodynamically unfavourable. The molecule therefore will spontaneously return to its original structure. Unlike collagen, elastin is nonglycosylated, contains no Hyl, and very little Hyp. The Lys-residues in the hydrophilic regions are cross-linked (see Fig. 14.8), stabilising the protein further against pulling forces. Cross-linking and the loss of positive charges make elastin very insoluble; it has a half-life of ≈ 70 a.

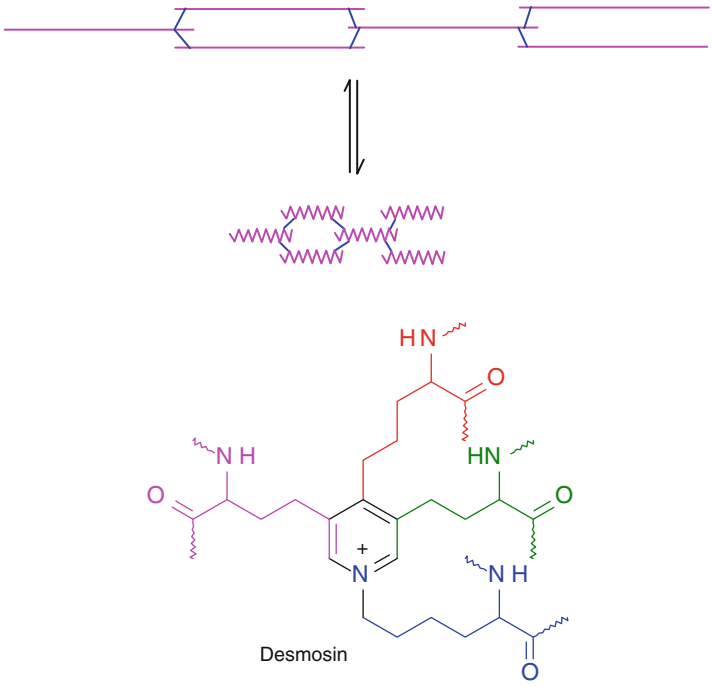


Fig. 14.8 *Top*: **Elastin** is a component of elastic fibres, which give tissue stretchability. The molecules are coiled up in the relaxed state; when stretched they become more linear. Cross-links between molecules prevent them from sliding past each other under force. *Bottom*: These cross-links are made by the unusual amino acid **desmosine**, which is posttranslationally synthesised from three allysine plus one unmodified Lys (marked in different colours) in a reaction similar to that in Fig. 14.3. In α-helices made of Lys-Ala-Ala-Lys repeats the Lys-residues are close together (see helical wheel projection in Fig. 16.9 on page 374); Lys-oxydase can form an intrachain dehydrolysinoleucin cross-link. This then reacts with a pair of allysines from a different chain, forming desmosine

During infections neutrophils release elastase, which cleaves the Val-Ala bonds between the cross-linked regions in elastin. This results in the appearance of desmosine in the urine, where it can be detected by LC/MS/MS. This can be used to monitor patients with infections in elastic tissue such as COPD [5].

Elastin is secreted into folds in the cell surface; there elastic fibers are formed by cross-linking between Lys-residues (see Fig. 14.8). The elastin molecules are normally curled; when a force is applied, the molecules straighten, gaining in length. The cross-links between molecules prevent them from sliding past each other, so they can regain their original shape when the force is released.

The elastin core in elastic fibers is surrounded by microfibriles made from glycoproteins such as **fibrillin**. During embryogenesis microfibriles are formed first, and elastin is introduced later. Apparently the microfibriles are required for correct organisation of elastin molecules. Fibrillin is also found in **elaunin** and in the **oxytalan fibres** that anchor our teeth in the jaw bone [1, 7]. Ushiki [11] has stunning images of these structures. The structure of fibrillin has not yet been elucidated; only some fragments have been crystallised. Fibrillin molecules are cross-linked to each other by transglutaminase (see Fig. 14.9).

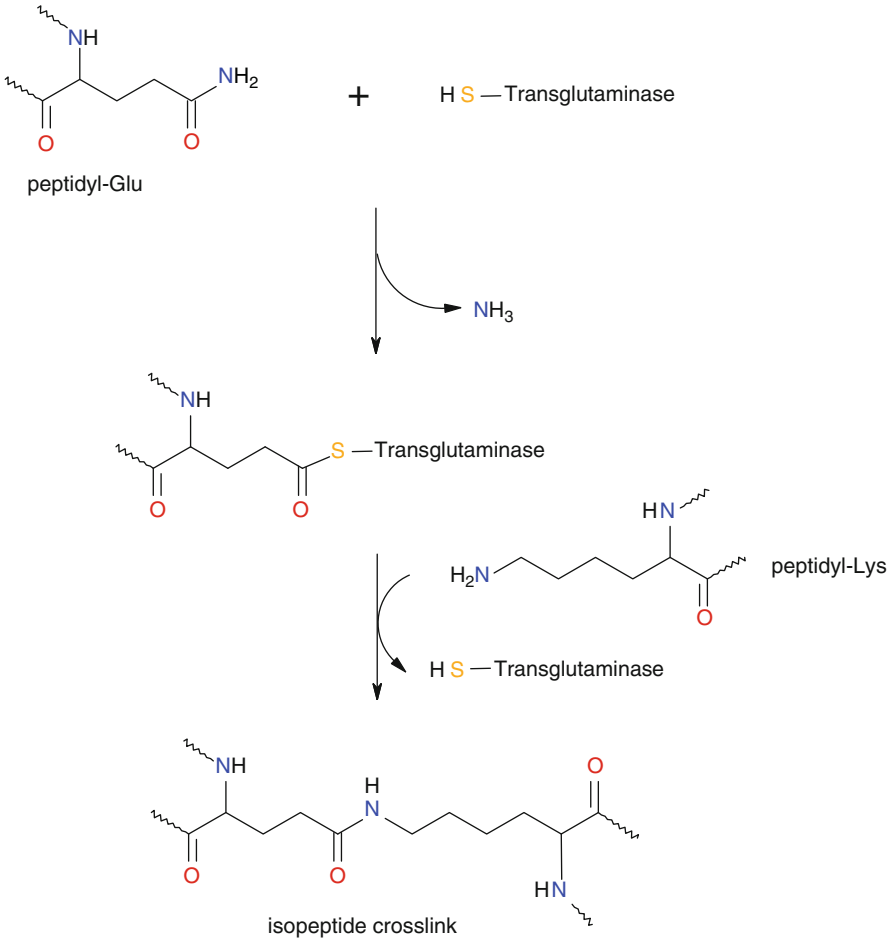


Fig. 14.9 Protein cross-linking by transglutaminase (EC 2.3.2.13). A Glu-residue from one protein is linked via an isopeptide bond to a Lys-residue from a different protein. The isopeptide bond is very resistant to proteolytic attack; it is used to cross-link proteins into a 3D network. Apart from fibrillin, we also find it in blood clotting (factor VIII). In "molecular cooking" bacterial transglutaminase is used to glue different pieces of meat together

MARFAN **Syndrome**
Mutations in fibrillin-1 are the cause of MARFAN syndrome ([6], OMIM #154700). Sufferers are unusually tall with long limbs and arachnodactyly, their joints are loose, chest and vertebral column are deformed, and two thirds require orthodontic treatment. Eyes are long, resulting in myopia. The lens may be dislocated. Mitral and aortic valves may be defective, resulting in murmur. The aorta may be weak, resulting in sudden aneurisms or dissection, especially in pregnant females. Although there is no cure for MARFAN syndrome patients may have near normal life expectancy if properly managed. Inheritance is autosomal-dominant with variable penetrance.

14.2 Cell Adhesion Molecules

In order to function in an organism, cells need to interact specifically with each other. This is the purpose of several proteins, which are collectively known as **Cell adhesion molecule (CAM)**. Some of these are found in electron-microscopically visible structures, the **cell junctions** (see Fig. 14.10).

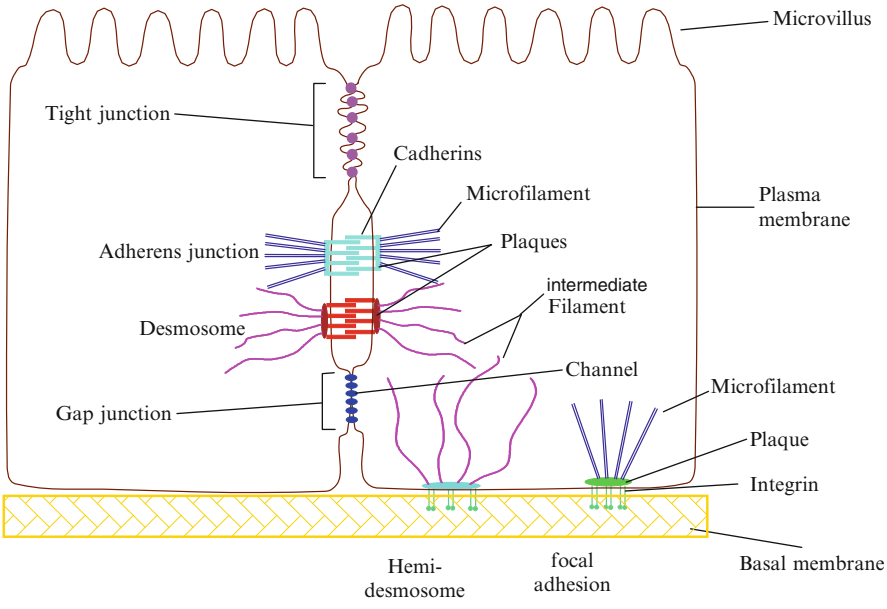


Fig. 14.10 Cell–cell and cell–matrix junctions

A number of important transmembrane proteins are involved in establishing cell–cell interactions:

Homophilic interactions between proteins of the same type:

Cadherins consist of five extracellular domains, with Ca^{2+}-binding sites between them. The N-terminal domain dimerises with a second cadherin in the same cell (Ca^{2+}-dependent); these dimers interact with cadherin-dimers on the other cell. More than 40 different types of cadherins are known; the best studied are:

E-cadherin is the most abundant cadherin and found in the preimplantation embryo and in nonneuronal epithelia. It occurs in the lateral part of the cell membrane and holds the epithelia together. In cell culture experiments the addition of antibodies against E-cadherin or (cheaper!) removal of Ca^{2+} causes cells to detach from each other.

P-cadherin is found in the trophoblast and required for nidation.

N-cadherin is found in neuronal tissue, lens, heart, and skeletal muscle.

Cadherins are type 1 membrane proteins (see Fig. 16.4 on page 367).

> Changes in cadherin expression often accompany metastasis and other pathological processes.

Cadherins prefer to interact with other cadherins of the same type.

Ig-superfamily CAMs, also called **nerve-cell adhesion molecules (N-CAMs)**, have type III fibronectin domains followed by several Ig-domains (see Fig. 11.3 on page 233). They cause Ca^{2+}**-independent** cell adhesion. They are all encoded by a single gene, but alternative splicing generates three isoforms. N-CAM-180 and -140 are type 1 transmembrane proteins, and N-CAM-120 has lost the transmembrane domain and is anchored by Glycosyl phosphatidylinositol (GPI) instead (type 6 membrane protein, see Fig. 16.4 on page 367). N-CAMs in embryos contain up to 25 % sialic acid. The many negative charges weaken N-CAM/N-CAM interactions, thus cell–cell interactions in embryos can be broken again during morphogenesis. Adult N-CAMs contain much less sialic acid. N-CAMs are required for nerve, glia, and muscle cell differentiation.

heterophilic interactions between different proteins:

Mucin-like CAMs are rich in carbohydrates and bind to **selectins**, which have a lectin (sugar-binding) domain. Selectins are important for **leucocyte/**tissue interactions. When leucocytes are to leave the bloodstream their **sialyl** LEWIS**-x-antigen** attaches to P-selectin on the blood vessel wall. The selectin is normally hidden in intracellular vesicles of the endothelial cells, but upon reception of **inflammatory mediators** they undergo rapid exocytosis.

They then trap leucocytes passing by, but because the interaction is weak, leucocytes are slowed down rather than stopped. Thus leucocytes "roll" along the endothelium.

Integrins are heterodimers with an α- (17 known isoforms) and a β-subunit (8 known isoforms). They bind mostly to extracellular matrix proteins including fibronectin or laminin. Integrin expression is also stimulated in leucocytes when **Platelet activating factor (PAF)**, a paracrin hormone from blood vessel endothelial cells, is received. Integrin then binds to I-CAM-I and I-CAM-II on the endothelium. This interaction is much tighter than that afforded by selectin; the leucocytes spread out on the endothelium and start moving into the tissue.

Deficiency in β-integrin expression prevents leucocytes from leaving the blood vessels, resulting in an inborn error, **leucocyte adhesion deficiency** (OMIM #116920).

14.3 Junctions

Tight junctions connect epithelial cells and prevent passage of fluid from one side of the epithelium to the other. For example, they separate the intestinal content from blood.

Gap-junction are distributed on the lateral surfaces of cells; they allow the exchange of small molecules (nucleotides, second messengers) between neighbouring cells. Thus they help to integrate the metabolism of the cells in a tissue. In EM the gap junctions look like close juxtaposition of cell membranes, with the narrow gap between them filled by particles. These "particles" are actually water-filled channels, linking the cytosols. The pore size of these channels is about 12 Å; this is sufficient to allow free passage of molecules up to 1200 Da, and to completely prevent passage of molecules greater than 2000 Da.

High cytosolic Ca^{2+} concentrations cause these channels to close. Remember that in a living cell [Ca^{2+}] is very low (below 0.1 μ M), whilst extracellular [Ca^{2+}] is much higher (1–2 mM). If a cell ruptures, cytosolic [Ca^{2+}] will increase and shut down the gap junction channels. Thus neighbouring cells are protected from the leak.

Cell–cell junctions have a structural role; they hold cells together by forming a bridge between their cytoskeletons. In cell–cell junctions the plasma membranes of the two cells run parallel with a distance of only 15–20 nm. The close spacing is achieved by cadherins; adaptor proteins link them to the cell skeleton. There are two types of cell–cell junction:

Adherens junctions Cadherin E is located in a continuous band just below the tight junctions. These molecules are cross-linked to actin filaments by

α- and β-catenins. This horizontal band of junctions links all cells within a layer of tissue and distributes forces between them.

Desmosomes Cadherins **desmocollin** and **desmoglein** are cross-linked to **keratin** filaments by adaptor proteins such as **placoglobin**. These form a cytoplasmic plaque which is visible in EM. Because the keratin IF reaches through the cells, desmosomes can distribute forces between different layers of a tissue.

Pemphigus vulgaris is an autoimmune disease caused by autoantibodies against desmoglein, one of the cadherins in desmosomes. This leads to disruption of desmosomes and thus to blistering of skin and mucus membranes.

Cell–matrix junctions Like cell–cell junctions, cell–matrix junctions maintain the integrity of a tissue. The most important cell membrane proteins involved are **integrins**. The combination of α- and β-isoforms determines the binding specificity of integrins. Most integrin isoforms are expressed in several cell types and most cell types express several integrins. However, some isoforms are specific to certain cell types; the α_L,β_2-type in **leucocytes**, which makes contact with blood vessel endothelial cells (rather than extracellular matrix) has already been discussed.

Another important example is the α_{IIb},β_3-type found in **platelets**. Normally it is inactive, but after stimulation of platelets it binds fibrinogen, fibronectin, VON WILLEBRAND-factor and other blood clot proteins. We do not yet know how this is regulated.

α_4,β_1-integrin keeps hæmatopoietic cells in the **bone marrow**, because it binds to fibronectin, which is secreted into the extracellular matrix by stromal cells. It also binds to the stromal cells directly, via V-CAM-I. At a late stage of their differentiation hæmatopoietic cells reduce integrin expression, so they may enter the blood stream.

Integrin/cell matrix interaction typically have low affinities (μ M), but because of the large number of interactions involved, cells bind firmly into their matrix. The weak individual interactions, however, are important for moving cells. Affinity appears to be subject to regulation by the cell, the mechanism is unclear. We distinguish two types of cell–matrix interactions:

focal adhesions are used by moving cells to temporarily attach themselves to the matrix. They consist of a cluster of integrins, which are linked to actin stress fibers by a focal plaque.

hemidesmosomes are found on the basal surface of epithelial cells, they attach the cells to the basal lamina. They consist of α_6,β_4-integrin which links extracellular laminin to keratin IF inside the cell.

14.3.1 Disintegrins

In some cases normally immobile cells need to move, for example, keratinocytes in wound healing. This is achieved by peptides, which compete with ECM proteins for binding to cellular integrins. These peptides are called **disintegrins**. Disintegrins may be secreted at the same time as matrix metalloproteases (MMPs), which dissolve either the extracellular matrix and/or the integrins on the cell's surface.

Some snake toxins also contain disintegrins to prevent blood clotting.

14.4 Exercises

14.4.1 Problems

14.1. Posttranslational modifications of which of the following amino acids are important for collagen stabilisation:

A Gly, Ala
B Tyr, Ser, Thr
C Lys, Pro
D Glu, Asp
E His, Arg

14.2. Explain the clinical symptoms of scurvy from the molecular action of vitamin C.

14.3. In a Dutch family whose parents were fourth cousins, three out of five siblings were found to have osteoarthritis-like degenerative joint disease: prominent interphalangeal joints, midface hyperplasia, a short, upturned nose, prominent supraorbital ridges, and protruding eyes. All three were deaf from birth.

If you sequenced their COL11A2-gene, what type of mutation are you likely to find?

A A Gly→Arg missense mutation in the triple-helical region
B A Trp→Arg missense mutation in the triple-helical region
C A nonsense mutation in an exon
D A nonsense mutation in an intron
E Loss of a TATA box

14.4.2 Solutions

14.1 Hydroxylation of Lys and Pro are important for the stability of collagen. Gly is an important component of collagen, but not modified.

14.2 Vitamin C is required to regenerate the enzyme prolyl-hydroxylase, which can become inactivated by oxydation of its catalytic iron centre to Fe^{3+}. If this is not corrected, prolyl groups of pre-pro-collagen will not become hydroxylated in the ER, and the resulting collagen triple helix will be unstable at physiological temperature. Thus the connective tissue will be weak, leading to the typical symptoms of scurvy such as petechia and loss of teeth.

14.3 The triple helix is formed by Gly-X-Pro repeats; only the small Gly in the first position allows the tight packing of the protein chains around each other, which is responsible for the high tensile strength of collagen (stronger than steel!). Any other amino acid in this place would weaken the triple helix. The Pro (which may be hydroxylated posttranslationally) causes a sharp angle so that the chains wrap around each other.

References

1. M.I. Bonetti, Microfibrils: a cornerstone of extracellular matrix and a key to understand Marfan syndrome. Ital. J. Anat. Embryol. **114**(4), 201–224 (2009)
2. P.M. Gallop, O.O. Blumenfeld, S. Seifter, Structure and metabolism of connective tissue proteins. Ann. Rev. Biochem. **41**, 617–672 (1972). doi:~10.1146/annurev.bi.41.070172.003153
3. M. Grossman, D. Tworowski, O. Dym, M.-H. Lee, Y. Levy, G. Murphy, I. Sagi, The intrinsic protein flexibility of endogenous protease inhibitor TIMP-1 controls its binding interface and affects its function. Biochemistry **49**(29), 6184–6192 (2010). doi:~10.1021/bi902141x
4. M. Levine, *Topics in Dental Biochemistry* (Springer, Berlin, Heidelberg, 2011). ISBN 978-3-540-88115-5
5. S. Ma, G.M. Turino, Y.Y. Lin, Quantitation of desmosine and isodesmosine in urine, plasma, and sputum by LC-MS/MS as biomarkers for elastin degradation. J. Chromatogr. B **879**(21), 1893–1898 (2011). doi:~10.1016/j.jchromb.2011.05.011
6. A. Marfan, Un cas de déformation congénitale des quartre membres, plus prononcée aux extrémitiés, caractérisée par l'allongement des os avec un certain degré d'amincissement. Bulletins et memoires de la Société medicale des hôspitaux de Paris **13**(3), 220–226 (1896)
7. G.S. Montes, Structural biology of the fibres of the collagenous and elastic systems. Cell Biol. Int. **20**(1), 15–27 (1996). doi:~10.1006/cbir.1996.0004
8. W.C. O'Neill, The fallacy of the calcium-phosphorus product. Kidney Int. **72**(7), 792–796 (2007). doi:~10.1038/sj.ki.5002412
9. R. Ramasamy, S. Yan, A. Schmidt, Advanced glycation endproducts: from precursors to rage: round and round we go. Amino Acids **42**(4), 1151–1161 (2012). doi:~10.1007/s00726-010-0773-2
10. J.G. Snedeker, A. Gautieri, The role of collagen crosslinks in ageing and diabetes - the good, the bad, and the ugly. Muscles Ligaments Tendons J. **4**(3), 303–308 (2014). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4241420/pdf/303-308.pdf

11. T. Ushiki, Collagen fibers, reticular fibers and elastic fibers. A comprehensive understanding from a morphological viewpoint. Arch. Histol. Cytol. **65**(2), 109–126 (2002). doi:~10.1679/aohc.65.109
12. D. Vauzour, A. Rodriguez-Mateos, G. Corona, M.J. Oruna-Concha, J.P.E. Spencer, Polyphenols and human health: Prevention of disease and mechanisms of action. Nutrients **2**(11), 1106–1131 (2010). doi:~10.3390/nu2111106
13. S. Willemsen, J.W.L. Hartog, M.R. Heiner-Fokkema, D.J. van Veldhuisen, A.A. Voors, Advanced glycation end-products, a pathophysiological pathway in the cardiorenal syndrome. Heart Fail. Rev. **17**(2), 221–228 (2012). doi:~10.1007/s10741-010-9225-z

Chapter 15
Aiding in Folding: Molecular Chaperones and Chaperonins

Abstract Molecular chaperones bind reversibly to unfolded and misfolded proteins, thereby preventing their aggregation. Molecular chaperonins can actively unfold misfolded proteins, using metabolic energy. Neither chaperones nor chaperonins actively fold proteins. The 70 and 90 kDa heat shock proteins and the crystallins are examples for chaperones, the foldosome (GroEL/GroES in bacteria, TRiC in eukaryotes) and the 100 kDa heat shock proteins are the most important examples for chaperonins.

C. ANFINSEN could show that denatured (boiled or urea-treated) RNase could resume its correct folding once the temperature or urea concentration had been lowered [1]. He postulated that all proteins contain in their primary structure (amino acid sequence) the complete information which determines their secondary and tertiary structure (folding in 3D space). This hypothesis has proven fruitful (NOBEL Price 1972).

The correct conformation of a protein is usually the one with the **lowest free energy**, and given enough time, all proteins will eventually assume their correct fold. However, many proteins would require a very long time, because they can become trapped in misfolded states. These states have a free energy which is higher than that of the native state, but lower than that of all neighbouring conformations. Unfolding such misfolded proteins therefore requires a high activation energy and hence is a slow process.

Additionally, inside a cell proteins are packed to very high densities (cytosolic [protein] $\approx$ 400 mg/mL); this increases the risk for inappropriate interactions between proteins, which would result in protein aggregation and precipitation. **Aggregation** is distinguished from **oligomer-formation** during protein assembly by the irregular, nonspecific, and disordered nature of protein–protein interactions; aggregates are devoid of biological function.

Protein aggregation is involved in many diseases (amyloidoses, see Sect. 10.2 on page 206).

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_15

Looking at the Cytosol
In *E. coli* cytosol the protein concentration is 300–400 mg/mL. Assuming an average molecular mass for proteins of 50 kDa, this is a molar concentration of 6–8 µM. Because one mol contains $N_A = 6.022 \times 10^{23}$ molecules and 1000 L is equivalent to 1 m^3, this is equivalent to $361–482 \times 10^{21}$ molecules/m^3, and each molecule has $2.1–2.8 \times 10^{-24}$ m^3 space available. This is a cube of 12.8–14.1 nm. The unit cell volume of proteins is 1.68–3.53 $Å^3$/Da [17], and 1 $Å^3$ is equivalent to 1×10^{-30} m^3. Thus a protein of 50 kDa would have a volume of $84–177 \times 10^{-27}$ m^3, equivalent to a radius of 2.72–3.48 nm. In other words, cytosol can be viewed as an assembly of protein spheres of 5.5–7.0 nm diameter with on average about 5.8–8.7 nm space between them. This space is filled with water, minerals, and metabolites.

Nature has solved these problems by inventing two classes of proteins, called **molecular chaperones** and **molecular chaperonins** [10, 25].

Chaperones, "like their human counterparts, prevent unfavourable interactions between proteins during critical periods of their existence" (U. HARTL). Thus they prevent misfolding and aggregation, but they cannot unfold misfolded proteins.

Chaperonins, on the other hand, will actively unfold misfolded proteins, using energy from ATP-hydrolysis to break offending bonds.

It is important to remember that neither chaperones nor chaperonins can actively fold proteins; folding is spontaneous according to the ANFINSEN-hypothesis (but see [18] for a different view). Both chaperonins and many chaperones require energy in the form of ATP to function.

Chaperones and chaperonins may be constitutively expressed (always present) or stress-regulated (so called **heat shock proteins**, even though cold, starvation, poisoning, and other forms of stress can also activate their synthesis). An interesting way to regulate the expression of some heat shock proteins has recently been described in *E. coli* and *Bradyrhizobium japonicum* [7]: the mRNA at low temperatures forms a secondary structure which masks its ribosome binding site (start codon and SHINE-DALGARNO-sequence). At elevated temperatures the secondary structure melts and translation becomes possible.

The same mechanism is also used to regulate the expression of virulence factors in certain human-pathogenic bacteria including *Yersinia pestis* and *Listeria monocytogenes*. A change of environmental temperature to the relatively high 37 °C indicates successful infection of the host and is followed by the synthesis of proteins required for pathogen-host interaction. Below 30 °C these proteins are not expressed.

Kinetic Parameters of Antiaggregation Activity
KURGANOV [13] describes how one can assess the antiaggregation activity of both proteins and small molecules (osmolytes, cyclodextrins) quantitatively. Aggregates are larger than the proteins they are made of; this increases light scattering (see [5, chapter 9] for details). If one does that in an aggregation experiment, one gets an S-shaped curve of scattering versus time. The foot of this curve is described by the equation

$$I_t = I_0 + \nu(t - t_0) \tag{15.1}$$

with I_t the signal at time t, I_0 the initial signal, t_0 the lag time, and ν the initial rate for aggregation. Both the rate and the lag time can be changed by chaperones. The order of the aggregation reaction n with respect to protein may be determined by plotting $\log(\nu)$ as a function of $\log([P]_0)$, the initial protein concentration because

$$\nu = \text{const} \times [P]_0^n \tag{15.2}$$

If aggregation is much faster than protein unfolding, n will be close to 1 as unfolding of course is a reaction of first order. $n > 1$ indicates that aggregation is slower than unfolding.

If substrate protein forms a complex with the chaperone, $[P]_0$ decreases. Then with $x =$ [chaperone]/[P] we get

$$\left(\frac{\nu}{\nu_0}\right)^{\frac{1}{n}} = 1 - \frac{x}{S_0} \tag{15.3}$$

The S_0, the intercept on the abscissa, is the apparent stoichiometry of the chaperone-substrate complex. Curved plots are obtained if the binding stoichiometry depends on x. For the discussion of special cases please refer to the original publication.

15.1 Hsp70 Is an Example for Molecular Chaperones

In the eukaryotes four isoforms of Hsp70 are found (see Fig. 15.1 for the structure and Fig. 15.2 for an evolutionary tree of these proteins. Fig. 15.3 shows the high homology between these proteins.):

Hsp70 proper which is expressed in the cytosol, but only in stress situations.
Hsc70 (70 kDa heat shock cognate protein) is also cytosolic, but constitutively (always) expressed.

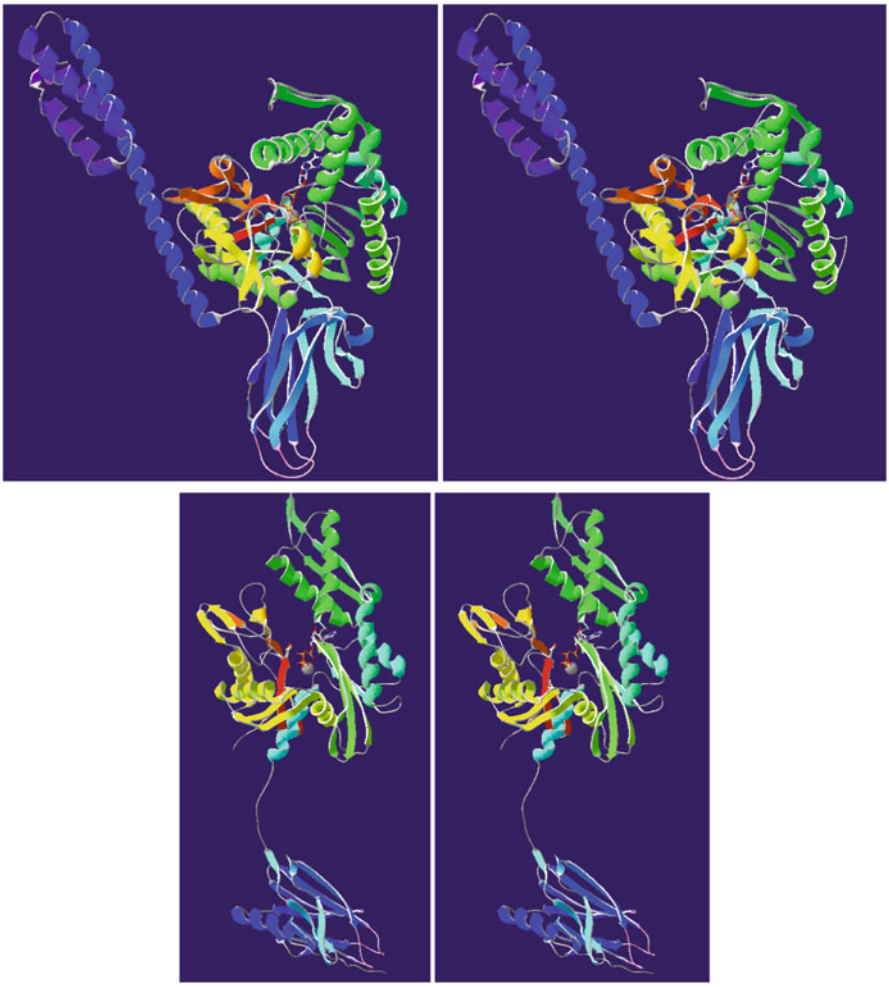


Fig. 15.1 *Top*: Stereo-image of Hsc70 with bound ATP (PDB-code 4b9q). The ATP-binding site is at the top right between the two green helices. The purple helices to the left form the lid domain, which is in the open position. Substrate binding occurs at the pink coils bottom right, at the foot of the β-sandwich core (*cyan*). *Bottom*: The ADP-Pi-form of DnaK (2v7y), shown as much as possible in the same orientation and with the same colour scheme as the ATP-bound molecule above. Note the extensive conformational changes

Grp78^{BiP} (78 kDa glucose regulated protein/immunoglobulin heavy chain binding protein) is found in the ER (see Sect. 16.4 on page 377) and has similar functions as Hsc70. It is always expressed, but its concentration rises in stress situations.

DnaK is the prokaryotic equivalent of Hsc70, found also in plastids and mitochondria.

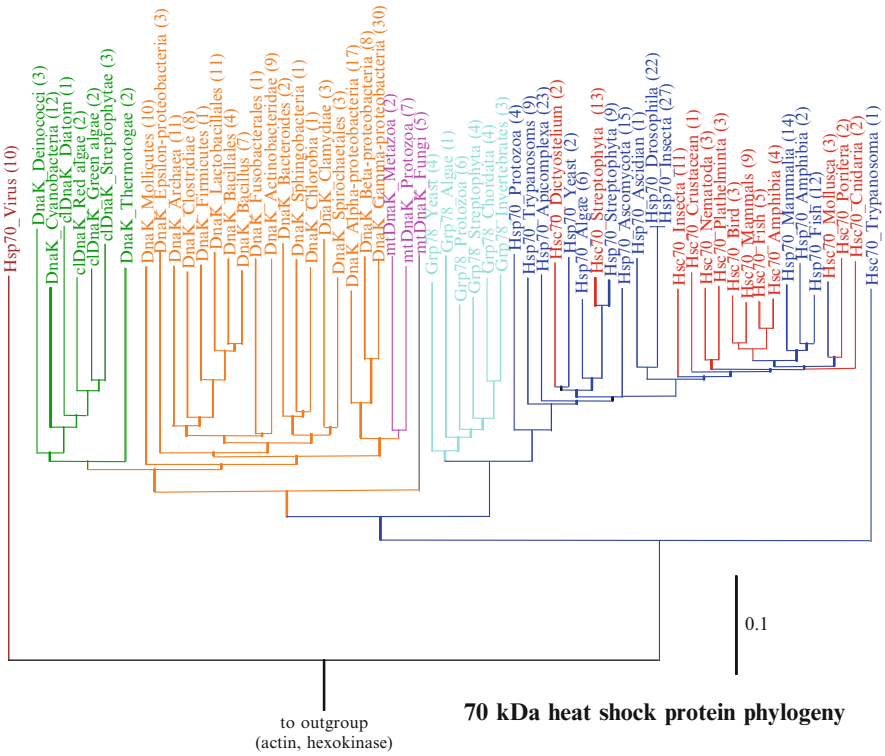


Fig. 15.2 Evolutionary tree of the 70 kDa heat shock proteins (Hsp70). The tree has two major branches, a eukaryotic and a prokaryotic one. Prokaryotic Hsp70s are often called **DnaK**. Each of these major branches has several recognisable subbranches. Mitochondrial and plastid versions of DnaK (*magenta and green*, respectively) can be distinguished from their bacterial homologues (*orange*). The eukaryotic subtree contains not only proteins expressed in stress situations (**Hsp70** proper, *blue*), but also a cytosolic (*red*, Hsc70, 70 kDa heat shock cognate) and an ER-version (*cyan*, **Grp78^{BiP}**, 78 kDa glucose regulated protein/Ig heavy chain binding protein) which are constitutively expressed. Certain plant virus have their own copy of Hsp70 (*brown*). The tree is rooted by inclusion of homologue proteins (actin, hexokinase) as outgroup. The number of proteins included in each group is given in brackets. Sequences were obtained from ExPASy (`http://au.expasy.org/`, clustering was performed with ClustalX (`http://www-igbmc.u-strasbg.fr/BioInfo/`)

Of these, Hsc70 is responsible for:

- Binding to newly synthesised proteins as they come from the ribosomes [16] to prevent them from misfolding (see Fig. 16.8 on page 373). About 15–20 % of all proteins produced by the cell transiently bind Hsc70; in pulse chase experiments it has been demonstrated that most protein/Hsc70 complexes dissociate within 10 min and essentially all within 30 min [10]. It has been speculated that mRNA stretches with rare codons slow down translation to align it with the folding rate [26].

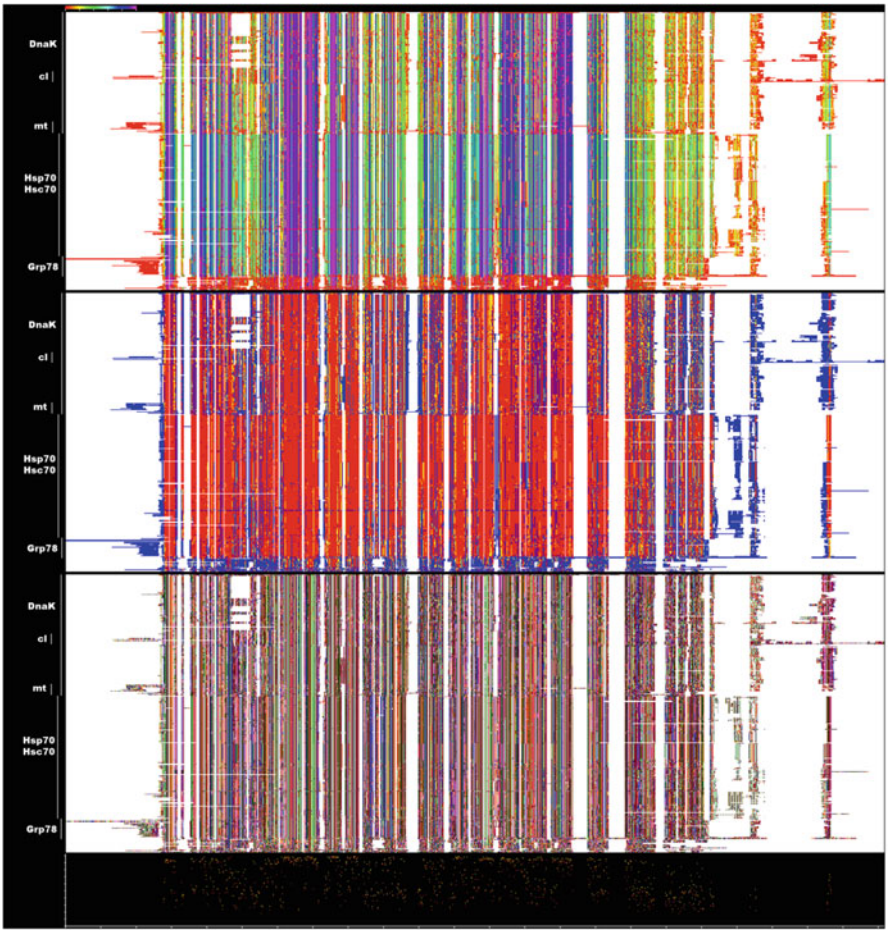


Fig. 15.3 Homology plot of the 70 kDa heat shock proteins. Sequences were aligned into a matrix as described for Fig. 15.2. Each row of the plot describes a protein sequence, and each column a position of the consensus sequence. The colour depends on the amino acid present. In the top diagram pixels are coded by the frequency of the amino acid in that position (*red* rare, *purple* frequent), in the second by identity with the consensus sequence (*red*: identity, *yellow*: conservative substitution, *blue*: nonconservative substitution), and in the third by Shapely colours. In the bottom diagram the identity and similarity are plotted. White denotes gaps in all diagrams. Regions of high conservation are mainly found in the N-terminal ATPase domain, the protein-binding C-terminal domain is less conserved and presumably determines substrate specificity. Various clusters of protein sequences can be identified, in particular, the separation between prokaryotic and eukaryotic sequences is apparent

- Binding to actin and tubulin during their release from the minus end of microfilaments and microtubules and their delivery to the growing (plus) ends (see Fig. 12.2 on page 289).

- Removal of clathrin from coated vesicles. This allows them to fuse with endosomes during endocytosis ([4], see Sect. 17.2.1 on page 396).
- Import of nascent proteins into organelles. Hsc70 and its prokaryotic counterpart, DnaK, work together as "molecular ratchet" (see page 371).
- Cross-linking of glutamate decarboxylase (which makes the neurotransmitter GABA) to synaptic vesicles. This allows efficient loading of the vesicle with neurotransmitter.
- Recognition of terminally misfolded proteins. One of the cochaperones, Chip, is a ubiquitin ligase which prepares such damaged proteins for degradation in the proteasome. Hsp70 and Hsc70 are also involved in the binding of ubiquitinated proteins to the proteasome, with the cochaperone Bag-1 (see pp. 258 and [8]).
- Shuttling of peptides from the proteasome to the ER (together with other chaperones), where they are taken up by the Tap1/Tap2 transporter and bind to the MHC-I for display at the cell surface (see Fig. 11.19 on page 260). Proteins bound to bacterial or tumor-derived Hsc70 and released into the interstitial fluid stimulate the immune system, however, the exact role of Hsc70 (adjuvant vs. immune modulator) is still a matter of debate [22].
- Cooperation with other heat shock proteins as required for their function (see, e.g., Fig. 15.5) [11]. Thus Hsc70 forms a central part of the folding machine in our cells; accepting proteins, for example, from small heat shock proteins and delivering them to the GroEL/GroES chaperonin.
- Maintaining the apo-proteins of [Fe-S]-dependent cytosolic enzymes in a folding competent state, until the [Fe-S]-cluster (synthesised in the mitochondria) has been bound.

The reaction mechanism of Hsc70 is probably the same in all these diverse functions. When Hsc70 binds ATP, its protein binding site opens (see Fig. 15.1). Any proteins bound are released, and new proteins can bind freely. Once the ATP is hydrolysed, however, the protein binding site closes; any bound protein is trapped until the ADP has been exchanged for fresh ATP.

Exchange of ADP for ATP and the hydrolysis of ATP is probably controlled by **protein cofactors**. In the bacterial homologue of Hsc70, DnaK, these cofactors have been identified: **DnaJ** binding to Hsc70 stimulates the hydrolysis of ATP, **GrpE**, on the other hand, stimulates the exchange of ADP for ATP.

Eukaryotic versions of these proteins have been isolated. DnaJ-like proteins (e.g., Hsp40 or auxillin) have in common the so-called J-domain, a motive of 4 α-helices with the conserved sequence HPD in the loop between helix 2 and 3. The J-domain is often, but not always followed by a Gly/Phe-rich region and 4 Cys-repeats that together form 2 Zn-binding sites. This conserved stretch is followed by a nonconserved C-terminus which is responsible for substrate protein and Hsp70 recognition. ADP/ATP exchange factors (such as Bag1 and HspBP1) have also been identified in eukaryotes; in addition, proteins that seem to stabilise the Hsp70/ADP/substrate complex such as Hip [15]. However, we do not yet understand the regulation of Hsp70 function in eukaryotes.

Note that Hsc70 is involved in many functions that involve **directed transport** of proteins (clathrin from the uncoated vesicle to the growing pit, peptides from the proteasome to the ER, actin from the minus to the plus end of microfilaments, and so on). If DnaJ-like cofactors were located at the source, and GrpE-like proteins at the destination, this vectorial action of Hsc70 could be explained.

Apart from the constitutively expressed Hsc70, cells under stress (heat, cold, lack of food, presence of certain poisons) can express the stress-regulated **70 kDa heat shock protein (Hsp70)**, to help maintain protein structure in the cell and to clear away any debris. For example, Hsp70 binds to poly-ubiquitinated proteins (with some E_3-ubiquitin ligases acting as ATPase-stimulating factors) and delivers them to the 26 S proteasome (see page 258ff for a description). There the Bag1 protein allows recognition of the Hsp70/protein complex by the proteasome and acts as ATP/ADP exchange factor.

15.2 Other Heat Shock Proteins also Have Chaperone Activity

There are several other families of heat shock proteins; in many cases we know little about them. The following examples, Hsp90 and sHsp, were selected because some coarse picture of their function and reaction mechanism is emerging:

15.2.1 Hsp90

Hsp90 acts as a homodimer (see Fig. 15.4) and is involved not so much in the *de novo* folding of proteins but in several regulatory processes inside the cell by binding to key proteins such as steroid hormone receptor (SHR, see Fig. 15.5) or src-kinase. It usually acts together with Hsc70 and several other chaperone proteins. Still unclear is to what extent Hsp90 is involved in the cellular stress response [9]. In eubacteria there is a homologue protein, high-temperature protein G (HtpG), which is absent in archaea. In eukaryotes, up to 3 % of cellular protein may be Hsp90, which occurs in two isoforms, called α and β. Deletion is lethal during early embryonic development.

Many details of its reaction cycle, and the function of several of its cochaperones, are still unclear [6, 14]. These other proteins often have a characteristic binding site for Hsp90 and substrate proteins, called the **tetratricopeptide repeat**, a couple of helices which form a grove into which substrate peptides can bind in extended conformation. One example of such proteins is a family of Hsp90-dependent **Peptidyl prolyl *cis/trans*-isomerases (PPIs)**, which catalyses the conversion between

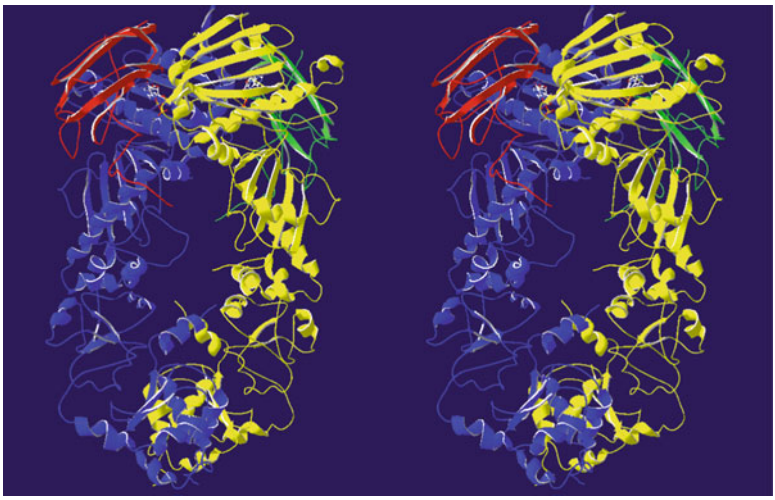


Fig. 15.4 Stereo image of the crystal structure of Hsp82, the stress-inducible isoform of Hsp90 in yeast. The protein (*yellow and blue*) has ATP bound and the cochaperone SBA1 (*red and green*). The PDB-code is 2cg9. There is a nucleotide-binding (N-)domain, a middle domain, and the substrate-binding C-domain. Initial data suggest that once ATP has been hydrolysed the N- and M-domains move apart, leading to a more open conformation of the protein (not shown)

the *cis*- and *trans*-conformation of proline residues in the protein chain (see Fig. 2.4 on page 18). Thus Hsp90 appears somewhat like a lathe, on which different tools (cochaperones) can be used to work on substrate proteins.

Heat Shock Proteins and Cancer
Geldanamycin, radicicol, and their derivatives are inhibitors of Hsp90 ATPase activity; celastrol and gedunin inhibit interactions with its cochaperones. They are used in cancer treatment (see Fig. 15.6). Other heat shock proteins are also investigated as targets for cancer therapy [23], as they are required for cancer growth and their overexpression indicates a poor prognosis [2, 19], in the case of Hsp90 especially if changed posttranslational modification leads to its expression on the cell surface or exosomal secretion. Extracellular Hsp90 may increase invasiveness of tumors by making target proteins more susceptible to MMP digestion.

In Fig. 15.5 interaction of Hsp90 and the steroid hormone receptor is explained in detail. As are other chaperones, Hsp90 is an ATPase; mutations that prevent ATP hydrolysis destroy the chaperone function of Hsp90.

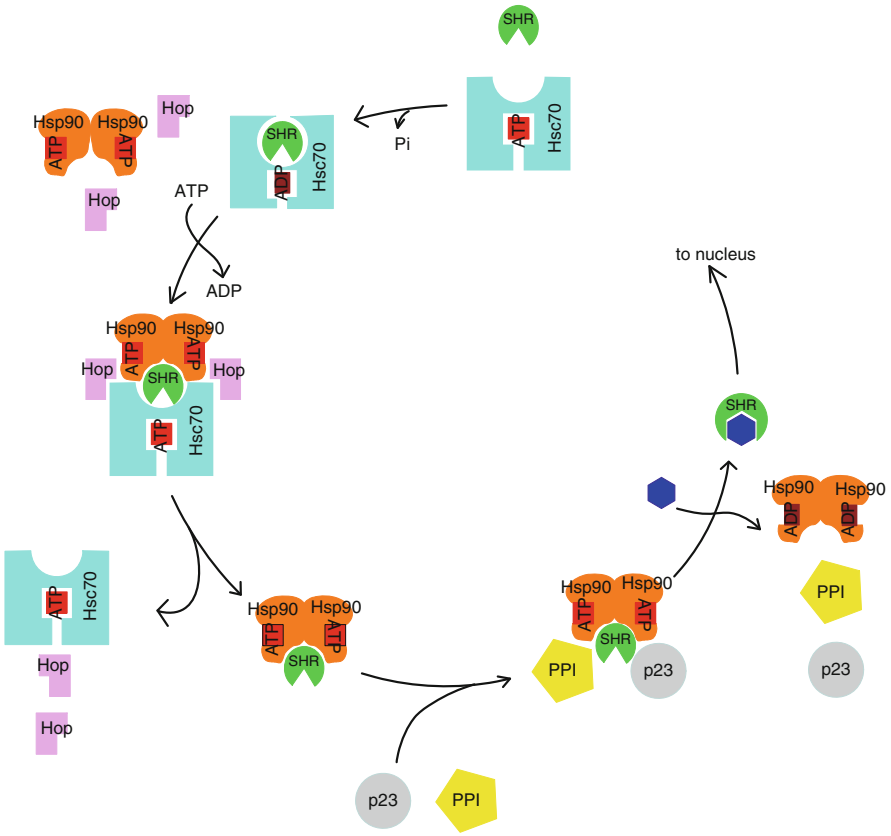


Fig. 15.5 Cooperation between Hsc70 and Hsp90 is required to keep the **SHR** in a binding-competent state. Steroid hormones are very hydrophobic; they passively diffuse through the plasma membrane into the cytosol, where they are bound by SHR. The resulting complex is transported into the nucleus, where it regulates the transcription of certain genes. In order to bind the hydrophobic steroids, SHR needs exposed hydrophobic patches which, given the high protein concentration inside a cell, might lead to aggregation. Also, bound steroid is required to keep the free receptor in its native conformation. Hsc70-ATP binds empty SHR (either newly translated or after it has lost previously bound steroid). Hydrolysis of the ATP bound to Hsc70 leads to the closing of its protein binding site, trapping the bound SHR. This complex binds Hsp90 and a cochaperone, **p60^{Hop}**, which stimulates ADP/ATP exchange in Hsc70. Thus its protein binding site opens and the SHR is transferred to Hsp90. Hsp90 not only keeps SHR in a binding competent state, but also acts as a lathe, where other proteins can bind and perform repair work on SHR. Examples include **PPI**, which catalyses the conversion between the *cis*- and *trans*-conformation of proline residues in the protein chain. Also involved is a protein of unknown function, **p23**. Binding of steroid to SHR leads to its dissociation from Hsp90; during this process Hsp90-bound ATP is hydrolysed

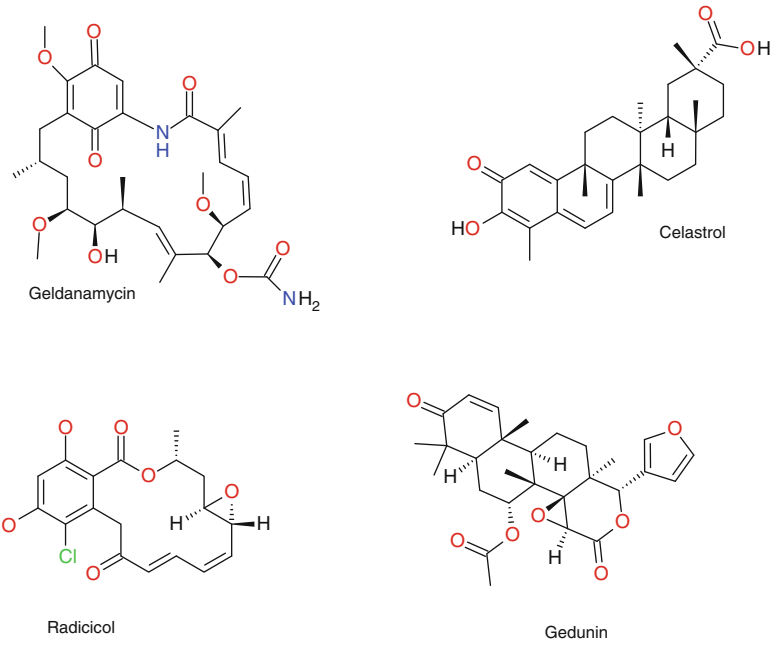


Fig. 15.6 Inhibitors of Hsp90 with antitumor effect. Geldamycin is a macrocyclic polyketide from *Streptomyces hygroscopicus*. Radicicol is a modified polyketide from the fungus *Pochonia chlamydosporia* ((GODDARD) ZARE & W. GAMS 2001). Celastrol is a pentacyclic triterpenoid from root extracts of *Tripterygium wilfordii* (HOOK.f., Thunder duke vine) and *Celastrus regelii* (SPRAG. & TEKEDA, REGEL's Threewingnut). Gedunin is isolated from the Neem tree *Azadirachta indica* A.JUSS., 1830

15.2.2 Small Heat Shock Proteins

In the presence of substrate proteins small heat shock proteins (sHsp) form globular oligomers with 12–42 subunits; these oligomers have a central cavity lined with hydrophobic amino acids. Lens α-crystallin, for example, consists of 32 subunits, forming a globule with 190 Å outer diameter and a cavity of 100 Å. Thus sHsp can bind considerable amounts of substrate proteins; substrate binding and oligomer formation is a highly cooperative process. sHsp cannot only prevent the precipitation of misfolded proteins, but to some extent even solubilise aggregates. However, there seems no set pathway for releasing bound proteins; some sHsps do not even bind ATP. Instead, substrate proteins are transferred to Hsc70 or Hsp70. Little is known about the reaction cycle of small heat shock proteins.

The most well-known sHsp are the **crystallins** found in our eye lenses [12, 21]. Eye lenses are made from fibre cells; these cells are added to the outside of the lens throughout life. Once made, the fibre cells stop protein turnover and there is no exchange of proteins between cells. Thus the proteins in the core of the lens

were made during embryogenesis. Protein concentration in fibre cells is very high (450 mg/mL) to achieve a high refractive index.

Any protein unfolding would result in aggregation; the aggregate would scatter light. Such a particle is called a **cataract**.

To protect the proteins if not from unfolding then at least from aggregation, during a life span of several decades, almost 90 % of the protein in fibre cells are crystallins. There are three major isoforms: α, β and γ. Of these α-crystallins are the most common, which come in two isoforms, α_A (75 %) and α_B (25 %). α_B-crystallin (encoded on chromosomes 11) is found in most cells of the body, α_A-crystallin (encoded on chromosome 21) only in the lens.

After birth most of the α-crystallin in the lens is water soluble, with increasing age more and more is in the water-insoluble fraction (i.e., bound to other proteins), beyond age 40 little soluble α-crystallin is found. Bound proteins include β- and γ-crystallins, housekeeping enzymes, cytoskeleton, and proteins of the plasma membrane.

Because of the large irregular oligomers crystallisation of crystallins is difficult. Of course, a low tendency to form crystals even at high concentration—or to undergo phase separation—is a very useful property for a crystallin. Phosphorylation of α-crystallin seems to reduce the size of the oligomers; its functional significance is unclear.

Although α-crystallin can bind ATP there is no ATPase activity, and the function of ATP-binding is not known.

Several mutations in α-crystallin have been linked to inherited cataracts, including R120G in α_B- and R116C in α_A-crystallin (desmin-related myopathy, the sensitivity of crystallins to proteolysis is increased). In some cancers and some infections (both viral and bacterial) expression of α_B-crystallin is upregulated, to ease cell division or pathogen multiplication, respectively. This also results in immune system stimulation, leading to hyperimmunity but, at the downside, also to autoimmunity. Pharmaceuticals that increase sHsp-expression as adjuvant for immunisation, or that decrease sHsp-function to suppress autoimmunity are under development.

15.3 The GroEL/GroES-Foldosome Is an Example for Molecular Chaperonins

In bacteria, the GroES/GroEL foldosome (the name indicates that mutations in these proteins inhibit the growth of certain bacteriophages) consists of 2 beakers of 45 Å inner diameter, each formed by 7 GroES (Hsc60) molecules (recently reviewed

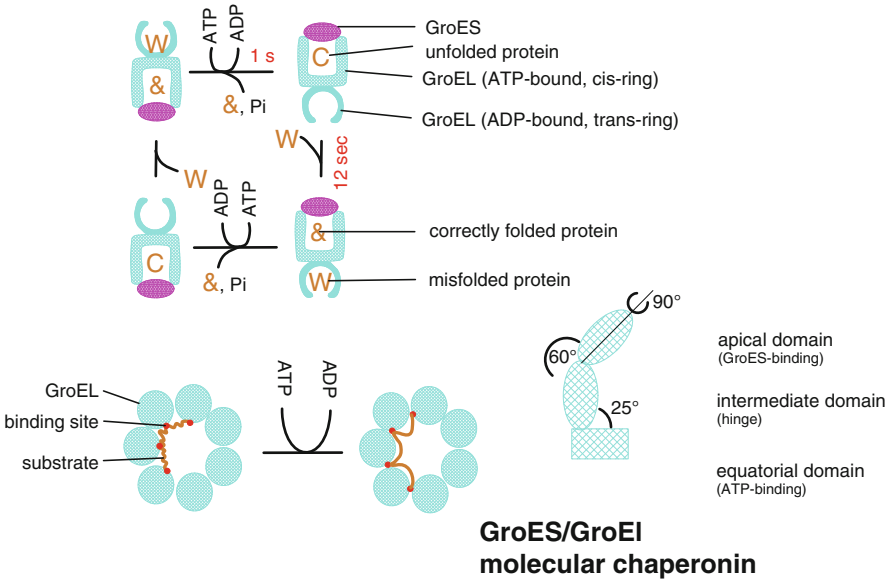


Fig. 15.7 *Top*: The foldosome GroEL/GroES is a typical chaperonin. It can unfold misfolded proteins (W) and provides a protected space for unfolded proteins (C) to achieve folding (&)

in [20]). One of these beakers is covered by a lid, which is formed by 7 GroEL (Hsc10) molecules. This is called the *cis*-ring; all its GroES-subunits have ATP bound. The other beaker is called the *trans*-ring; it does not have a GroEL-lid and all its GroES-subunits contain bound ADP. In thermophilic archaea such as *Thermococcus sp.* the chaperonin is a homo-octamere with a built-in lid, apparently to increase stability against thermal dissociation. In eukaryotes a similar complex, called Tailless complex polypeptide 1 Ring Complex (TRiC) exists, which is a hetero-oligomer of 8 different subunits. It is assumed that function and mechanism of the various chaperonins are similar.

Each GroES-subunit consists of three domains, called equatorial, intermediate and apical (see Figs. 15.7 and 15.8). The equatorial domain contains the nucleotide binding site, the intermediate domain binds the substrate protein, and the apical domain binds GroEL.

The two GroES beakers go through the same reaction sequence, but with a phase difference of 180°. A misfolded protein (indicated by W in Fig. 15.7) is bound in the cavern of the *trans*-ring. Each of the 7 GroES-molecules forming the ring cooperatively exchanges an ATP-molecule for its bound ADP. This results in a change of conformation that leads to the unfolding of the substrate protein (the unfolded substrate is indicated by a C). By this conformational change what was the *trans*-ring now becomes the *cis*-ring.

The *cis*-ring is covered by a GroEL-hat. Thus the substrate protein is enclosed in a protected space (ANFINSEN-cage), where it has time to fold correctly under

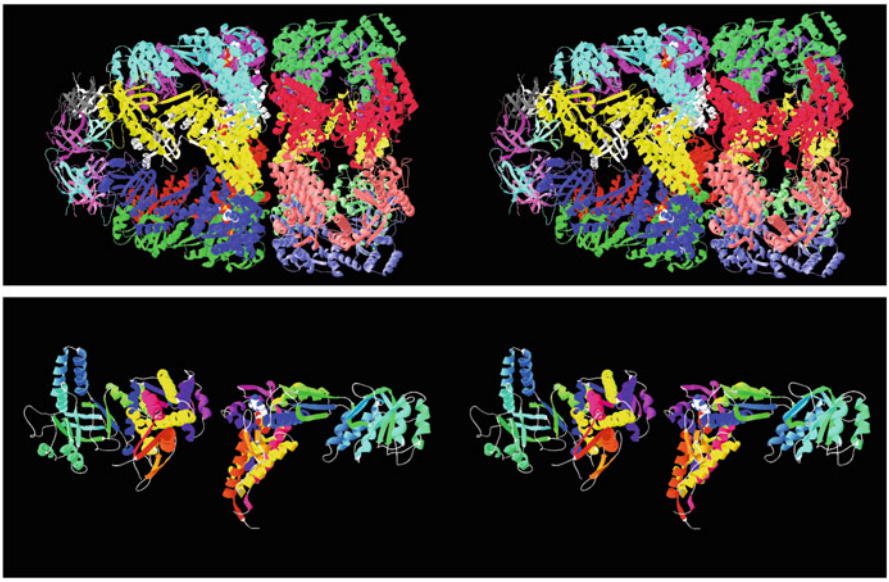


Fig. 15.8 *Top*: Crystallographic structure of the GroEL/GroES-foldosome from *E. coli* at 2.9 Å resolution (PDB-code 1PF9). One ring of GroES has ADP bound in each subunit and is covered with the GroEL lid; the other ring is nucleotide free. *Bottom*: Comparison of the GroES-subunits with and without nucleotide (same molecule as above, but only one subunit from each ring shown). Note the considerable change in tertiary structure

"**infinite dilution**". Note that the GroES/GroEL foldosome does not catalyse folding: it only prevents access of other proteins, which might aggregate with the substrate. Additionally, enclosure of the protein in a small space changes the thermodynamics of folding [24]. Correct folding is achieved by the substrate protein alone, in line with the ANFINSEN-hypothesis.

Once the substrate protein has achieved correct folding (& in the drawing), bound ATP is hydrolysed, the GroEL-lid is removed and the substrate released from what has become the *trans*-ring again. One ATPase-cycle takes about 15–30 s.

How can the foldosome decide when a protein is correctly folded? Misfolded proteins usually have hydrophobic amino acid residues exposed at the surface; these are buried inside correctly folded proteins. Each GroES subunit has a hydrophobic patch, which can interact with hydrophobic residues in the substrate protein in the ATP-bound *cis*-state. In the ADP-bound (*trans*-)state these residues face the interface between the proteins of the beaker.

ATP/ADP-exchange in GroES results in a considerable conformational change, which twists the substrate binding sites (red patches in the lower left half of Fig. 15.7) from the centre of the cavity to the side, thereby increasing the distance between these patches. This tears apart the inappropriate folding of the substrate. Once the conformational change in GroES is completed, the hydrophobic patches are buried between the interfaces of the GroES-subunits. In other words: no binding

sites are available for the substrate protein, which lies now unfolded and unbound inside a protective cavity, which is lined exclusively with hydrophilic amino acids. In this state GroES also binds the GroEL lid, thus the protective cavity is totally closed. In this womb the substrate protein can try to fold correctly.

If the protein manages to achieve a correct conformation, it will no longer have exposed hydrophobic residues. Thus it can be released once the GroEL-lid is removed. If correct folding is not obtained, the substrate will rebind to the GroES in the next cycle. This can be repeated several times, until the substrate is correctly folded.

In *E. coli* about 10 % of all proteins produced transiently interact with GroES/-GroEL; this figure triples under heat stress. The complex preferentially binds proteins between 10–55 kDa; about 60 kDa proteins would fit into the beaker. Larger proteins can, however, also be bound if they consist of several independently folding domains. GroES/GroEL has a preference for proteins with α/β-structure (see Sect. 2.3.1 on page 34); it may be required to produce the correct long-range hydrogen bonds required for β-sheet formation and for bringing the α-helices into a correct orientation with respect to the β-sheet.

Although we have looked at chaperones and chaperonins separately, in a living cell they work closely coordinated and sequentially as "folding machines". GroES/GroEL can accept substrate proteins from DnaK, which in turn can accept them from other heat shock proteins.

15.4 Exercises

15.4.1 Problems

15.1. BERRY and coworkers [3] described a four-generation British family suffering from congenital bilateral posterior polar cataract, that was 0.5–3 mm in diameter and did not progress during the lifetime of the patients. No other visual or systemic problems were coinherited. Inheritance was autosomal dominant.

In a mapping study, which of the following proteins would be most the likely culprit?

A Hsc70 (70 kDa Heat shock cognate)
B Hsp70 (70 kDa Heat shock protein)
C HspBP1 (nucleotide exchange factor)
D α B crystallin
E $\text{Grp78}^{\text{BiP}}$

15.4.2 Solutions

15.1

D Crystallins in the eye lens have two functions: Their high concentration increases the refractive index of the lens, and they serve as molecular chaperones that prevent the aggregation of lenticular proteins and hence cataract formation.

References

1. C. Anfinsen, Principles that govern the folding of protein chains. Science **181**, 223–230 (1973). doi:~10.1126/science.181.4096.223
2. J.J. Barrott, T.A.J. Haystead, Hsp90, an unlikely ally in the war on cancer. FEBS J. **280**(6), 1381–1396 (2013). doi:~10.1111/febs.12147
3. V. Berry and P. Francis and M.A. Reddy and D. Collyer and E. Vithana and I. MacKay and G. Dawson and A.H. Carey and A. Moore and S.S. Bhattacharya and R.A. Quinlan. Alpha-B Crystallin Gene (CRYAB) Mutation Causes Dominant Congenital Posterior Polar Cataract in Humans. Am. J. Hum. Genet. **69**(5), 1141–1145 (2001). doi:~10.1086/324158
4. T. Böcking, F. Aguet, S.C. Harrison, T. Kirchhausen, Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. Nature Struct. Mol. Biol. **18**(3), 295–301 (2011). doi:~10.1038/nsmb.1985
5. E. Buxbaum, *Biophysical Chemistry of Proteins: An Introduction to Laboratory Methods* (Springer, New York, 2011). ISBN 978-1-4419-7250-7
6. S.K. Calderwood, Molecular cochaperones: Tumor growth and cancer treatment. Scientifica **2013**, Article ID 217513 (2013). doi:~10.1155/2013/217513
7. S. Chowdhury, C. Ragaz, E. Kreuger, F. Narberhaus, Temperature-controlled structural alterations of an RNA thermometer. J. Biol. Chem. **278**, 47915–47921 (2003).~doi:10.1074/jbc.M306874200
8. D.M. Cyr, J. Höhfeld, C. Patterson, Protein quality control: U-box containing E3 ubiquitin ligases join the fold. Trends Biochem. Sci. **27**(7), 368–375 (2002). doi:~10.1016/S0968-0004(02)02125-4
9. J.M. Eckl, K. Richter, Functions of the Hsp90 chaperone system: lifting client proteins to new heights. Int. J. Biochem. Mol. Biol. **4**(4), 157–165 (2013). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3867702/pdf/ijbmb0004-0157.pdf
10. D.E. Feldman, J. Frydman, Protein folding *in vivo*: the importance of molecular chaperones. Curr. Opin. Struct. Biol. **10**, 26–33 (2000). doi:~10.1016/S0959-440X(99)00044-5
11. O. Genest, J.R. Hoskins, J.L. Camberg, S.M. Doyle, S. Wickner, Heat shock protein 90 from Escherichia coli collaborates with the DnaK chaperone system in client protein remodeling. Proc. Natl. Acad. Sci. USA **108**(20), 8206–8211 (2011). doi:~10.1073/pnas.1104703108
12. J. Horwitz, Alpha-crystallin. Exp. Eye Res. **76**, 145–153 (2003). doi: 10.1016/S0014-4835(02)00278-6. URL http://www.life.illinois.edu/ib/426/handouts/Alpha_Crystallin_Review_Horwitz_EER2003.pdf
13. B.I. Kurganov, Review: Antiaggregation activity of chaperones and its quantification. Biochemistry (Mosc). **78**(13), 1554–1566 (2013). doi:~10.1134/S0006297913130129
14. J. Li, J. Buchner, Structure, function and regulation of the Hsp90 machinery. Biomed. J. **36**(3), 106–117 (2013). doi: 10.4103/2319-4170.113230
15. Z. Li, F.U. Hartl, A. Bracher, Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. Nature Struct. Mol. Biol. **20**(8), 929–935 (2013). doi:~10.1038/nsmb.2608

16. B. Liu, C.S. Conn, S.-B. Quian, Viewing folding of nascent polypeptide chains from ribosomes. Expert Rev. Proteomics **9**(6), 579–581 (2012). doi:~10.1586/EPR.12.57
17. B.W. Matthews, Solvent content of protein crystals. J. Mol. Biol. **33**, 491–497 (1968). doi:~10.1016/0022-2836(68)90205-2
18. R.U.H. Mattoo, P. Goloubinoff, Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. Cell. Mol. Life Sci. **71**(17), 3311–3325 (2014). doi:~10.1007/s00018-014-1627-y
19. J.R. McConnell, S.R. McAlpine, Heat shock proteins 27, 40, and 70 as combinational and dual therapeutic cancer targets. Bioorg. Med. Chem. Lett. **23**(7), 1923–1928 (2013). doi:~10.1016/j.bmcl.2013.02.014
20. N.A. Ryabova, V.V. Marchenkov, S.Yu. Marchenkova, N.V. Kotova, G.V. Semisotnov, Molecular chaperone GroEL/ES: Unfolding and refolding processes. Biochem. (Mosc.) **78**(13), 1405–1414 (2013). doi:~10.1134/S0006297913130038
21. C. Slingsby, G.J. Wistow, A.R. Clark, Evolution of crystallins for a role in the vertebrate eye lens. Protein Sci. **22**(4), 367–380 (2013). doi:~10.1002/pro.2229
22. P. Stocki, A.M. Dickinson, The immunosuppressive activity of heat shock protein 70. Autoimmune Dis. **2012**, Art. No. 617213 (2012). doi:~10.1155/2012/617213
23. F. Stricher, C. Macri, M. Ruff, S. Muller, HSPA8/HSC70 chaperone protein: Structure, function, and chemical targeting. Autophagy **9**(12), 1937–1954 (2013). doi:~10.4161/auto.26448
24. F. Takagi, N. Koga, S. Takada, How protein thermodynamics and folding mechanism are altered by the chaperonin cage: Molecular simulations. Proc. Natl. Acad. Sci. USA **100**, 11367–11372 (2003). doi:~10.1073/pnas.1831920100
25. S. Walter, J. Buchner, Molecular chaperones – cellular machines for protein folding. Angew. Chemie Int. Ed. **41**, 1098–1113 (2002). doi:~10.1002/1521-3773(20020402)41:7%3C1098::AID-ANIE1098%3E3.0.CO;2-9
26. G. Zhang, M. Hubalewska, Z. Ignatova, Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nature Struct. Mol. Biol. **16**, 274–280 (2009). doi:~10.1038/nsmb.1554

Part IV
Membrane Transport

Chapter 16
Protein Transport Across Membranes

Abstract Mitochondria contain their own DNA, they divide throughout interphase and are distributed randomly between daughter cells during cell division. Most mitochondrial proteins are encoded in the nucleus and produced in the cytosol. They are then imported into the mitochondria via the TOM/TIM transport system, which recognises them by a specific localisation sequence which is cleaved off once a protein has entered a mitochondrium. Mitochondrial lipids are made by the ER. Similar pathways exist for plastids, peroxisomes and nucleus. However, nuclear transport sequences are not removed after transport as the nuclear membrane breaks down during mitosis, releasing nuclear proteins into the cytosol. They need to be reimported afterwards. Insertion of proteins into the ER occurs cotranslationally. While they are transported into the ER, proteins are glycosylated. The oligosaccharides added to Thr or Ser or Hyl (O-linked) are simple, but some are important as blood group antigens. Very complex sugar trees are added to Asn (N-linked). These trees play an important role in the quality control of the ER, which prevents the production of misfolded proteins. Additional sugar residues are added during the protein's passage through the GOLGI apparatus.

16.1 Structure of Membrane Components

16.1.1 Membrane Lipid

The **fluid mosaic model** of biological membranes (see Figs. 16.1 and 16.2) was proposed by SINGER & NICOLSON in 1972 [30]. However, the membrane is not homogeneous. Phospholipids, sphingolipids and cholesterol mix only at certain ratios. Thus, the outer leaflet of the cell membrane has areas rich in cholesterol and sphingolipid that are in the liquid ordered (L_o-)phase and swim in a sea of phospholipid in the lamellar crystalline (L_α-) phase [18, 23] (see Fig. 16.3). The composition of the inner leaflet is different and more homogeneous. The cholesterol/sphingolipd-rich areas are called **rafts** [29]; each cell contains about 10^6 of them. They are thicker than the surrounding membrane (48 vs. 39 Å) and are more resistant to solubilisation with detergent. The also contain glucosylceramide, cholesterylglucoside, and glycosylphosphatide; these glucosylated lipids seem essential for life, but we are only now figuring out why [17]. Membrane proteins

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_16

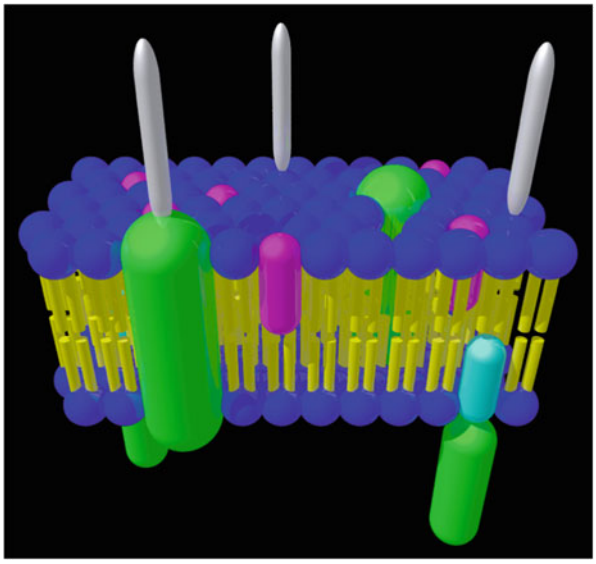


Fig. 16.1 **Fluid-mosaic model** of the biological membrane. **Phospholipids** form a bilayer with the hydrophobic tails of the lipids (*yellow*) buried inside, and the hydrophilic head groups (*blue*) exposed to the aqueous medium on both sides. **Cholesterol** (*magenta*) is also present. In this matrix **transmembrane proteins** are embedded (*green*); additionally proteins are bound to the membrane by **GPI-anchors** on the extracellular side (*top*) and **acyl**- or **prenyl-residues** (*cyan*) on the cytoplasmic side (*bottom*). **Sugar side-chains** (*grey*) may be attached to proteins and lipids on the extracellular side of the membrane

associate with either rafts or "normal" membrane areas [27], depending on the length of their transmembrane α-helices. Proteins with acyl or GPIgroups prefer rafts, proteins with prenyl groups phospholipid-rich membrane segments. There are several subpopulations of rafts, that have different proteins associated with them and differ in size and lifetime. **Caveolae** are specialised rafts that occur in membrane invaginations stabilised by the protein caveolin. Lateral membrane structure is highly dynamic; cell signalling and apoptosis are but two processes influenced by it [16].

16.1.2 Membrane Proteins

There are 6 types of membrane proteins, as depicted in Fig. 16.4:

Type 1 protein synthesis starts in the same way as that of proteins destined to the ER lumen. The ribosome/mRNA/protein complex attaches to the translokon, and the signal sequence binds to TRAM. The nascent protein is imported

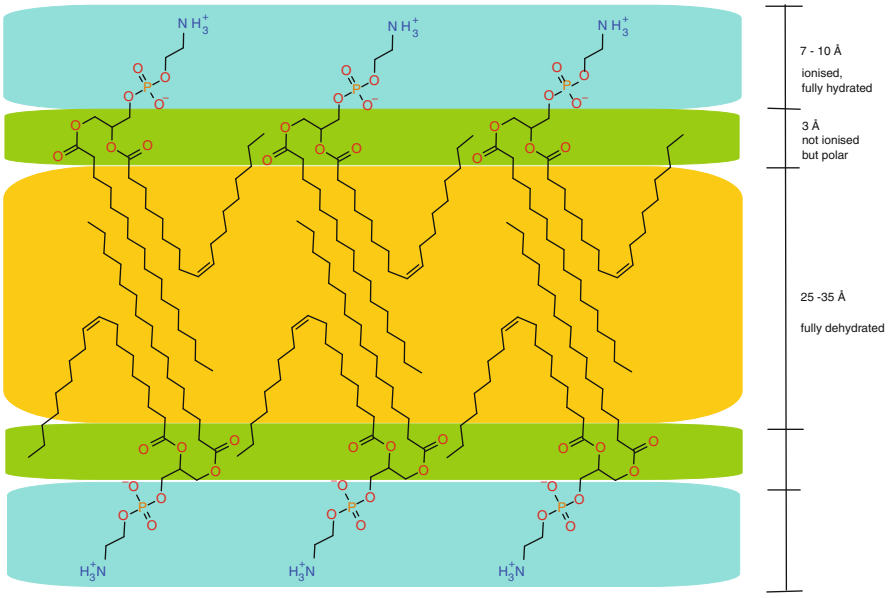


Fig. 16.2 Arrangement of phospholipid molecules in the membrane. The lipid chains are tilted at an angle ϕ^0 from the normal to the membrane surface. This angle depends on the packing of the head groups, and hence on their composition and on the temperature. Fatty acids on the 2-position of glycerol tend to be longer (C18–22) than those in 1-position (C16–18), they also are more likely to be unsaturated. Unsaturation prevents dense packing of lipids and hence increases fluidity of the membrane. Fatty acids from both leaflets interdigitate in the centre of the membrane; the hydrophobic interactions between them stabilise the membrane

into the ER, until a **stop-transfer signal** is encountered. This is a stretch of 20–25 hydrophobic amino acids, which form a transmembrane α-helix. Inasmuch as this sequence, once inserted into the membrane, cannot be pushed into the aqueous environment of the ER lumen, the amino acids' C-terminal of it stay in the cytosol. Cleavage and digestion of the signal sequence results in a type 1 integral membrane protein, with the C-terminus on the cytosolic side and the N-terminus in the lumen of the ER. If the stop-transfer signal is deleted from the gene of a type 1 protein, or if one of the amino acids encoded by it is mutated into a charged one, the protein enters the ER lumen and is eventually secreted.

Type 2 proteins have an internal **signal-anchor sequence**, so the N-terminus stays on the cytosolic side of the membrane. All amino acids' C-terminal of the signal-anchor sequence are translocated into the ER-lumen; the orientation of type 2 proteins is opposite to that of type 1. Signal-anchor sequences also form 20–25 amino acid transmembrane helices. As a general rule the cytosolic end of a membrane helix has positively charged amino acids. This situation is stabilised by the negative potential of the cytosol with respect to the ER-lumen. This

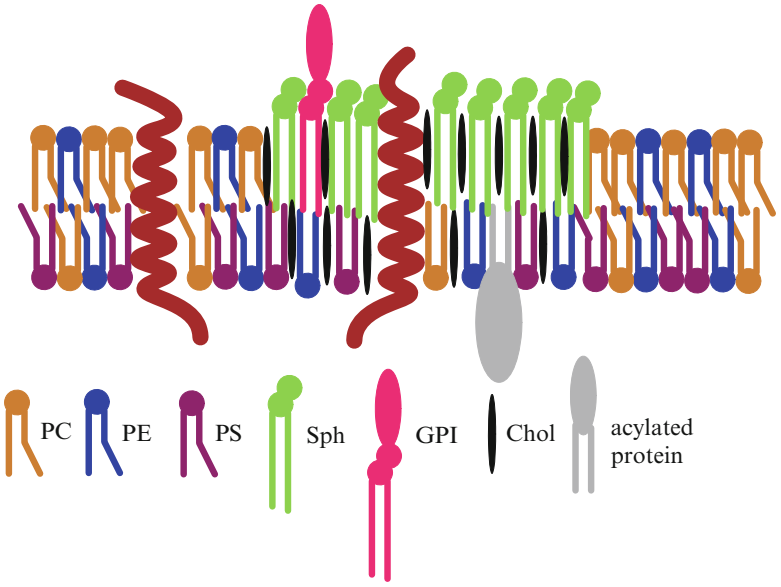


Fig. 16.3 Rafts swim in a sea of glycerophospholipids. The outer leaflet of the plasma membrane is rich in phosphatidylcholine (PC) and phosphatidylethanolamine (PE); the inner leaflet in addition contains phosphatidylserine (PS). Both in the inner and outer leaflet the acyl group in the 2-position of glycerol is usually unsaturated, resulting in a disordered (L_α-) structure. In rafts, the outer leaflet is made of sphingolipids (Sph) that contain saturated, long-chain fatty acids. Cholesterol partitions between these molecules. The inner leaflet in rafts is made up of glycerophospholipids, but with longer, saturated acyl-chains. As a result rafts are thicker than membranes made of glycerophospholipids (48 vs 39 Å), and have a more ordered (L_o-)structure

"**positive inside**" rule determines whether a hydrophobic helix acts as a stop-transfer or signal-anchor sequence. Long hydrophobic stretches tend to function as stop-transfer signals; signal-anchor sequences tend to be shorter.

Type 3 proteins span the membrane several times. They have repeated signal- and stop-transfer sequences. Once the first of these pairs has left the translocon, anchoring the protein in the membrane, protein synthesis continues in the cytosol, further signal-/stop-transfer sequence pairs are inserted into the membrane using the translocon, but without SRP and SRP-receptor. Originally both N- and C-termini face the cytosol, however, posttranslational proteolysis can change that.

Type 4 proteins are made by the combination of several polypeptides (either type 1 or 2) into a single quaternary structure.

Type 5 proteins are synthesised in the cytosol; they are chemically linked to hydrophobic molecules anchored in the cytosolic leaflet of the ER-membrane. Three subtypes are known:

- Internal Cys or Ser residues of the protein are chemically linked to palmityl-groups (C16:0).
- N-terminal Gly-residues are linked to N-myristoyl-groups (C14:0).

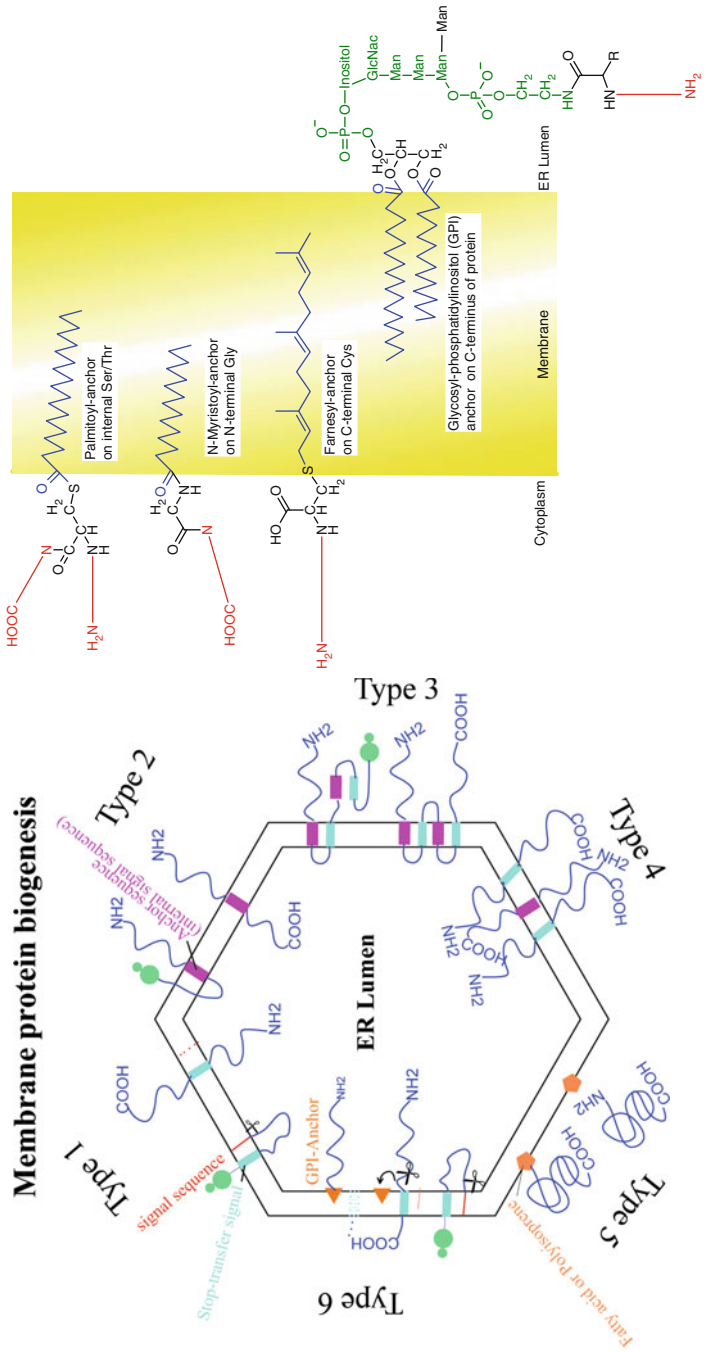
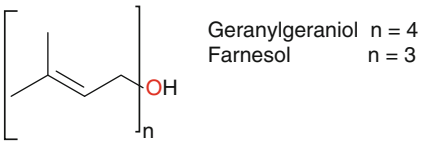


Fig. 16.4 *Left*: There are 6 classes of membrane proteins. *Right*: Linkage of proteins to the membrane by hydrophobic anchors (for type 5 and 6)

- C-terminal Cys is linked to farnesyl- or geranylgeranyl-groups.



Linking proteins to acyl or prenyl groups can convert an inactive cytosolic protein into a membrane-bound active one. This regulatory function of the enzymes responsible for prenylation or acylation has excited considerable pharmacological interest. For example, statins may work in part by preventing prenylation of proteins [32]. Farnesyltransferase inhibitors are currently in clinical trial against cancers with abnormal Ras-protein, and for children with progeria.

Type 6 proteins are linked to the luminal leaflet of the ER membrane by a **Glycosyl phosphatidylinositol (GPI)** residue bound to their C-terminus. These proteins are synthesised like type 1 proteins; later a C-terminal signal peptide of $\approx$ 30 amino acids is cleaved off by an endoprotease, and the new C-terminus is transferred to the already fully synthesised GPI-anchor at the same time. This step removes the cytosolic bit of the original protein, which would interact with the cytoskeleton and prevent free diffusion of the protein in the membrane plane. The GPI-anchor in plasma membrane proteins also acts as a **sorting signal**, directing the protein to the **apical** domain of the cell.

16.2 Transport of Proteins into Mitochondria

16.2.1 The Mitochondrium in the Cell Cycle

Mitochondria and plastids were derived from **endosymbiont** bacteria about 1–1.5×10^9 a ago, and contain their own DNA, which is circular as in bacteria (see Fig. 16.5). There may be several mtDNA molecules in each mitochondrium; the exact number is species-specific (see Fig. 16.6). The mtDNA is replicated and the mitochondria divide during the entire interphase. Division of mitochondria starts with an invagination of the inner membrane; the outer follows later. Mitochondrial division is not necessarily coupled to cell division: if, for example, a muscle is exercised, the number of mitochondria per muscle cell increases. The number of mitochondria per cell can be quite large, about 1000 in a rat liver cell.

Because of their large number, mitochondria are distributed randomly ("stochastically") between the daughter cells during mitosis, just like the vesicles

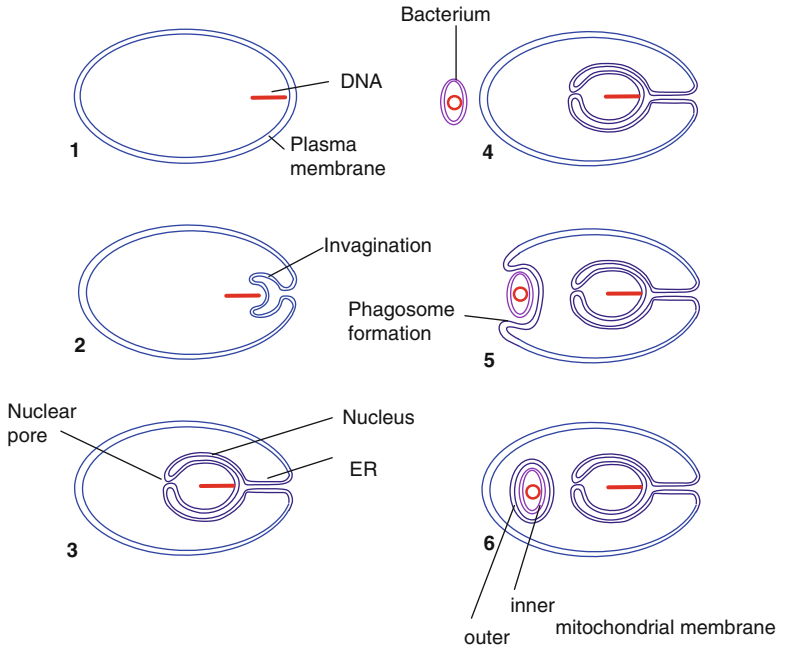


Fig. 16.5 Origin of the modern eukaryotic cell: Organelles are derived either by invagination of the original prokaryotic cell membrane (*left*, ER, Golgi, nucleus) or by the uptake of endosymbionts (*right*, mitochondria, plastids) [19]. The inner mitochondrial and plastid membranes are derived from the cell membranes of the symbiotic organisms; the outer membranes are remnants of the phagosome. Note that the intermembrane space, as well as the inside of ER and Golgi apparatus, are topologically "outside". The environment in these compartments resembles that of the interstitial fluid (high Na^+ and Ca^{2+}, low K^+ and Mg^{2+}, oxydising)

derived from nucleus, ER, and other internal membrane systems. Thus both daughter cells receive approximately, albeit not exactly, the same number of mitochondria.

16.2.2 Mitochondrial Proteins and Lipids

All proteins encoded by mtDNA are synthesised on mitochondrial ribosomes, which are associated with the inner membrane. However, most mitochondrial proteins are encoded by nuclear DNA, synthesised in the cytosol and posttranslationally imported into the mitochondrium. No proteins encoded by mtDNA are exported from the mitochondrium. The mitochondrial genome contains 37 genes:

- the genes for 13 subunits of OxPhos enzymes
- 2 rRNAs and 22 tRNAs

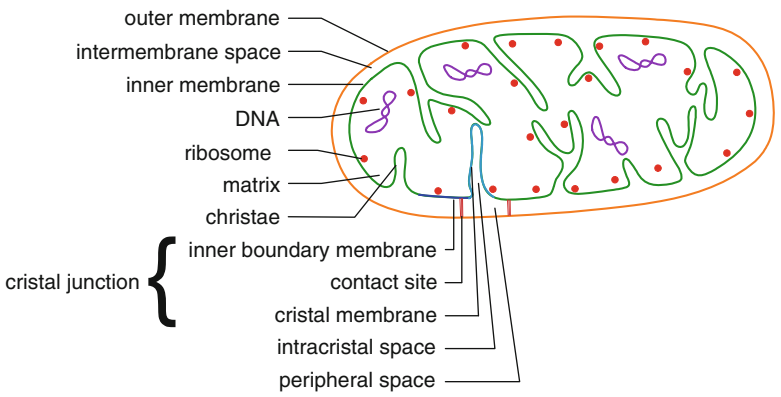


Fig. 16.6 Schematic diagram of a mitochondrium

Enzyme	Complex	mtDNA	nuDNA
NADH + H^+ oxydase	I	7	38
$FADH_2$ oxydase	II	0	4
Succinate dehydrogenase	III	1	10
Cytochrome c oxydase	IV	3	10
ATP synthase	V	2	14

Mitochondrial Inheritance of Disease

Because of the 423 proteins found inside the mitochondrium only 13 are encoded on the mtDNA, most genetic diseases affecting the mitochondrium have autosomal recessive inheritance. Mutations in mtDNA lead to maternal inheritance as few if any of the mitochondria of the sperm enter the egg during fertilisation. Random segregation of mitochondria during cell division in these cases leads to the phenomenon of **homoplasmy** (a cell contains either only affected or only unaffected mitochondria) and **heteroplasmy** (a cell contains a mixture of affected and unaffected mitochondria). A cell is affected if a certain percentage of its mitochondria are defective (threshold effect). As a result, the severity of mitochondrially inherited diseases can vary significantly not only from person to person, but also from tissue to tissue in the same person. Mitochondrial diseases in a predisposed individual can break out at any time and in any organ, but organs with heavy respiratory activity are most likely to be affected:

eye LEBER hereditary optic neuropathy (LHON), chronic progressive ophtalmoplegia

nervous system KEARNS-SAYRE-syndrome

(continued)

muscle "ragged red fibre" myopathies

Exposure to toxic substances can increase the risk of outbreak (*e.g.*, alcohol and tobacco).

Mitochondrial lipids are synthesised on the ER. Mitochondria only modify them later; cardiolipin is found only in mitochondria and bacteria (see Fig. 16.7.)

Mitochondria and Antibiotics
Mitochondria have 70 S ribosomes, just as do bacteria. Thus they can be affected by antibiotics which react with bacterial 70 S-, but not eukaryotic 80 S ribosomes (*e.g.* chloramphenicol).

16.3 Synthesis and Sorting of Mitochondrial Proteins in the Cytosol

Many mitochondrial proteins are encoded by nuclear DNA; they are translated on ribosomes in the cytosol.

Proteins coming out of a ribosome are unfolded. In order to prevent fatal protein aggregation, and maintain them in an unfolded, transport-competent state, they are complexed with Hsc70, a molecular chaperone (see Fig. 16.8 and Sect. 15.1 on page 345).

On the N-terminal side of mitochondrial proteins is a signal sequence, which is recognised by a receptor in the outer mitochondrial membrane (TOM). This sequence forms an amphipatic helix, with a positively charged hydrophilic, and an uncharged hydrophobic side (see Fig. 16.9). This structural motive (rather than the amino acid sequence) is recognised. If this motive is transferred to a cytosolic protein by genetic engineering the chimeric proteins get imported into the mitochondria. If the signal sequence is removed from mitochondrial proteins, they can no longer be imported.

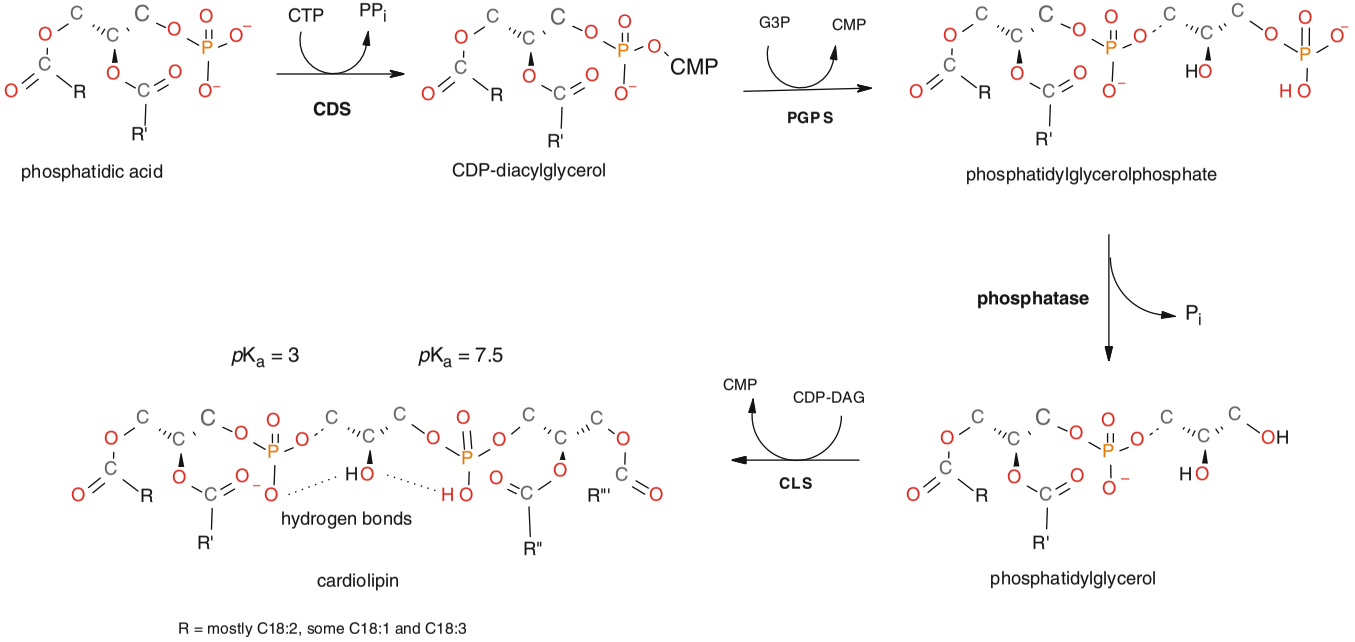


Fig. 16.7 Synthesis of cardiolipin in mitochondria. Cardiolipin makes up 20 % of the inner membrane of mitochondria. It consists of 2 phospholipids, linked by a glycerol bridge. Hydrogen bond formation between the glycerol-OH and the phosphate groups results in a bicyclic structure, that becomes integral part of some membrane enzymes ("core lipid"; see [5, chapter 15]). In addition, because of its pK_a of 7.5, cardiolipin can buffer the protons transported during oxydative phosphorylation. Most of the acyl-groups are C18:2 (linoleic acid), with some C18:1 (oleic acid) and C18:3 (linolenic acid). Cardiolipin was first isolated from beef heart, a tissue rich in mitochondria (hence its name)

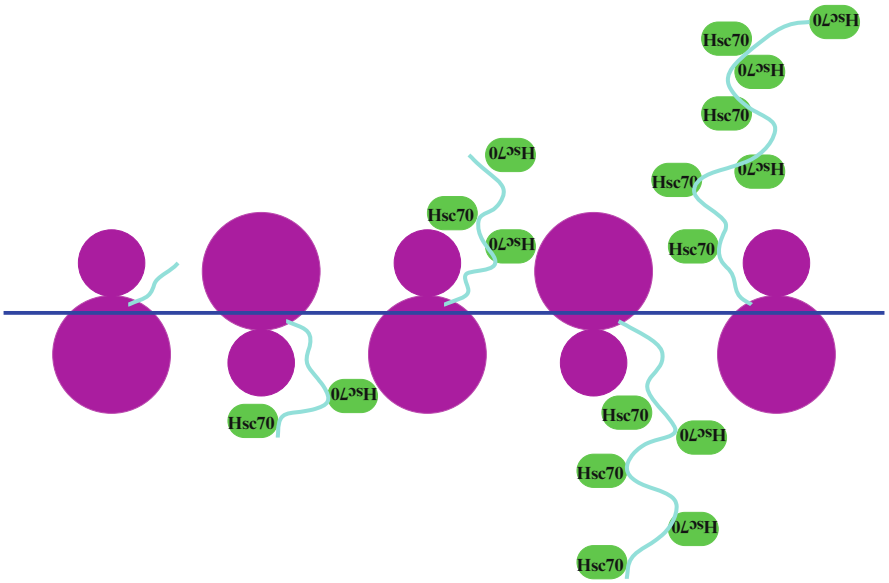


Fig. 16.8 Proteins synthesised on the ribosome are immediately captured by the 70 kDa heat shock cognate protein. This prevents misfolding

Mitochondria have about 1000 TIM/TOM pairs each (transporter of the inner-/outer) membrane; see Fig. 16.10) [14]. The import of proteins from the cytosol into the mitochondria occurs in two distinct steps.

First the signal sequence of the protein is bound to a receptor on the outer membrane of the mitochondrium. This signal sequence, as we have seen, has a number of positive charges. The matrix of the mitochondrium is negatively charged with respect to the cytosol (−200 mV), because endoxydation leads to the transport of protons from the matrix to the cytosol. This is equivalent to a field strength of 400 000 V/cm. As a result the signal sequence is electrophoretically pulled through the outer and inner membrane transporter. Experimentally, this step does not require ATP and can proceed at low temperatures. However, uncouplers such as dinitrophenol (DNP, see page 459), which dissipate the electrical potential difference across the inner membrane, prevent signal sequence transfer.

The rest of the protein is imported in the second step, which depends on ATP-hydrolysis and can occur only at 37, but not at 5 °C. The protein inside the transporters can move either into or out of the mitochondrium, and it will do so randomly (BROWNIAN motion). Thus, in the absence of other factors, no net transport of protein in either direction would occur. However, any stretch of protein that comes into the mitochondrium is immediately bound by mtDnaK. This prevents the protein from slipping back. On the other hand, on the cytosolic side the protein is released from Hsc70, which allows it to move into the mitochondrium. Thus a random movement is used to achieve vectorial transport by the binding and unbinding of 70 kDa heat shock proteins. This mechanism is called **molecular ratchet**.

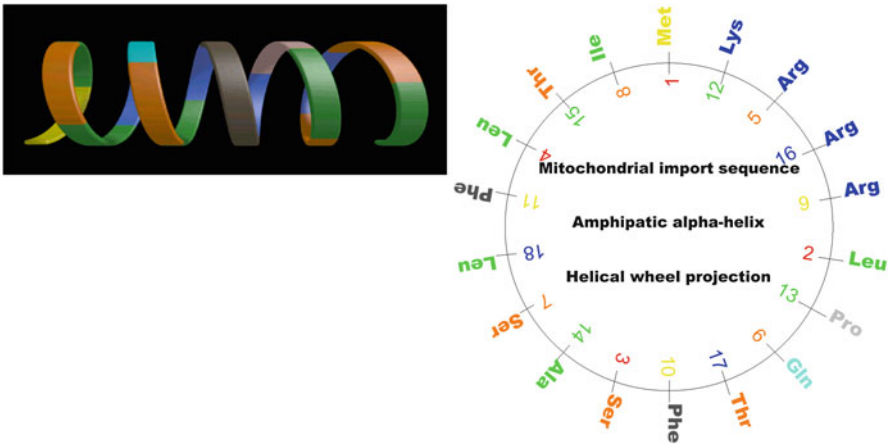


Fig. 16.9 N-terminal signal sequence for import into mitochondria, here the subunit 4 of cytochrome c oxydase from yeast. The sequence MLSLRQSIRFFKPATRTL- forms an α-helix. Every third or fourth amino acid is positively charged (Lys or Arg); because there are 3.6 amino acids per turn in an α-helix all positively charged amino acids (*blue*) are in the same quadrant. The remaining quadrants are occupied by hydrophobic (*green*) or polar (*orange*) amino acids, also a few aromatic (*dark grey*) amino acids are present. This type of helix is called amphipatic. *Left*: Ribbon diagram of the alpha helix; the N-terminal Met is coloured yellow. Turning this helix by 90° around the y-axis results in *right*: Helical wheel projection. The numbers inside the circle denote the position of the respective amino acid in the sequence. These numbers are coloured red for the first turn, orange for the second, yellow for the third, green for the fourth, and blue for the fifth

Some researchers claim that binding of mtDnaK to the imported protein leads to a change in the way mtDnaK interacts with the inner membrane, allowing it to pull the protein actively through the transporter. Evidence for this additional mode of action of mtDnaK is currently limited though.

Once the protein has been imported into the mitochondrial matrix, the signal sequence is cleaved off by **Matrix-processing protease (Mpp)**. Matrix proteins then fold, as mtDnaK dissociates from them. Folding may involve GroEL/GroES chaperonin action. There are additional transport systems for mitochondrial import, however, they are not covered here. The interested reader is referred to [14].

Proteins of the inner mitochondrial membrane or the intermembrane space are transported to their destination directed by a second signal sequence, which is unmasked once the first has been cleaved off. In other words, these proteins (e.g., ATP synthase subunits) are transported from the cytosol through both mitochondrial membranes into the matrix, and from there into the inner membrane or intermembrane space, respectively. This is a reminder of the origin of mitochondria as endosymbionts. Other proteins such as cytochrome c are small enough that in their unfolded form (*e.g.* apocytochrome c, without the hæm group) they can enter the intermembrane space directly through the outer membrane, presumably through a porin-like channel. Folding of these proteins either around a prosthetic group or by disulphide bond formation prevents them from returning to the cytosol.

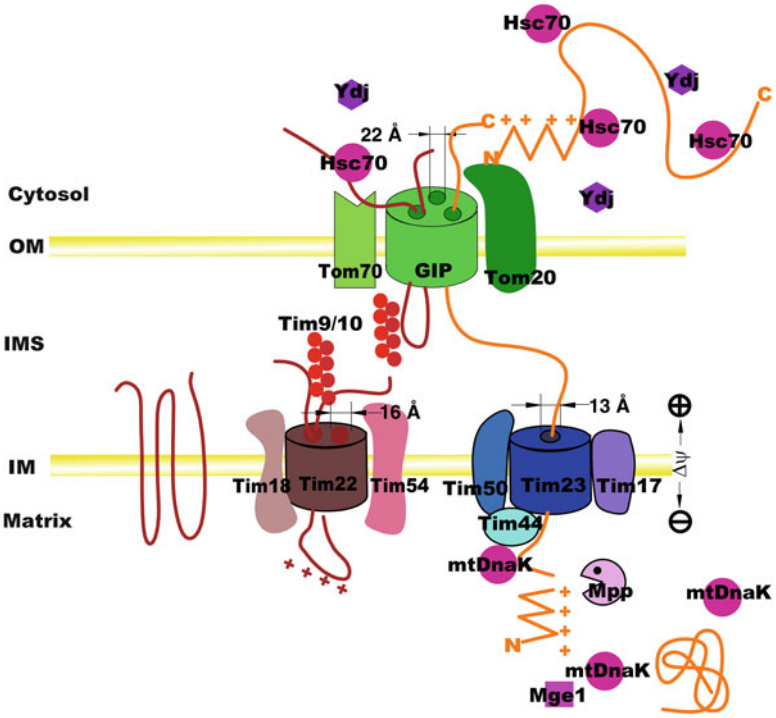


Fig. 16.10 Mitochondrial proteins are transported from the cytosol into the mitochondria by two membrane transporters, called TIM and TOM, respectively. Cytosolic Hsc70 and mitochondrial DnaK are also required. Import occurs only in places where the inner and outer membrane are close together, allowing concurrent transport of the protein across both membranes. The pore through the transporter has a diameter of approximately 2 nm

Synthesis of proteins encoded by mitochondrial and nuclear DNA is synchronised, but the mechanism is still unknown.

16.4 Transfer of Protein into the ER Lumen

The signal for protein import into the ER is a positively charged amino acid near the N-terminus of the protein, followed by a box of 6-15 hydrophobic amino acids (mostly Leu, Ala, Val) followed in turn by polar amino acids plus a Gly or Ala marking the cleavage site [28, 37].

Once this signal sequence emerges from the ribosome (because it is right at the N-terminus that happens early in synthesis) it is bound by the **Signal recognition particle (SRP)**, which halts protein synthesis at the ribosome by also binding to its large subunit. This binding is GTP-dependent. SRP is a ribonucleoprotein, composed of six polypeptides and one RNA-molecule.

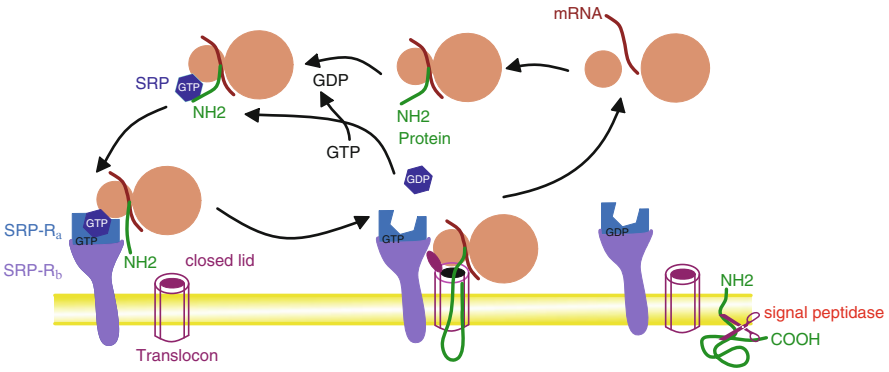


Fig. 16.11 Model for the transfer of nascent proteins into the ER. For explanations see text

The entire complex of SRP, ribosome, mRNA, and nascent protein (about 70 amino acids long at this stage, of these about 40 protrude from the ribosome) is then bound by the **SRP receptor** which interacts with the **peptide translocation complex (translocon)** in the ER membrane (see Fig. 16.11). This leads to hydrolysis of SRP-bound GTP, dissociation of SRP, and resumption of protein synthesis. The hydrolysis of GTP provides the energy required for the formation of the ribosome/nascent chain/translocon complex.

Bound ribosomes give ER-membranes a characteristic appearance under the electron microscope, forming the **rough ER**. Thus rough ER is the site of membrane protein biosynthesis; **smooth ER** lacks ribosomes and fulfills other functions (*e.g.*, lipid biosynthesis, biotransformation).

In the absence of SRP *in vitro* synthesis of proteins destined for the ER can occur normally; if the SRP is added later the transport of the completed proteins into the ER is not possible. However, there are SRP-independent pathways of ER insertion [3].

If the *in vitro* synthesis is performed in the presence of isolated ER membranes (which form vesicles, so-called **microsomes**), the newly synthesised protein is immediately resistant to proteases added to the outside medium. In other words, they are sequestered inside the enclosed space of the microsomes during the process of synthesis.

Interaction of the SRP/mRNA/ribosome/nascent protein complex with the translocation complex is via the hydrophobic box in the **signal sequence**; removal of this box from the gene sequence of an ER protein (or replacement of one of its amino acids with a charged one) prevents targeting to the ER; and addition of that box to a cytosolic protein leads to translocation into the ER. Translocation is accompanied by removal of the signal sequence by the enzyme **signal peptidase**, which is located inside the ER.

The key component of the peptide translocation complex is a transmembrane protein, **Sec61p**, the product of the mammalian homologue to the yeast *sec61* gene. This protein forms a channel through the membrane. The growing polypeptide is

threaded from the ribosome through this channel and reaches the lumen of the ER without exposure to the cytosol. Also required is the **Translocating chain associated membrane protein (TRAM)**, which binds the signal sequence once it has been handed over by SRP. Two further proteins in this complex, **Sec61β** and **Sec61γ**, bind the ribosome. These proteins are sufficient to reconstitute protein translocation into liposomes *in vitro*. Electron microscopy indicates that the translocon is 50–60 Å high and has a diameter of 85 Å, and the channel is about 20 Å in diameter. In the absence of ribosomes, the cytosolic entrance of the channel is closed by a loop of Sec61p, which gets out of the way after ribosome binding. About 40 % of proteins that go through the Sec61 complex do so independently of SRP, some because the targeting sequence is at the C-terminal end of the protein, and others because they are too short to bind to SRP before they leave the ribosome. The pathways used by these proteins, however, are not as well characterised [2, 12].

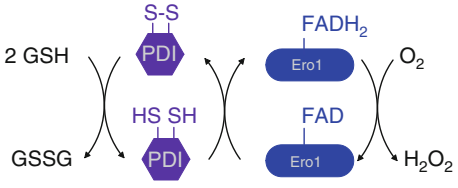
In the ER lumen, the nascent protein is bound by an ER-specific homologue of the 70 kDa heat shock protein family, **78 kDa glucose-regulated protein (Grp78)**. This maintains the nascent protein in a folding-competent state until translocation is complete.

16.5 Folding and Quality Control of Membrane Proteins

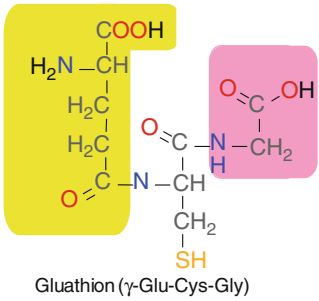
16.5.1 Posttranslational Modification in the ER

16.5.1.1 Disulphide Bond Formation

The cytosol is a **reducing** environment, and Cys-residues carry thiol (−SH) groups. The lumen of the ER, on the other hand, corresponds to the extracellular space, and is **oxydising**. Some cysteine are oxydised to cystine. Disulphide formation does not utilise oxygen directly, as thiol oxydation by oxygen is a slow process. Instead, redox buffers such as glutathione (γ-Glu-Cys-Gly; see Fig. 16.12) and also Protein disulphide isomerase (PDI) transfer electrons to the flavo-enzyme Ero1. From there, electrons are transferred onto molecular oxygen, forming H_2O_2 [35]:



Substrate proteins can also donate electrons to PDI, just as glutathione. In the intermembrane space of mitochondria and in the bacterial periplasm the same process occurs, but the electrons are transferred to cytochrome c by Erv1 and DsbB, respectively. Thus, in these cases disulphide bonds can be formed without producing reactive oxygen species (ROS). Interestingly, Ero1, Erv1, and DsbB are



Gluathion (γ-Glu-Cys-Gly)

Fig. 16.12 Glutathione is an important redox-coupler in cells

not evolutionarily related, even though they serve the same purpose. This is an example for convergent evolution (reinventing the wheel).

Vaccinia virus contains an Erv-like protein which presumably allows disulphide formation in coat proteins that fold in the cytosol.

Disulphide formation is important for correct **tertiary structure** of proteins. This means that neither are all thiols converted nor is it unimportant which –SH groups are involved in the formation of a disulphide bond. One way of ensuring that correct thiols enter disulphide formation is to oxydise Cys-residues in the N-terminal part of the protein before more C-terminal Cys-residues have been incorporated. Immunoglobulins are an example for this method.

In other proteins disulphide bonds are formed across the molecule (e.g., insulin, Cys 1-4, 2-6, and 3-5). Wrongly formed disulphide bonds can be corrected by **Protein disulphide isomerase (PDI)**, an abundant enzyme in the ER of secretory tissues.

Inclusion Bodies

Bacteria do not have an oxydising compartment like the eukaryotic ER, even though the periplasm is oxydising and contains PDIs (encoded by Dsb). If one tries to express human proteins in bacteria for the production of **recombinant therapeutics** problems may occur. Proteins that require disulphide bond formation often do not fold correctly; instead they aggregate in the form of **inclusion bodies**. In some cases it is possible to isolate these inclusion bodies, completely unfold the proteins with urea or similar agents, and refold *in vitro* under proper oxydising conditions by slow removal of the denaturant. However, in many cases such proteins will be produced in cultured animal cell lines to avoid this complication.

16.5.1.2 Chaperones of the ER

We have seen how certain proteins, molecular chaperones, and chaperonins, aid in the folding of cytosolic proteins (see Chap. 15 on page 343) and how they can in addition also help with import of proteins into compartments such as mitochondria (see Sect. 16.3 on page 373). Given the importance of the ER for protein synthesis it is hardly surprising that this compartment is particularly rich in chaperones.

One of them is the **78 kDa glucose-regulated protein (Grp78)^{BiP}**, a homologue of the cytosolic Hsp70/Hsc70 proteins (see Fig. 15.2 on page 347). Grp78^{BiP} receives nascent proteins when they come out of the translocon and maintains them in a folding-competent state. ERdj serves as J-protein for BiP. Further chaperons required for protein folding in the ER are discussed below.

16.5.1.3 Peptidyl Prolyl *cis*-trans Isomerase

Proline is an unusual amino acid because its α-amino group is part of a ring structure. In Pro the *cis*-conformation occurs more frequently than in other amino acids (see Fig. 2.5 on page 19). Proper protein folding requires the conformation of the Pro-residues to be correct; this is ensured by **Peptidyl prolyl *cis/trans*-isomerase (PPI)**. This reaction is often the rate-limiting step in protein folding in the ER. PPIs are often specific for a particular protein.

> Folding of the major histocompatibility complex (MHC) requires a specific PPI, which is one of the targets for the immunostatic drug **Cyclosporin A**, which is used to prevent organ rejection in transplant patients. Hsp70s are also targets for this drug.

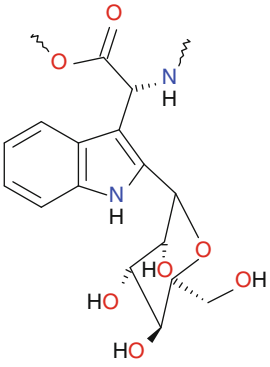
16.5.2 Glycosylation

Most secreted and transmembrane proteins contain sugar trees, which

- Aid in protein folding. For this reason substances interfering with glycosylation (desoxynojirimycin, tunicamycin) may be used to **treat viral infections.**
- Target proteins to the correct subcellular compartment.
- Serve as receptors for cell–cell recognition and as **immunological determinants** and blood group antigens.
- Form the binding sites for bacteria and virus.

Nuclear and cytosolic proteins, on the other hand, are not N-glycosylated; rarely they have a single GlcNAc-residue on Ser or Thr. Inherited disorders of the glycosylation pathway lead to severe disease [8, 9, 15].

The sugar trees can be added to the OH-groups of Ser, Thr, and hydroxylysine (**O-linked oligosaccharides**) or to the amino group of Asn (**N-linked oligosaccharides**). Whereas O-linked oligosaccharides are generally short, N-linked oligosaccharides are quite complex. The same protein may have different N-linked oligosaccharides depending on the tissue where it was produced, and even the protein molecules from one cell type can be heterogeneous. Because some of the sugar groups (**sialic acids**) contain negative charges, the number of negative charges in a glycoprotein may vary, making their separation in the laboratory difficult (see Sect. 3.1.4 on page 71). The reasons for these differences are not understood, but subject to intensive study.



C-linked glycosylation is a single mannose linked to the first Trp in the sequence W-X-X-W or W-S/T-X-C [10]. This is a rare modification; we know nothing about its function.

16.5.2.1 O-Linked Oligosaccharides

O-linked oligosaccharides are produced in the GOLGI-apparatus by adding one sugar at a time to the growing tree on the protein. There is no consensus-sequence for O-glycosylation. The sugars are first activated by the addition of nucleotides (UDP for **galactose** (Gal) and **N-acetylgalactosamine (GalNAc)**, CMP for **N-acetylneuraminic acid** (NANA) and sialic acid (Fig. 16.13). Synthesis of these activated sugars occurs in the cytosol; they are transported into the GOLGI-apparatus by antiporters, which exchange them for the free nucleotides left after formation of oligosaccharides.

Transfer of the sugar residues is catalysed by membrane-bound glycosyltransferases, which are specific both for acceptor OH-group and donor sugar-nucleotide.

In most cases, synthesis of O-linked oligosaccharides starts with the transfer of **GalNAc** (from UDP-GalNAc) to Ser or Thr OH-groups by GalNAc-transferase bound to the luminal leaflet of the membrane of the **rough ER** or ***cis*-GOLGI network**. After transport of the nascent protein to the ***trans*-GOLGI network** **Gal** (from UDP-Gal) (in $\beta 1 \rightarrow 3$-linkage) and two **NANA**-residues (from CMP-NANA) are added in sequence.

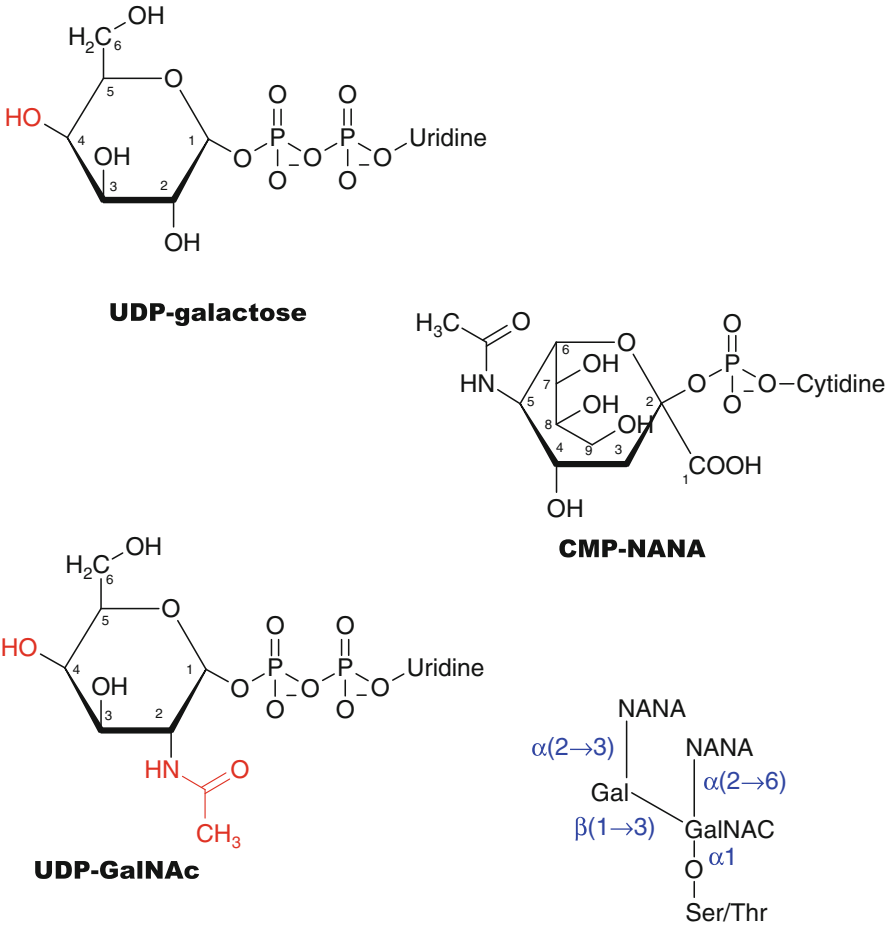


Fig. 16.13 Activated intermediates for the formation of O-linked sugars in the GOLGI-complex

Transfusion

Not only proteins, but also lipids are O-glycosylated. The **AB0-blood group antigens** are oligosaccharides linked to glycoproteins and glycolipids of erythrocytes (see Fig. 16.14). The **0-antigen** has the sequence −Glc−Gal−GlcNAc−Gal−Fuc, which can be synthesised by all humans. People with blood **group A** produce a **GalNAc-transferase** that transfers a GalNAc-residue to the second Gal; people with blood **group B** produce a galactose-transferase instead. Those with blood **group AB** have both these enzymes and hence produce both the A- and the B-antigen; those with blood

(continued)

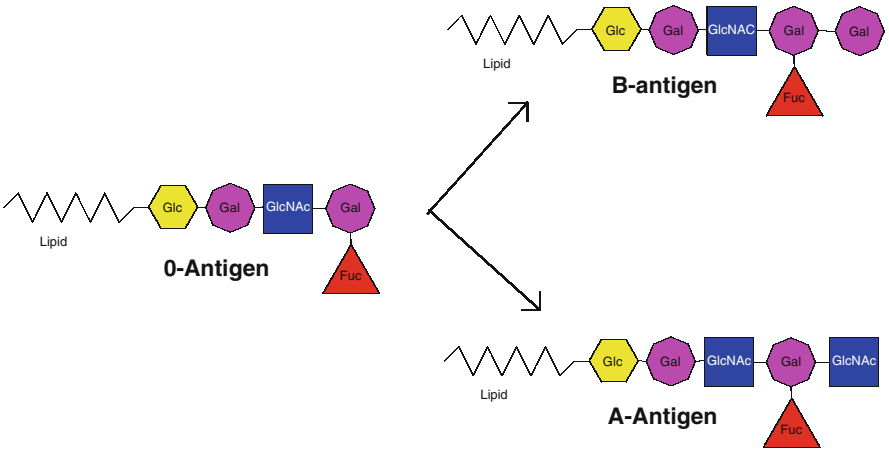


Fig. 16.14 The AB0-blood group antigens are oligosaccharides bound to plasma membrane proteins and lipids. For details see text

group 0 produce neither of these transferases and hence have the O-antigen only. The GalNAc- and Gal-transferases responsible for synthesis of A- and B-antigen, respectively, are very similar proteins which differ in only 3 amino acids.

Intestinal bacteria (*E. coli*) also produce these antigens, hence each type of bacteria can grow only in humans with matching blood groups. This exposure to the common bacteria leads to immunisation to foreign blood group antigens even without exposure to foreign blood. Therefore a severe immune reaction is produced if a patient is transfused blood from a nonmatching donor, even the first time.

Note that because donors with blood group O can produce neither the A- nor the B-antigens and because the O-antigen is present in all people, these donors can donate erythrocytes to patients of any blood group (**universal donor**). On the other hand, patients with blood group AB do not produce antibodies against O-, A-, or B-antigen and can receive erythrocytes from donors of any blood group (**universal recipient**). This picture is made more complicated by the existence of several other antigens (most notably the rhesus-factor), which are caused by variations in the amino acid sequence of proteins in the erythrocyte membrane.

16.5.2.2 N-Linked Oligosaccharides

N-linked oligosaccharides are transferred to proteins as a preformed block and then modified (see Fig. 16.15) [1]. The frequency and complexity of glycosylation increases from prokaryotes over single-cell eukaryotes to multicellular organisms. In prokaryotes glycosylation tends to occur after protein folding, in eukaryotes before. Glycosylation in bacteria, archaea, plants, and various animals have been reviewed and compared in [7, 11, 22, 26, 33]. We focus on vertebrates.

On the cytosolic side of the ER membranes nucleotide-activated sugar residues are added to the long-chain alcohol **dolichol** via a pyrophosphate linkage. Dolichol is a polyisoprene with 75–95 carbon atoms, which is enough to span the membrane four or five times. Hence dolichol is anchored very firmly in the lipid bilayer.

A special **flippase** transports this initial tree to the luminal side of the membrane, where further sugar residues (activated by linkage to dolichol) are added. These dolichol-intermediates too are produced on the cytosolic leaflet of the membrane, and flipped to the luminal side by a specialised transporter.

Transfer of the resulting **oligosaccharide** $(Glc)_3(Man)_9(GlcNAc)_2$-tree from dolichol pyrophosphate to Asn-residues of a nascent protein occurs during its entry into the ER. The receiving Asn must occur in an Asn-X-Thr sequence; the Thr may be replaced sometimes with Ser and rarely with Cys. X can be any amino acid except Pro. Only about two thirds of the potential glycosylation sites are used *in vivo*; the reason is unclear. The **oligosaccharyltransferase** is an multisubunit integral membrane protein which binds to the ribosome on the cytosolic side of the ER membrane and to its substrates on the luminal side. Thus the transferase is held near the nascent protein during its transfer into the ER.

The sugar tree is then trimmed back by glucosidases to $(Man)_8(GlcNAc)_2$- (**high mannose type**) by specific glucosidases. Glucose residue N in Fig. 16.15 is removed once the sugar tree has been transferred from dolichol to protein. This prevents rebinding to the oligosaccharyltransferase, and promotes binding to **malectin**, a protein that retains incorrectly folded proteins in the ER for folding or destruction. Once a protein has passed this checkpoint, glucose residue M is removed, so the protein can bind to calreticulin (luminal chaperone) and calnexin (membrane-bound chaperone). These proteins also bind prolyl-isomerase and PDI; together these proteins help the substrate achieve proper folding. Then glucose L is removed, freeing the substrate from the chaperones. Just to make sure, UDP-Glc:glycoprotein glucosyltransferase binds to the substrate and adds the L glucose back if it finds solvent-exposed hydrophobic patches. The substrate then returns to the chaperons for further folding. This cycle can repeat several times.

The protein is then exported to the *cis*-GOLGI-apparatus, (see Fig. 16.16) where further trimming to $(Man)_5(GlcNAc)_2$- (**core oligosaccharide**) occurs. Again, quality control is performed at this stage, and incorrectly folded substrates are sent back to the ER instead by a receptor that recognises the presence of the Man residues. If everything went well, the substrate is transported to *median*- and *trans*-GOLGI for further trimming and addition of different sugar residues to the final **complex** glycoprotein.

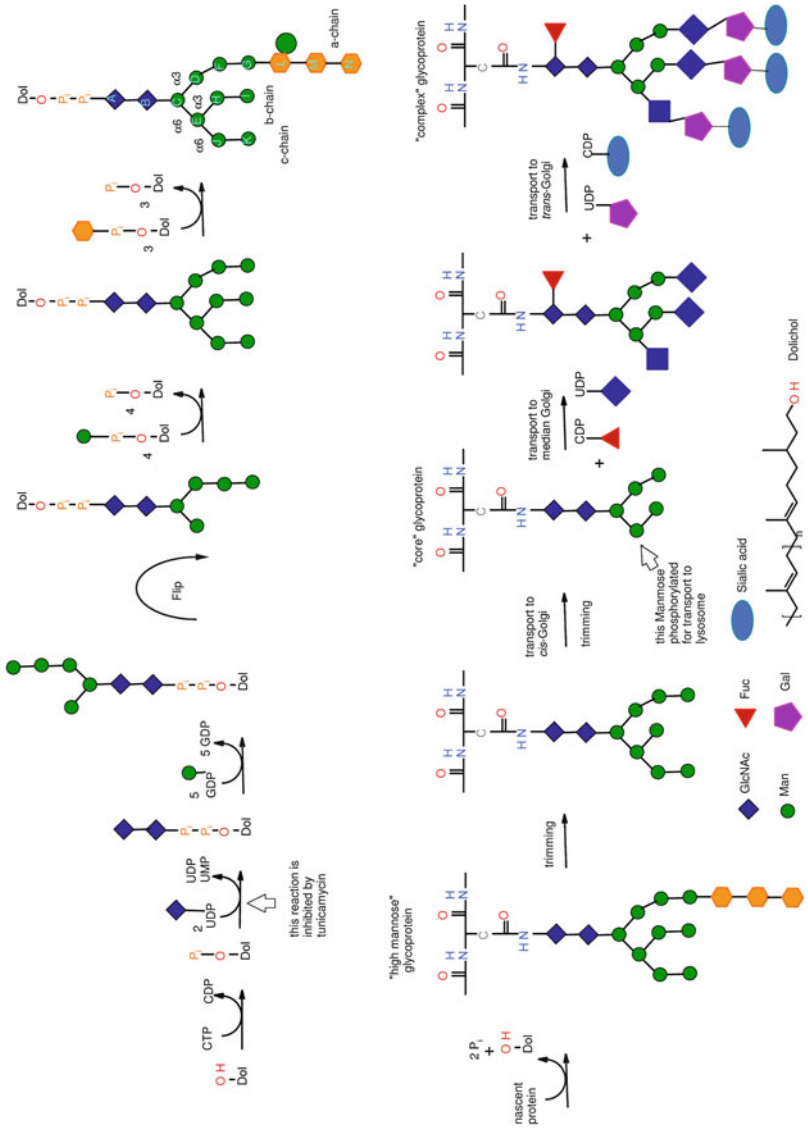


Fig. 16.15 Synthesis of N-linked sugar trees in the ER and GOLGI-apparatus. Nomenclature for sugar residues is marked top right, this nomenclature is referred to in the text

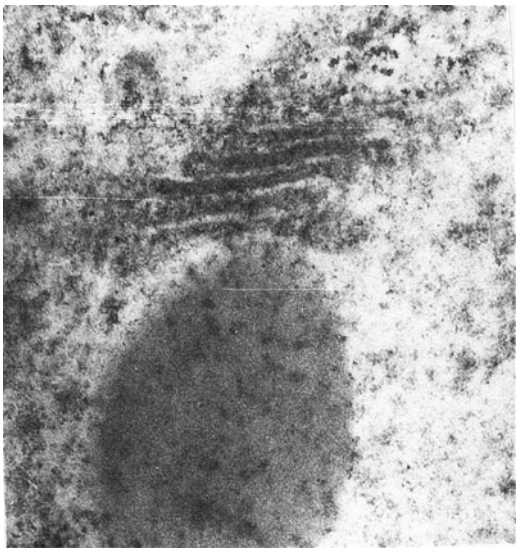


Fig. 16.16 GOLGI apparatus in a tomato fruit (magnification 30 000 times)

16.5.3 Protein Quality Control in the ER

Proteins that cannot be folded correctly (and orphan subunits whose partner cannot be folded) need to be destroyed, because their accumulation in the ER would eventually destroy the cell. This is called **ER associated protein degradation (ERAD)** [6, 20, 24, 25]. The history of research on ERAD has been reviewed in [21]. Removal of the mannose residues "I" and "K" on the b- and c-chains of the Asp-linked sugar tree (see Fig. 16.15) by a highly conserved, ER-specific mannosidase I specific for unfolded proteins controls entry into the ERAD pathway for soluble proteins. Removal of residue "G" prevents readdition of Glc and hence further attempts of refolding by caltreticulin/calnexin.

Misfolded membrane proteins bind to lectins **OS-9** and **XTP-3B** and are transported back into the cytosol through the **Sec61**-translocon. During this **retrograde transport** membrane-bound E3 **ubiquitin**-ligases (see page 57) mark these proteins for destruction by the 26 S**-proteasome**. There are different E3-ligases for:

ERAD-C substrates have a defect in the cytosolic loop
ERAD-L substrates have a defect in the luminal domain
ERAD-M substrates have a defect in the transmembrane domain

Some proteins are partially degraded by ER-resident proteases before they undergo retrograde transport. Presumably, they would not fit through the Sec61 pore in a partially folded state. Nonglycosylated proteins that are not correctly folded are recognised by $Grp78^{BiP}$ and also handed over to ERAD.

Apart from Sec61, other transmembrane proteins may also transport misfolded proteins back from the ER to the cytosol, but they remain to be characterised. Indeed, some authors question that Sec61 is responsible for retrotransport at all

[25]. There also appears to be a proteasome-independent pathway for proteolysis, about which little is known yet. Protein aggregates are destroyed by **autophagy**.

Pathogens and the ERAD-Pathway
Some virus can use the ERAD system to prevent MHC-1 display of virus antigens on the cell surface (see page 259). A/B-toxins can enter the cytosol abusing the ERAD pathway (see page 408). Some polyoma virus are routed to the ER after endocytosis. In the ER the capsid proteins change conformation (e.g., after interaction with PDIs). The capsids are then retrotransported to the cytosol and bind to the nuclear pores, where they release their DNA into the nucleus.

ERAD in Genetic Disease
Some mutations lead to the inability of a protein to attain its native conformation. **Cystic fibrosis** (see page 450) is an example for a disease caused by this mechanism. But lack of essential proteins is not the only possible consequence of such mutations: accumulation of misfolded proteins in the ER of secretory cells can cause serious damage. An example is α_1-antitrypsin, secreted mainly by the liver, but also by macrophages. Lack of this protein results in uncontrolled hydrolysis of elastin and is a cause of lung tissue destruction (emphysema). Accumulation of unprocessed antitrypsin results in the formation of a para-crystalline precipitate in the ER of the liver.

Regulation of Metabolism by ERAD

Some enzymes are subject to feedback control by ERAD. For example, HMG-CoA reductase and squalene monooxygenase are subjected to ERAD when [24,25]-dihydrolanosterol (an intermediate of cholesterol synthesis) exceeds certain levels. Then HMG-CoA reductase binds to Insig-1 or -2, which targets the enzyme to an E3-ligase [31]. Insig-1 in turn is destroyed by the ERAD pathway when [cholesterol] is low.

Some soluble cytosolic proteins also appear to be degraded by the ERAD pathway, or at least some part of it.

16.5.3.1 The Unfolded Protein Response

Correct folding of proteins is vital to the cell; there is extensive quality control on proteins before they leave the ER. Most proteins produced do not pass this check and are degraded. Experimentally, this can be activated by growing cells in the

presence of DTT, which prevents disulphide bond formation. Overwhelming the ERAD response leads to **ER stress** [36].

Bound chaperones (Grp78 and calnexin) are one way for the quality control machine to recognise misfolded proteins. Grp78, PPI, and PDI expression are upregulated in cells experiencing stress conditions (e.g., limited glucose supply, hence glucose regulated protein Grp). This is called **unfolded protein response** [34] and is caused by a lowered concentration of free Grp78, as more and more of this protein becomes bound to misfolded substrates. This causes Grp78 to dissociate from its receptor **Atf6** which is then allowed to move into the GOLGI-apparatus, where it is split by two GOLGI-resident endoproteases, S1P and S2P. The proteolytic peptides increase the transcription of ER chaperones.

If this does not correct the situation, Grp78 dissociates from **Ire1**, a protein found in the inner nuclear membrane (which is continuous with the ER). Ire1 looks a little like MHC-1 with a protein binding site made of a β-sheet with two α-helical sidewalls. Unfolded proteins can bind to that site. Ire1 then oligomerises, autophosphorylates, and splices HAC1 pre-mRNA to its mature form. **Hac1** activates the transcription of mRNA encoding not only for ER chaperones, but also for enzymes involved in the degradation of terminally misfolded (glyco-)proteins.

The third step of unfolded protein response is dissociation of Grp78 from **Perk**, which then phosphorylates the **elongation factor 2α (eIF2α)**, resulting in a stop of translation. This should give the cell time to deal with the problem, before it gets overwhelmed with unfolded protein. Continued activation of Perk finally results in **apoptosis** (cell suicide), to protect the rest of the organism.

In virus-infected cells, large amounts of glycoproteins are produced in a very short time. Because some of these are unavoidably misfolded, the ERAD system can get overwhelmed. This seems to be an important part of the pathology caused by virus [13]. Apoptosis due to the unfolded protein response is important in protein folding diseases (see Sect. 10.2 on page 206). Cancer treatment may make use of this pathway [4] to shift the balance between cell death and survival.

16.6 Exercises

16.6.1 Problems

16.1. A 16-year-old female high-school student visits her family doctor because of a missed period. She admits having had “one or two” boyfriends, who “almost always used condoms”. She is found pregnant but misses most of her appointments during pregnancy. Her blood type is AB cde/cde. At 32 weeks gestation, she delivers

a male infant by vaginal delivery under regional anesthesia, 1400 g, APGAR-score 3 after 1 min and 10 after 5 min. His blood group was B cDe/cde. Bilirubin in the infant rose to 22.4 mg/dL, there was jaundice, and a positive COOMBS test. Blood smear showed anisopoikilocytosis with polychromasia; hematocrit was 34 %. The case was managed with 5 exchange transfusions over two days. There were no further complications.

The hemolytic anæmia in this newborn was caused by maternal antibodies against which of the following?

A Galactose on an O-linked oligosaccharide
B N-acetylglucosamine on an O-linked oligosaccharide
C A protein in the erythrocyte membrane
D Cardiolipin
E Lipopolysaccharide

16.2. Which amino acid side-chains bear O-linked carbohydrate trees?

A Threonine
B Tyrosine
C Glutamine
D Asparagine
E Aspartic acid

16.3. Cardiolipin is characteristic for the membrane of

a) erythrocytes.
b) mitochondria.
c) the endoplasmic reticulum.
d) myelin sheets.
e) lysosomes.

16.4. Patients with mutations in pyruvate dehydrogenase (PDH) E1β-subunit present with lactic acidosis, mildly dysmorphic face, and with varying degrees of CNS damage (seizures, spaticity, hypotonia, nystagmus, poor visual contact). PDH oxydatively decarboxylates pyruvate to acetyl-CoA in the mitochondria.

What is the most likely inheritance of the disease?

A autosomal dominant
B autosomal recessive
C X-linked dominant
D X-linked recessive
E mitochondrial

16.6.2 Solutions

16.1 The mother is Rh-negative, whereas the child is heterozygote for the Rh D-antigen; after the ABO-blood group antigens this is the most significant for incompatibility. Unlike the ABO-antigens, which are O-linked oligosaccharides, the ≈ 50 Rh-antigens are proteins. Their function is largely unknown; in the case of the D-antigen the protein shows homology to transmembrane-transporters and is assumed (!?) to transport ammonia. Maternal IgG cross the placenta; antibodies against the fetus' erythrocytes lead to hemolysis and increased bilirubin (both conjugated and unconjugated). Note that sometimes maternal anti-D antibodies can already be formed during the first pregnancy. This can be prevented by RhoGAM injections during pregnancy to catch any fetal D that might have passed into the maternal blood stream. This is one of the things done during routine exams. In rare instances, the mother is affected by a fetus with hemolysis (BALLANTYNE or mirror syndrome).

16.2 O-linked carbohydrates occur on Ser and Thr. Asn bears N-, not O-linked sugar trees.

16.3 Cardiolipin occurs in the mitochondrial membrane, specifically the inner leaflet of the inner membrane. Because heart muscle cells contain many mitochondria cardiolipin was first isolated from heart tissue, hence its name.

16.4 Most mitochondrial enzymes are encoded in nucleus; mutations therefore have usually an autosomal recessive inheritance. PDH-E1β happens to be encoded on chromosome 3.

References

1. M. Aebi, N-linked protein glycosylation in the ER. Biochim. Biophys. Acta **1833**(11), 2430–2437 (2013). doi: 10.1016/j.bbamcr.2013.04.001
2. T. Ast, G. Cohen, M. Schuldiner, A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. Cell **152**(5), 1134–1145 (2013). doi: 10.1016/j.cell.2013.02.003
3. J.J. Barrott, T.A.J. Haystead, Hsp90, an unlikely ally in the war on cancer. FEBS J. **280**(6), 1381–1396 (2013). doi: 10.1111/febs.12147
4. J.A. Boelens, S. Lust, F. Offner, M.E. Bracke, B.W. Vanhoecke, The endoplasmic reticulum: A target for new anticancer drugs. In Vivo **21**(2), 215–226 (2007). URL http://iv.iiarjournals.org/content/21/2/215.full.pdf+html
5. E. Buxbaum, *Biophysical Chemistry of Proteins: An Introduction to Laboratory Methods* (Springer, New York, 2011). ISBN 978-1-4419-7250-7
6. C.M. Cabral, Y. Liu, R.N. Sifers, Dissecting glycoprotein quality control in the secretory pathway. Trends Bichem. Sci. **26**, 619–624 (2001). doi: 10.1016/S0968-0004(01)01942-9
7. D. Calo, L. Kaminski, J. Eichler, Protein glycosylation in archaea: Sweet and extreme. Glycobiology **20**(9), 1065–1076 (2010). doi: 10.1093/glycob/cwq055

8. B. Cylwik, K. Lipartowska, L. Chrostek, E. Gruszewska, Congenital disorders of glycosylation. Part I. Defects of protein N-glycosylation. Acta Biochim. Pol. **60**(2), 151–161 (2013a). URL http://www.actabp.pl/pdf/3_2013/361.pdf
9. B. Cylwik, K. Lipartowska, L. Chrostek, E. Gruszewska, Congenital disorders of glycosylation. Part II. Defects of protein O-glycosylation. Acta Biochim. Pol. **60**(3), 361–368 (2013b). URL http://www.actabp.pl/pdf/3_2013/361.pdf
10. T. de Beer, J.F.G. Vliegenthart, A.S. Loeffler, J. Hofsteenge, The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase u_s is α-mannopyranose. Biochemistry **34**(37), 11785–11789 (1995). doi: 10.1021/bi00037a016
11. A. Dell, A. Galadari, F. Sastre, P. Hitchen, Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. Int. J. Microbiol. **2010**, Article ID 148178 (2010). doi: 10.1155/2010/148178
12. V. Denic, A portrait of the GET pathway as a surprisingly complicated young man. Trends Biochem. Sci. **37**(10), 411–417 (2012). doi: 10.1016/j.tibs.2012.07.004
13. J. Dudley, H. Byun, Y. Gou, A. Zook, M. Lozano, ERAD and how viruses exploit it. Frontiers Microbiol. **5**, Art. No. 330 (2014). doi: 10.3389/fmicb.2014.00330
14. T. Endo, K. Yamano, Transport of proteins across or into the mitochondrial outer membrane. Biochim. Biophys. Acta **1803**(6), 706–714 (2010). doi: 10.1016/j.bbamcr.2009.11.007
15. S.P. Ferris, V.K. Kodali, R.J. Kaufman, Glycoprotein folding and quality-control mechanisms in protein-folding diseases. Dis. Models Mech. **7**(3), 331–341 (2014). doi: 10.1242/dmm.014589
16. K.S. George, S. Wu, Lipid raft: A floating island of death or survival. Toxicol. Appl. Pharmacol. **259**(3), 311–319 (2012). doi: 10.1016/j.taap.2012.01.007
17. Y. Ishibashi, A. Kohyama-Koganeya, Y. Hirabayashi, New insights on glucosylated lipids: Metabolism and functions. Biochim. Biophys. Acta **1831**(9), 1475–1485 (2013). doi: 10.1016/j.bbalip.2013.06.001
18. E. Klotzsch, G.J. Schütz, A critical survey of methods to detect plasma membrane rafts. Phil. Trans. Roy. Soc. Lond. B: Biol. Sci. **368**(1611), (2012). doi: 10.1098/rstb.2012.0033
19. B. Franz Lang, Mitochondria and the origin of eukaryotes, in ed.by W. Löffelhardt, *Endosymbiosis* (Springer, Vienna, 2014), pp. 3–18. ISBN 978-3-7091-1302-8. doi: 10.1007/978-3-7091-1303-5_1
20. M. Mehnert, T. Sommer, E. Jarosch, ERAD ubiquitin ligases. BioEssays **32**(10), 905–913 (2010). doi: 10.1002/bies.201000046
21. P.G. Needham, J.L. Brodsky, How early studies on secreted and membrane protein quality control gave rise to the ER associated degradation (ERAD) pathway: the early history of ERAD. Biochim. Biophys. Acta **1833**(11), 2447–2457 (2013). doi: 10.1016/j.bbamcr.2013.03.018
22. H. Nothaft, C.M. Szymanski, Bacterial protein N-glycosylation: New perspectives and applications. J. Biol. Chem. **288**(10), 6912–6920 (2013). doi: 10.1074/jbc.R112.417857
23. D.M. Owen, K. Gaus, Imaging lipid domains in cell membranes: the advent of super-resolution fluorescence microscopy. Front. Plant Sci. **4**(503), (2013). doi: 10.3389/fpls.2013.00503
24. A.J. Parodi, Protein glycosylation and its role in protein folding. Annu. Rev. Biochem. **69**, 69–93 (2000). doi: 10.1146/annurev.biochem.69.1.69
25. A. Ruggiano, O. Foresti, P. Carvalho, ER-associated degradation: Protein quality control and beyond. J. Cell Biol. **204**(6), 869–879 (2014). doi: 10.1083/jcb.201312042
26. B. Schiller, A. Hykollari, S. Yan, K. Paschinger, I.B.H. Wilson, Complicated N-linked glycans in simple organisms. Biol. Chem. **393**(8), 661–673 (2012). doi: 10.1515/hsz-2012-0150
27. A. Shah, D. Chen, A.R. Boda, L.J. Foster, M.J. Davis, M.M. Hill, Raftprot: mammalian lipid raft proteome database. Nucleic Acids Res. (2014). doi: 10.1093/nar/gku1131
28. S. Shao, R.S. Hegde, Membrane protein insertion at the endoplasmic reticulum. Annu. Rev. Cell . Dev. Biol. **27**(1), 25–56 (2011). doi: 10.1146/annurev-cellbio-092910-154125
29. K. Simons, E. Ikonen, Functional rafts in cell membranes. Nature **387**, 569–572 (1997). doi: 10.1038/42408
30. S.J. Singer, G.L. Nicolson, The fluid mosaic model of the structure of cell membranes. Science **175**(4023), 720–731 (1972). doi: 10.1126/science.175.4023.720

31. B.-L. Song, N. Sever, R.A. DeBose-Boyd, Gp78, a membrane-anchored ubiquitin ligase, associates with Insig-1 and couples sterol-regulated ubiquitination to degradation of HMG CoA reductase. Molecular Cell **19**(6), 829–840 (2005). doi: 10.1016/j.molcel.2005.08.009
32. S.R. Spindler, R. Li, J.M. Dhahbi, A. Yamakawa, P. Mote, R. Bodmer, K. Ocorr, R.T. Williams, Y. Wang, K.P. Ablao, Statin treatment increases lifespan and improves cardiac health in *Drosophila* by decreasing specific protein prenylation. PLoS ONE **7**(6), e39581 (2012). doi: 10.1371/journal.pone.0039581
33. R. Strasser, Biological significance of complex n-glycans in plants and their impact on plant physiology. Frontiers Plant Sci. **5**(363), (2014). doi: 10.3389/fpls.2014.00363
34. L. Tagliavacca, T. Anelli, C. Fagioli, A. Mezghrani, E. Ruffato, R. Sitia, The making of a professional secretory cell: Architectural and functional changes in ER during B lymphocyte plasma cell differentiation. Biol. Chem. **384**, 1273–1277 (2003). doi: 10.1515/BC.2003.141
35. B.P. Tu, J.S. Weissman, Oxidative protein folding in eukaryotes: mechanisms and consequences. J. Cell Biol. **164**(3), 341–346 (2004). doi: 10.1083/jcb.200311055
36. S. Wakabayashi, H. Yoshida, The essential biology of the endoplasmic reticulum stress response for structural and computational biologists. Comput. Struct. Biotechnol. J. **6**, e201303010 (2013). doi: 10.5936/csbj.201303010
37. R. Zimmermann, S. Eyrisch, M. Ahmad, V. Helms, Protein translocation across the ER membrane. Biochim. Biophys. Acta **1808**(3), 912–924 (2011). doi: 10.1016/j.bbamem.2010.06.015

Chapter 17
Vesicular Transport in Eukaryotic Cells

Abstract Transport from the ER through GOLGI-apparatus to plasma membrane, secretory vesicles, or lysosomes occurs in specialised vesicles. Proteins in the plasma membrane, some with bound ligand, are recycled in clathrin-coated vesicles, which after uncoating fuse with the endosome. From there proteins can return to the plasma membrane or move on to the lysosome for degradation.

Each of the cellular compartments has a special function, and an appropriate set of enzymes to meet them. However, for the cell to survive, there has to be constant transport between the compartments (Fig. 17.1). For example, newly synthesised proteins need to be transported from the ER to the Golgi apparatus for glycosylation, and from there to the target membranes. That needs to be carefully orchestrated, the proteins destined to the Golgi need to be included, but any proteins that belong into the ER need to be excluded from transport. Transport, once initiated, has to go specifically to the Golgi, and not to, say, the endosome.

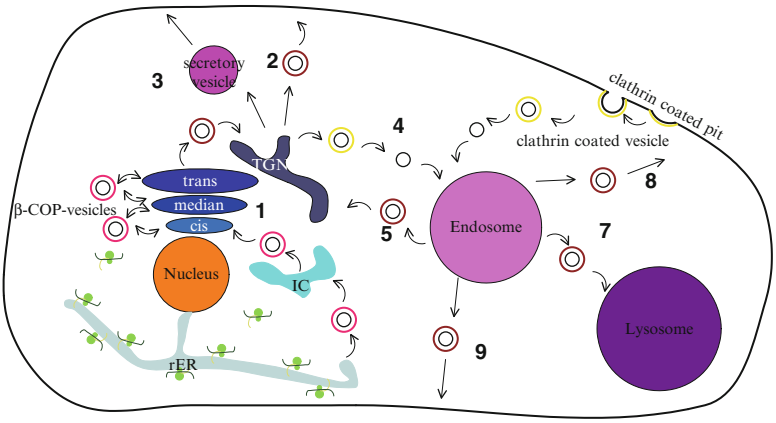
17.1 Two Models for Transport Between Compartments

Transport of proteins between cellular compartments can be achieved either by **transport vesicles** or by **maturation of compartments**. The maturation model is discussed in particular for the GOLGI-apparatus and the endosomal-lysosomal compartments. As I show in a moment, these models need not be mutually exclusive.

17.1.1 Maturation

For the GOLGI-apparatus the **cisternal maturation model** would state that as more and more lipid is produced on the smooth ER, new *cis*-GOLGI-cisternae are formed. As this happens, the old *cis*- becomes the new *median*-, and the old *median* the

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_17



	Pathway	coat protein
1:	rER - intermediate compartment - *cis-*, *median-* and *trans-*GOLGI-apparatus	β-COP
2:	*trans-*GOLGI-apparatus - *trans-*GOLGI-network - cell membrane	unknown
3:	*trans-*GOLGI-network - secretory vesicle, plasma membrane	unknown
4:	*trans-*GOLGI-network - endosome	clathrin
5:	endosome - *trans-*GOLGI-network	unknown
6:	apical cell membrane - endosome	clathrin
7:	endosome - apical cell membrane	unknown
8:	endosome - basolateral cell membrane	unknown
9:	endosome - lysosome	unknown

Fig. 17.1 Vesicular protein transport in eukaryotic cells. For details see text

new *trans*-GOLGI-cisternae. In this model, transport of proteins between GOLGI-cisternae is required only for **retrograde transport** of resident enzymes back into the compartment where they belong, *for example*, *cis*-GOLGI-resident glucosidases from the *median*- back to the *cis*-compartment. **Anterograde transport** of newly synthesised proteins, however, would occur by **bulk flow**.

In the case of the **endosome/lysosome** compartment, similar ideas have been put forward. Endocytic vesicles fuse to form new early endosomes; these mature into late endosomes and finally lysosomes (removal of proteins by transcytosis and recycling, acidification).

17.1.2 Vesicular Transport

Whereas in the maturation model compartments such as the GOLGI-apparatus and the endosome are fairly dynamic structures, which are produced, age, and are recycled, the vesicular transport model sees them as static entities, factories which receive raw materials, process them, and release their final products, without changing much in the process.

Transport vesicles are produced by budding of membranes, caused by assembly of a cage of special proteins on the cytosolic side of the membrane. Two such proteins have been characterised so far: **clathrin** and **β-coat protein (COP)**.

17.1.3 Experimental Evidence

With some very large material, such as **pro-collagen fibers**, vesicular anterograde transport seems impossible because these proteins do not fit into vesicles. This would favour the cisternal maturation model.

With smaller proteins, such as the **vesicular stomatitis virus glycoprotein (VSV-G-Protein)**, observations seem to favour the vesicular transport model; transport through the compartments is faster than could be accounted for by cisternal maturation.

Note, however, that these models are not necessarily mutually exclusive: there may be a vesicular fast track for smaller proteins within the framework of a maturing cisternae.

17.2 Clathrin

Clathrin-coated vesicles are involved in endocytosis and in the transport of newly synthesised lysosomal proteins from the TGN to the endosome. **Clathrin triskelia** consist of three copies each of the heavy and light clathrin proteins. They are tetrahedral in shape, with an apex where all the C-termini of the protein chains meet. From there the proteins extend in three L-shaped arms downward and sideways (see Fig. 17.2). Free clathrin molecules can spontaneously assemble into sheets and baskets in vitro; no ATP is required at this step [9]. In the cell, clathrin coat assembly is organised by a 4-subunit adaptor protein (AP).

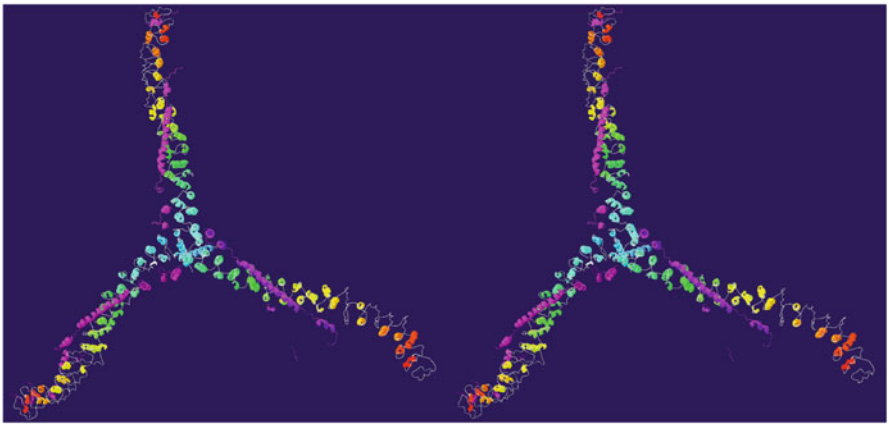


Fig. 17.2 Combined electron microscopic and low-resolution X-ray structure of clathrin, at 7.9 Å (PDB-code 3lvg). The three heavy chains are colour-coded red (N-terminal) to cyan (C-terminal). Those fragments of the light chains that are resolved are drawn in shades of purple

17.2.1 Endocytosis, Membrane Protein Recycling, and Transcytosis

Endocytosis is a process by which cells collect segments of plasma membrane, transmembrane proteins, and ligands bound to these proteins and moves them to the endosome and lysosome. There are several reasons why a cell might want to do this:

- The plasma membrane grows by fusion with exocytotic vesicles. This excess lipid needs to be collected and recycled for vesicle formation. Such recycling is particularly important for **nerve** and **gland** cells.
- Cell surface proteins may have become damaged. They are collected and brought into the lysosomes for degradation to amino acids and sugars, which are reused.
- Cell surface receptors may have ligands bound which need to be brought into the cell. Examples would be iron or cholesterol.
- Certain plasma membrane proteins are stored in pools of intracellular vesicles in a resting cell; after hormonal activation of the cell these proteins are brought back to the plasma membrane. This avoids the delays involved in the resynthesis of proteins, if they were destroyed after each use. Na^+/K^+-ATPase in the kidney would be an example for this type of regulation.

The best-characterised pathway for endocytosis involves clathrin-coated vesicles, and we focus on this below. However, membrane rafts are endocytosed with different proteins (caveolin, flotillin, Arf6, RhoA, GTPase regulator associated with focal adhesion kinase-1 (GRAF1)) [6]. The latter pathways we ignore here.

In addition, macropinocytosis and phagocytosis bring extracellular material into the endosome without coat proteins; the large membrane areas involved contain both rafts and nonraft areas. These processes are actin-driven.

17.2.1.1 Formation of the Clathrin-Coated Pit

Proteins destined for endocytosis accumulate in special structures of the plasma membrane, the clathrin-coated pits, where they may occur in a concentration about 1000 times higher than in the rest of the cell membrane. They bind **adaptor proteins (adaptins)** on the cytosolic side of the membrane, and clathrin assembles on these adaptors.

Three different types of adaptor proteins are known: AP1 is required for TGN to lysosome transport, AP2 for endocytosis, and AP180 for recycling vesicles in brain.

AP1 and -2 consist of 2 large and 2 small subunits (see Fig. 17.3). One of the large subunits is required for interaction with clathrin whereas one of the small subunits interacts with the internalisation signal (Tyr-polar-Arg-hydrophobic) on

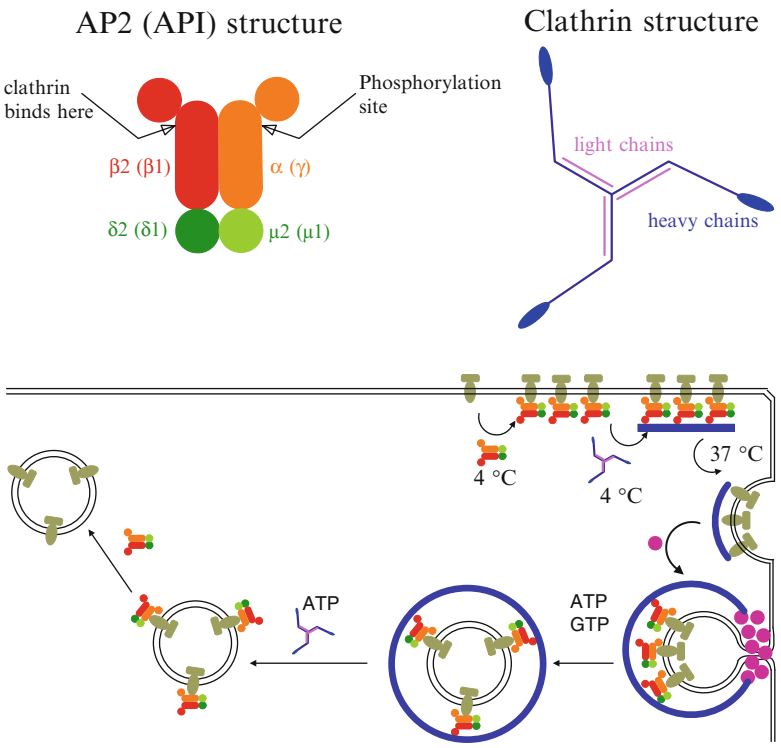


Fig. 17.3 *Top*: Structure of adaptors and clathrin. *Bottom*: Role of clathrin-coated vesicles in endocytosis

the membrane proteins. The second large subunit becomes phosphorylated during the uncoating reaction; dephosphorylation is required before the adaptor becomes available again for coated vesicle formation. For this reason only adaptors isolated from CCV, but not from cytosol, can reassemble into coated pits in in vitro assays. Nothing is known about the function of the remaining small subunit.

The clathrin molecules assemble in the clathrin-coated pits initially as curved sheets, a process that can occur at 4 °C and in the absence of nucleotides. In these sheets clathrin molecules are arranged in a hexagonal lattice. As the sheet grows, this hexagonal lattice is converted into a pentagonal one, increasing curvature to form a globe, which is connected to the plasma membrane by a stalk. This process is blocked by **anti-Hsc70 antibodies** and requires warming to 37 °C, but no nucleotides.

17.2.1.2 Formation of the Clathrin-Coated Vesicle

The stalk is surrounded by the protein **dynamin**, which is responsible for cutting the stalk (GTP dependent), so that free clathrin-coated vesicles are formed [15]. Dynamin binds to clathrin in its GDP form, but upon GDP/GTP exchange moves to the stalk where it self-assembles into helical stacks of rings, containing about 20 molecules of dynamin. Apparently then coordinated hydrolysis of GTP to GDP occurs to cut the stalk (thus freeing the CCV) and to release the dynamin-GDP molecules. In *Drosophila melanogaster* a temperature-sensitive mutant of dynamin (*shibire*^{ts}) (Japanese for "paralysed") is known. At the nonpermissive temperature these flies can no longer recycle lipids in their nerve cells, which leads to immotility. This makes it easy to collect affected animals for investigation. In mammals two isoforms of dynamin are known: dynamin-1 occurs only in neurons whereas dynamin-2 occurs in all cells.

17.2.1.3 Uncoating of Clathrin-Coated Vesicles

Because of their clathrin coats, clathrin-coated vesicles cannot fuse directly with the endosome. And because clathrin coats form spontaneously, energy is required to disassemble them again. This reaction is catalysed by a molecular chaperone **Hsc70** (see page 345), also known as the uncoating ATPase. There is an equilibrium between bound and free clathrin arms. Hsc70·ATP binds to exposed clathrin arms, preventing them from rejoining the coat. This shifts the equilibrium from coat assembly to coat dissolution. Hydrolysis of bound ATP makes this process irreversible, until the ADP produced is exchanged against ATP. It is easy to see that specific ATP/ADP exchange factors at the plasma membrane would stimulate the formation of new clathrin-coated pits, whereas ATP-hydrolysis stimulating factors at the CCV would stimulate the uncoating of CCV. Effective uncoating of CCV requires the presence of **auxilin** in the coat, and auxilin has a DnaJ-like domain. It is presently unclear, however, whether it actually functions as an ATP-hydrolysis stimulating factor.

It has been shown that the formation of CCV requires a **phosphatidyl-inositol-3-phosphate kinase activity** in the plasma membrane to increase the stability of clathrin coats. A corresponding phosphatase (synaptojanin 1) is included in the CCV [10]; its action provides a kind of timed fuse, decreasing the stability of the clathrin coats sometime after vesicle formation. This starts the uncoating process.

One of the proteins endocytosed from the plasma membrane is Na^+/K^+-**ATPase**, which is therefore present in early endosomal membrane. It is recycled to the cell membrane and does not enter late endosomes and lysosomes. This has an important consequence:

The Na^+/K^+-ATPase is electrogenic, pumping 3 Na^+ from the cytoplasm into the exterior of the cell (or the endosomal lumen) for every 2 K^+ pumped into the cytoplasm. Na^+/K^+-ATPase activity therefore leads to a positive charge inside the early endosome, which limits the accumulation of protons—which also bear a positive charge—by a V-type H-ATPase. For this reason the *p*H inside early endosomes is 5.5–6.0. In late endosomes and lysosomes (which do not contain Na^+/K^+-ATPase) *p*H can be decreased to 4.5 by the H-ATPase. The *p*H-gradient between the extracellular environment or TGN, respectively, and the lysosome provides the driving force for differential ligand binding affinities in those compartments.

Endocytosis and exocytosis are fast, high-throughput processes; about 2 % of the cell membrane are covered with coated pits, and each such pit turns into a CCV within about 1 min. About 2500 CCV are formed in a cell every minute; the whole plasma membrane surface is endocytosed (and exocytosed) statistically every 30 min.

17.2.1.4 Specific Materials Taken Up by Clathrin-Coated Vesicles

The Transferrin Cycle

Transferrin is a soluble protein found in serum. Its function is to bind iron and transport it to the cells in which this nutrient is needed, for example, in the bone marrow. These cells express a receptor on their surface, which, under neutral conditions, has a high affinity for the iron-carrying (holo-), but a low affinity for the iron free (apo-) transferrin. This is important, because in healthy people only about one third of the transferrin molecules in serum carry iron. Once the receptor has bound holo-transferrin, it is collected in clathrin-coated pits and is transported to the endosome via clathrin-coated vesicles (Fig. 17.4).

Inside the endosome the *p*H is acidic, around $\approx$ 5.5. Under these conditions, Fe is released from the receptor and becomes available to the cell. The acidic conditions have a second effect, however: the affinity of the transferrin receptor for the apo-transferrin increases. Thus receptor and transferrin stay together while they are brought back to the cell surface by transport vesicles.

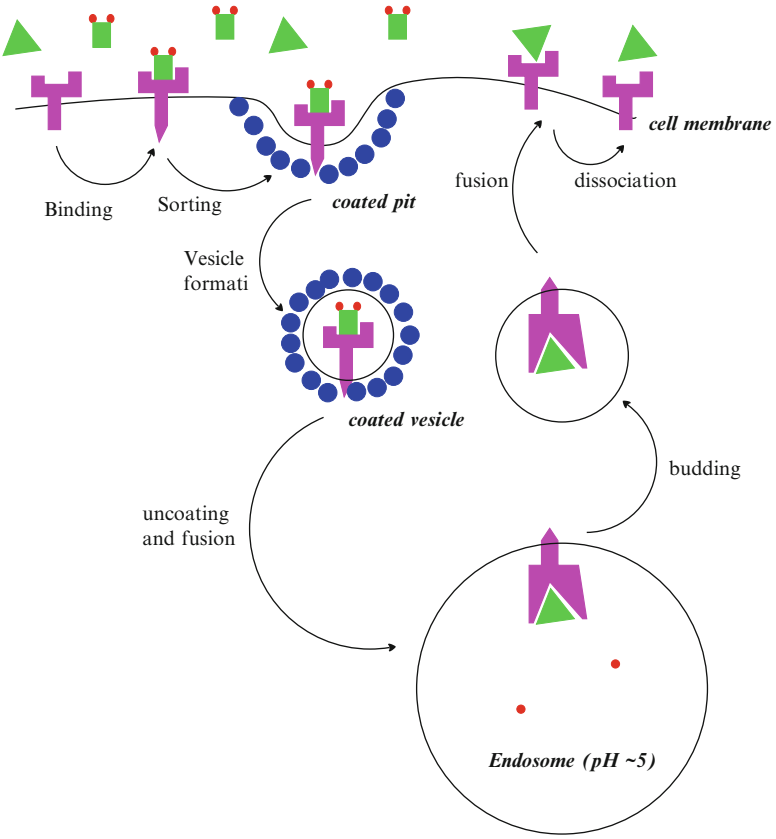


Fig. 17.4 The transferrin cycle: Transferrin (*green*) with bound iron (*red*) binds to the transferrin receptor (*pink*), which is then sorted into clathrin (*blue*)-coated pits. The complex moves into endosomes via clathrin-coated vesicles; the low *p*H inside this organelle releases the Fe and changes the conformation of the transferrin receptor. After return to the cell surface the complex is exposed to a neutral *p*H and the receptor assumes its normal conformation, leading to dissociation of the apo-transferrin

On the cell surface the receptor/apo-transferrin complex is exposed to a neutral *p*H again. As mentioned before, the transferrin receptor has a low affinity for apo-transferrin at neutral *p*H, thus the complex dissociates. The receptor is now available again for binding of holo-transferrin, and the apo-transferrin returns to the liver to bind new iron.

Cholesterol Uptake

Cholesterol is an essential component of the cell membrane in eukaryotes. It is either taken up from food and transported to the liver, or synthesised there. Cholesterol in blood is transported as **Very low-density lipoprotein (VLDL)**, which contain cholesterol, phospholipid, and proteins. They are bound on the surface of cells by a specific Low-density lipoprotein (LDL)-receptor. Receptors with bound VLDL accumulate in clathrin-coated pits and are endocytosed. The lipid portion goes the endosome-lysosome route and is made available to the cell. The receptors are recycled to the cell surface; proteins and some remaining lipids are released back into the bloodstream. Removal of the (low density) lipids results in an increased concentration of (relatively high density) proteins, thus **High density lipoprotein (HDL)** can be distinguished from the VLDL by centrifugation. HDL returns to the liver for reloading with cholesterol. If the cell cannot use all the cholesterol supplied, some of it is returned to the surface too and lipoproteins of intermediate density are formed: **Low-density lipoprotein (LDL)**, "bad cholesterol"). LDL is involved in the formation of arteriosclerotic plaques. HDL can bind free cholesterol and thus prevents arteriosclerosis (hence "good cholesterol").

Patients with mutated LDL-receptors cannot efficiently take up cholesterol by this route. This leads to the accumulation of LDL in the blood, known as **familial hypercholesterolæmia**. The cause can be the inability to produce LDL-receptor, the inability of the receptor to bind LDL, or failure to sort the receptor into clathrin-coated pits.

Downregulation of Receptors

Some cell types require **Epidermal growth factor (EGF)** to initiate cell division. Binding of EGF to its receptor leads to activation of its **tyrosine kinase** activity, signalling the cell to divide (see Fig. 8.1 on page 186). The receptor-hormone complex is internalised by endocytosis; unlike the transferrin or LDL receptors it is not recycled to the cell membrane again, but both hormone and receptor are digested in the lysosomes. Thus an increase in EGF concentration in blood will lead to an increase in receptor internalisation and a decrease in the receptor number in the cell membrane. This "**downregulation**" of receptor density prevents overstimulation of the cells. G-protein coupled receptors also are endocytosed in a manner that appears similar to tyrosine kinase receptors such as EGF-R.

Clathrin-Coated Vesicles in Receptor Internalisation
Internalisation of receptors occurs by clathrin-coated vesicles, but the mechanism appears to be somewhat different than that of the normal endocytotic pathway. Cytosols from different cell types have similar ability to support transferrin receptor endocytosis, but differ in their ability to support the endocytosis of EGF-R. EGF-R endocytosis is not limited by the availability of AP-2.

Recycling Vesicles in Brain
Release of neurotransmitter in the synapses of nerve cells occurs by fusion of transmitter-carrying vesicles with the synaptosomal membrane. That obviously results in an increased lipid content of the præsynaptic membrane. This lipid needs to be recycled, and clathrin-coated vesicles are used to achieve this.

According to one model, the synaptic vesicles are then formed by budding from the endosome, and filled with neurotransmitter *en route* to the plasma membrane. A second model predicts that synaptic vesicles are formed directly from uncoated endocytic vesicles, without going through the endosome [14]. In any case, for exocytosis in nerve cells the GOLGI-apparatus is not directly involved (although of course the proteins required ultimately come from there).

If one looks carefully [3–5, 14], recycling vesicles isolated from brain have some properties which are quite different from those of "normal" clathrin-coated vesicles (which occur in all tissues including brain):

- Analytical ultracentrifugation of CCV from most tissues (placenta is a convenient source) shows a fairly homogeneous population of about 157 S. If the same experiment is performed with brain CCV, two populations with sedimentation constants of 145 and 124 S are seen. The larger vesicles presumably correspond to the "normal" CCV, but the smaller ones are the recycling vesicles.
- Clathrin light chains in brain have a different size from those in other organs, due to alternative splicing.
- Recycling vesicles are less stable than normal CCV against *p*H changes.
- Recycling vesicles are a better substrate for Hsc70, which acts catalytically on brain, but stoichiometrically on liver or placenta-coated vesicles.
- Uncoating of recycling vesicles by Hsc70 can occur even under conditions where ATP hydrolysis is blocked (low temperature or nonhydrolysable

(continued)

ATP analogues; see Fig. 17.5), this is not the case for CCV from placenta or liver.

It would thus appear as if the clathrin-coated recycling vesicles from brain are unusual and adapted to rapid turnover. This makes evolutionary sense, as it reduces the refractory period in nerve cells.

17.2.1.5 Transcytosis

In the examples discussed above receptors were internalised from the apical surface of the cell and returned there after unloading their ligand in the endosome. However, in some cases the proteins are sent to the basolateral surface instead. For example, antibodies in milk bind to specific receptors in the apical (luminal) surface in the intestine of a suckling, are internalised with their receptor in CCV, and transported to early endosomes. There the receptor-antibody complex is packed into transport vesicles which fuse with the basolateral surface of the cell. The neutral *p*H of the extracellular fluid leads to receptor-ligand dissociation and the antibody is released into the suckling's bloodstream.

In the mother the transport occurs by a similar mechanism, but in the opposite direction: from blood antibodies are bound to the basolateral surface of the epithelium in the milk gland, transported to the apical surface and from there into the milk.

Epithelial cells have different early endosomes for their apical and basolateral surfaces; these are linked by a common late endosome. In other cells, transcytosis seems to involve caveolin instead of clathrin.

17.3 β-Coat Protein

17.3.1 Vesicular Transport Between ER and GOLGI *Stacks*

Vesicles budding from the ER *en route* to the Golgi and vesicles budding from the Golgi for anterograde and retrograde transport are coated with **COP** [7]. We distinguish COP-I and COP-II. **COP-I** (see Fig. 17.6) is responsible for **retrograde transport** between GOLGI-cisternae and between *cis*-GOLGI and rER. **COP-II** is responsible for **anterograde transport** from the rER to the *cis*-GOLGI-cisterna. There is a high degree of homology between the equivalent proteins from COP-I and COP-II.

Formation of COP-coated vesicles starts with a complex of seven different coat proteins assembling on the cytoplasmic side of the membrane ([2, 12, 13, 16], see Table 17.1, Figs. 17.7 and 17.8). A selector protein binds to a specific recognition motive on the cargo protein with one arm and the adaptor protein with the other.

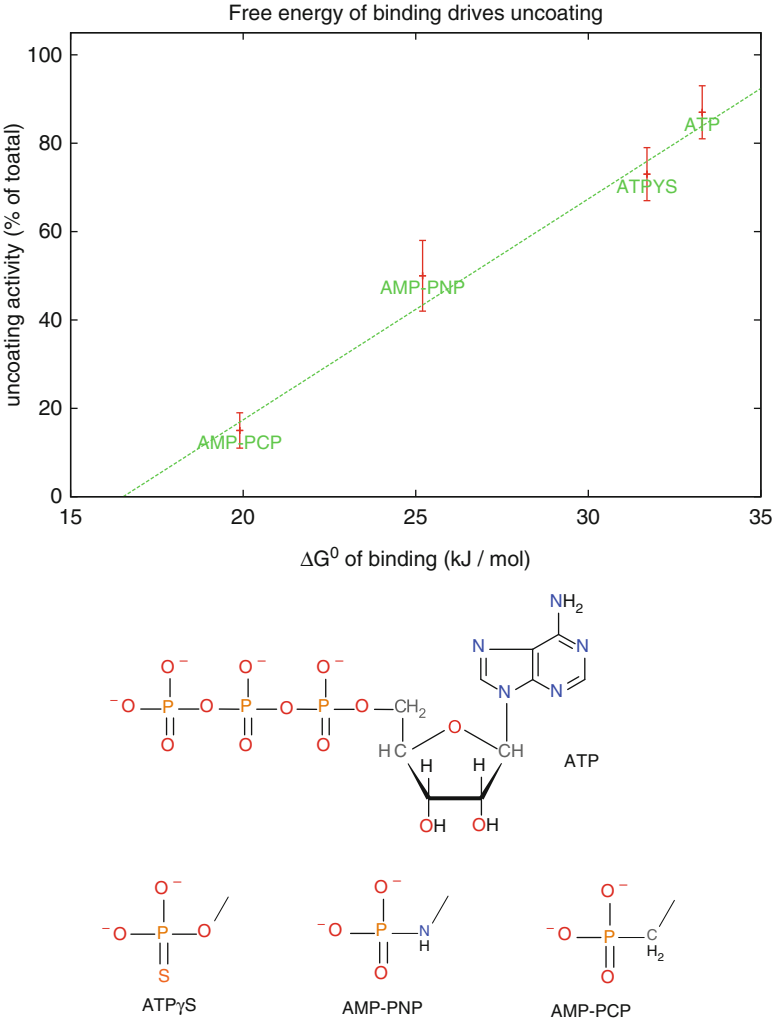


Fig. 17.5 Effectiveness of ATP-derivatives supporting uncoating of brain clathrin-coated vesicles by Hsc70, plotted against the free energy of binding of that analogue to Hsc70. ATPγS, AMP-PNP, and AMP-PCP are nonhydrolysable analogues of ATP with modified γ-phosphate group. The binding angle between β- and γ-phosphate is about 90° in ATP, 101° in AMP-PNP, and 109° in AMP-PCP. Thus the analogues fit less well into an ATP binding site. In ATPγS an oxygen atom is replaced by the bigger sulfur. Because the analogues cannot be hydrolysed, but do support uncoating, the driving force of the uncoating reaction cannot be the free energy of hydrolysis. The linear relationship between energy of nucleotide binding (which can be calculated from the K_D) and the effectiveness of uncoating indicates that it is nucleotide binding, not its hydrolysis, that powers uncoating. The minimum energy required can be estimated from the x-intercept ($\approx$ 16 kJ/mol). Data from [3, 4]

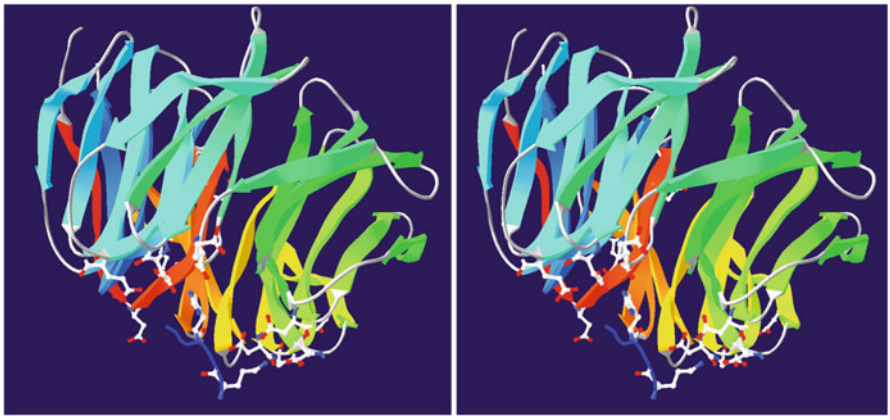


Fig. 17.6 3D-structure of β-COP, (PDB-code 4j82). At the bottom in blue a fragment from Insig2, one of the proteins transported from the GOLGI-apparatus back to the ER in COP-I coated vesicles. Such proteins have a KxKx- or KKxx-motive near the C-terminus; here the second Lys was replaced with His. The two positively charged amino acids point into a negative pocket of the coat protein. Insulin-induced gene 2 protein regulates steroid synthesis by keeping HMG-CoA reductase in the ER, where it is a substrate of the ERAD-pathway

Table 17.1 Comparison of COP-I and COP-II, using yeast nomenclature

Function	COP-I	COP-II
Adaptor	ARF	Sar1
GEF	Gea1/2, Sec7	Sec12
GAP	Gcs1, Glo3	Sec23
GAP stimulator	Sec26, Sec21	Sec31
GAP inhibitor		Sec16
Cargo selection	Erd2, Rer1	Sec24
Cargo motive	KxKx, KKxx, KDEL/HDEL	various
Coat	Sec21-Ret3, $(Sec27\text{-}Ret1)_3$	Sec23-Sec24, $(Sec13\text{-}Sec31)_4$
v-SNARE		Bed1
t-SNARE	Use1	Sed5, Sec22, Bos1
direction	retrograde	anterograde

The adaptor protein is a small GTPase; if the cargo protein is incorrect, the GTPase-activating protein (GAP) stimulates GTP hydrolysis in the adaptor, leading to dissociation of the coat. In its GDP-bound form, the adaptor is a soluble cytosolic protein; in the GTP-bound form it is membrane associated due to exposure of a N-terminal amphophilic α-helix. Hence entry of incorrect cargo into the COP-vesicle is prevented. COP-proteins and their cargo are so close together on the membrane that noncargo proteins are crowded out of the area that will form a new vesicle. These **exit sites** are located in areas of strong curvature (*i.e.*, edges) rather than on membrane sheets. Insertion of GTPase amphipatic tails destabilises the membrane. While the coat is formed, GAP inhibitors on the membrane prevent GTP

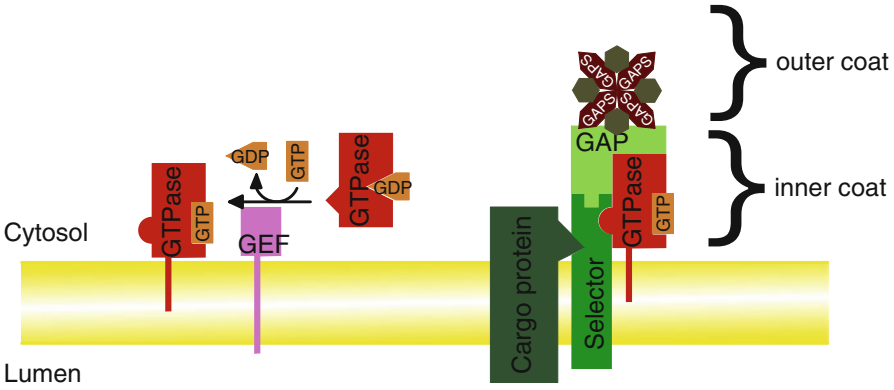


Fig. 17.7 Cartoon of COP-I and COP-II. In COP-I the outer coat is a triprotomer rather than a tetraprotomer. For explanations see text

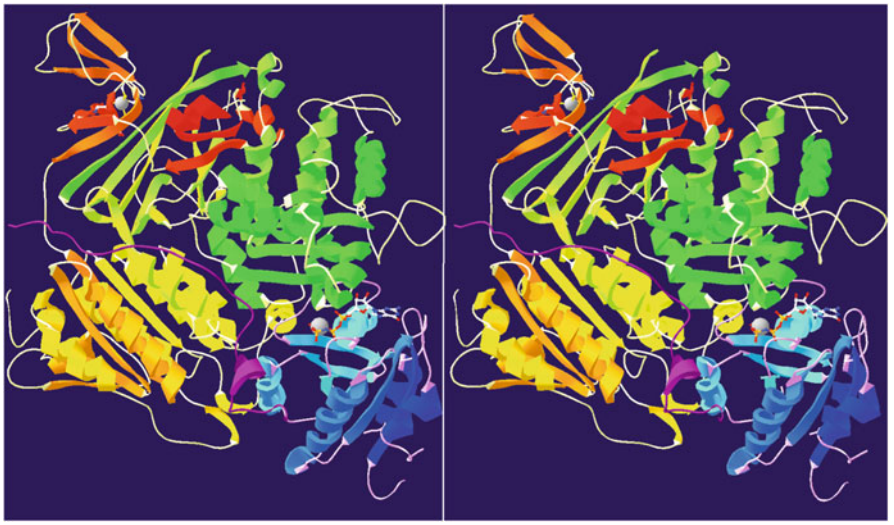


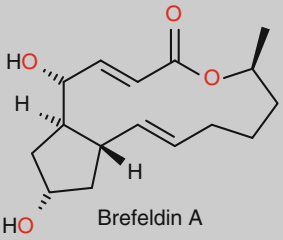
Fig. 17.8 Crystal structure of Sec23 (*red to green*) bound to Sar1 (*cyan to blue*). The purple coil is the Sec23 binding site of Sec31. Sar1 has Mg^{2+}-GMP-PNP bound to it, a nonhydrolysable GTP analogue (oxygen between β- and γ-phosphate replaced by an imino-group (−NH−). Sec23 has Zn bound to it, held by Cys-groups

hydrolysis and hence coat destabilisation. Once a vesicle has left the membrane from which it originated, the GAP and GAP-stimulator (GAPS) trigger GTP hydrolysis in the adaptor and the coat disassembles. This allows complex formation between v-SNARE and t-SNARE, that starts the fusion process with the target membrane.

New *cis*-GOLGI stacks are formed by fusion of COP-II vesicles originating from the ER; they also receive material from COP-I vesicles that recycle material from median and *trans*-GOLGI. As new *cis*-GOLGI stacks form, they push older stacks into the median and *trans*-position. Any proteins that got swept too far out get transported back by the COP-I vesicles. This is part of the maturation process of the older stacks. Thus forward transport within the GOLGI-apparatus occurs mainly by cisternal maturation.

The ER is not only the place of cotranslational membrane insertion and modification of proteins, but also of phospholipid biosynthesis. The lipids are distributed to the other membrane systems by vesicular transport, just as the proteins are.

Brefeldin A Is an Inhibitor of COP-II Transport



Brefeldin A

Brefeldin A is an antibiotic produced by the mold *Penicillium brefeldianum* (Trichocomaceae, STOLK & D.B. SCOTT 1967). It prevents the anterograde transport from GOLGI-apparatus to ER [11] by inhibition of GDP/GTP exchange in Sar1. Brefeldin A is used not only to study vesicular transport itself, but also processes that depend on it such as antigen presentation of virus assembly.

A/B-Toxins

The retrograde pathway of protein transport was hijacked by a group of bacterial and plant toxins, known as **A/B-toxins**. Examples include the **cholera-toxin** (from *Vibrio cholerae*), and **ricin** from the castor bean, *Ricinus communis*. These toxins consist of two protein chains, linked by a disulphide bond. The nontoxic β-subunit binds to the plasma membrane; the entire toxin is endocytosed and transported to the early endosome. From there it is transported back to the ER, because the β-subunit contains the K/HDEL-sequence required for this route. In the ER α- and β-subunit sep-

(continued)

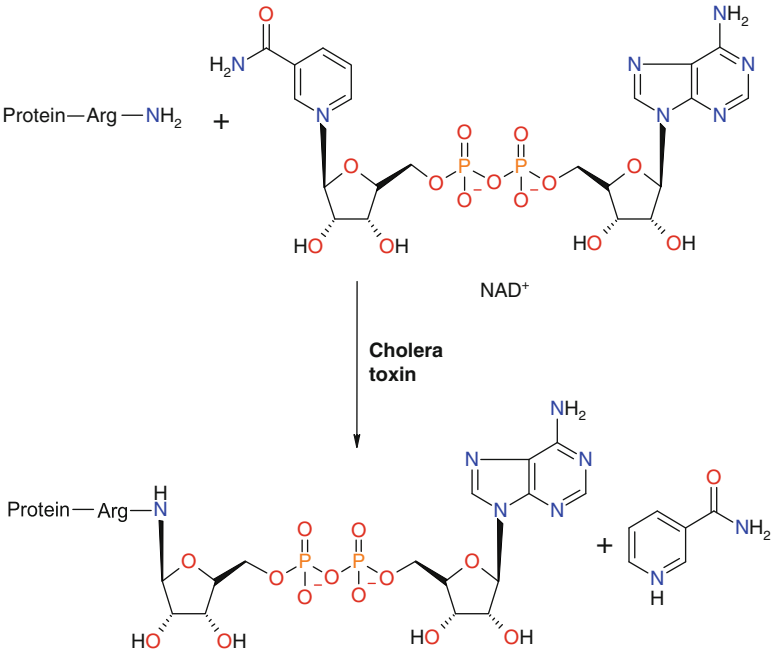


Fig. 17.9 ADP-ribosylation of the α-subunit of G_s by cholera toxin blocks its GTPase function, leading to constant activation

arate, the α-subunit is ubiquitinated and transported into the cytosol using the ERAD-pathway. By an unknown mechanism the α-subunit evades hydrolysis in the proteasome; instead it causes havoc in the cell.

Cholera toxin acts by transferring ADP from NAD^+ to an Arg-residue in the α-subunit of G_s (see Fig. 17.9), blocking its GTPase activity and rendering G_s permanently active. The chronic activation of cAMP-synthesis leads to the secretion of bicarbonate and chloride ions into the intestinal lumen; the resulting passive water efflux into the lumen leads to diarrhoea, which aids in the spread of the pathogen (but can kill the patient from dehydration and mineral loss).

A single molecule of ricin can prevent protein synthesis in a cell by enzymatically depurinating a specific adenosine in the 23 S-rRNA, making it one of the most potent poisons known to man. This high toxicity has kindled the interest of researchers working on anticancer drugs. If the α-subunit could be brought specifically into a cancer cell (e.g., using antibodies against tumor markers as the vehicle), it would act as a "magic bullet". Unfortunately despite intense efforts of many groups, no success has been achieved yet.

17.4 The Specificity of Membrane Fusion

17.4.1 v-SNAREs and t-SNAREs

With all the vesicles budding of the various intracellular membrane systems and fusing with others, **specificity** of the fusion event must be maintained. In other words: each vesicle must carry an **address label** that allows it specifically to fuse with its target membrane, and only with that one.

The first thing that became known about several types of fusion events in the eukaryotic cell was that a cytosolic factor is involved. Finally a small protein, **NEM-sensitive factor (NSF)**, was isolated. If this protein is inactivated with the SH-reagent **N-ethylmaleimide (NEM)**, or if it is removed from cytosol by immunoabsorbtion, fusion no longer works.

It also became known that fusion of synaptic vesicles with the presynaptic membrane could be prevented by incubation with either **tetanus** or **botulinus toxin**, because these act as specific proteases that destroy **synaptobrevin** and **syntaxin**, respectively. Synaptobrevin is a protein in the plasma membrane; syntaxin occurs in synaptic vesicles.

Synaptobrevin and syntaxin are special cases of a group of proteins known as **SNAP receptor (SNARE)**). These always occur in pairs: one on the vesicle membrane (v-SNARE, syntaxin in the case of exocytosis) and the other on the target membrane (t-SNARE, synaptobrevin in the case of exocytosis). Four-helix bundle formation between v-SNARE and its corresponding t-SNARE into a *trans*-SNARE complex (see Fig. 17.10) is believed to account for the specificity of membrane fusion. After fusion, both proteins are in the same membrane as *cis*-SNARE complex. This complex is disassembled by **Soluble NSF-attachment protein (SNAP)** and **NEM-sensitive factor (NSF)**. vSNAREs are enriched into special vesicles and transported back to the donor membrane. SNAPs and NSF-tetrameres can be isolated after solubilisation of the membranes as 20 S complex. ATP hydrolysis in bound NSF is supposed to be involved in the actual fusion event, possibly by changing the conformation of the complex and bringing the membranes closer together.

17.4.2 Rab Proteins

Vesicle formation, movement through the cell, and fusion with target membranes are controlled in a not fully understood manner by another set of 70 small (Ras-like) G-proteins, which are called **Rab** [8] (see Table 17.2). They are attached to the cytosolic side of vesicles by prenyl groups. Rab-proteins act in concert with guanine-nucleotide exchange factors (GEFs), GTPase-activating proteins (GAPs), GDP dissociation inhibitors (GDIs), and GDI displacement factors (GDFs). Rab escort proteins (REPs) bind to the GDP-bound form, Rab effectors—the proteins through which Rab acts—interact with the GTP-bound form.

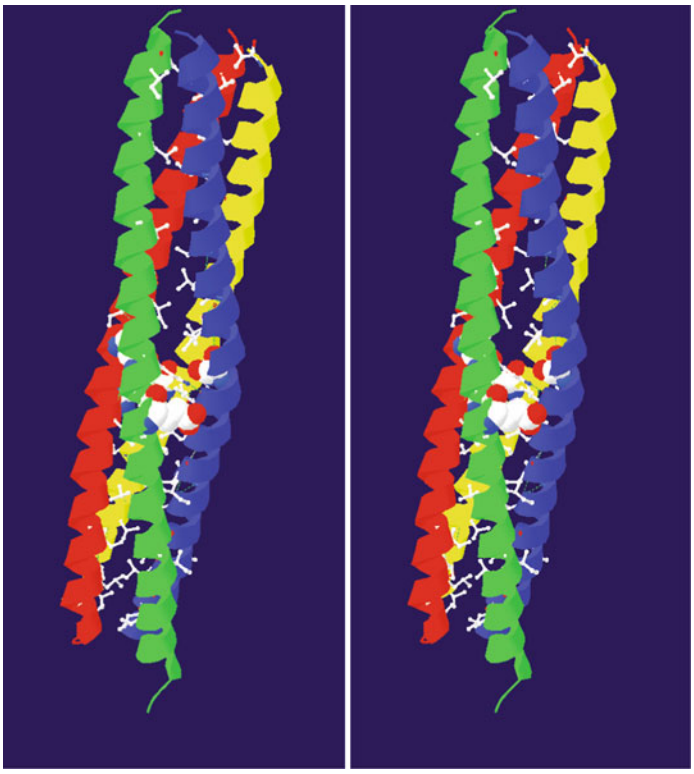


Fig. 17.10 4 helix bundle SNARE core complex, here between syntaxin 7, syntaxin 8, vti1b, and endobrevin/VAMP-8. These SNAREs are required for late endosome fusion (PDB-code 1gl2). The SNARE motive is intrinsically disordered (see Sect. 10.1 on page 203); the helices are present only when v- and t-SNAREs form a complex. They are held together by hydrophobic amino acids (shown as ball and stick) arranged in layers like the rungs of a ladder. In the centre of the bundle is the "hydrophilic layer 0" composed of three Gln and one Arg (shown as VAN DER WAALS). SNAREs with a Gln in this position are called Q-SNAREs; those with an Arg R-SNAREs

17.4.3 How Do Proteins Know Where They Belong?

There are several hypotheses for this question, which again need not be mutually exclusive.

ER-resident proteins are prevented from moving with the **bulk flow** by a special sorting signal (KDEL in mammals, HDEL in yeast) which is also responsible for their transport from the GOLGI-apparatus back to the ER, should they accidentally move to the GOLGI-apparatus. Proteins not bearing this signal will automatically move with the bulk flow from the rER to the GOLGI-apparatus.

Proteins may form large oligomeric complexes with other proteins in the compartment where they belong. These complexes could be too big to be included in transport vesicles. Such aggregation can be demonstrated experimentally: if one

Table 17.2 Important Rab proteins

Type	Found on	Function	PDB code	OMIM
Rab1	ER, GOLGI	Autophagy, ER-GOLGI, and intra-GOLGI transport	3svf	*179508
Rab2	ER, *cis*-GOLGI	ER-GOLGI transport	1z0a	*179509
Rab3A	Secretory and synaptic vesicles	Ca^{2+}-mediated exocytosis, synaptic plasticity	3rab	*179490
Rab4	Early endosomes	Dynein recruitment, recruitment of Rab7 during endosome maturation	1yu9	*179511
Rab5A	Clathrin-coated vesicles, plasma membrane,	Contact to actin, endosome enlargement	1r2q	*179512
Rab5C	Early endosome		1huq	*604037
Rab6	Medial and *trans*-GOLGI	Retrograde transport	1d5c	*179513
Rab7A	Late endosome	Endosomal maturation and movement on microtubules, melanosome position	1t91	*602298, #600882
Rab8	Secretory vesicles (basolateral)	Transport of plasma membrane proteins, cilia organisation	3tnf	*165040
Rab9	Late endosome, TGN	Retrograde transport of Man6-P receptor to GOLGI	1yzl	*300284
Rab11	Recycling endosomes	Neurite formation, recycling of endocytic vesicles	1yzk	*605570
Rab18	Synaptic vesicles, secretory granules,	Secretion in neuroendocrine cells	1x3s	*602207, #614222
Rab23		Negative regulator of Sonic hedgehog signaling	1z22	*606144, #201000
Rab27		Melanosome transport	2f72	
Rab39a	Inflammasome, binds Caspase-1	Pinocytosis, intracellular lysosome positioning		*610917
Sec4	Secretory vesicles	Exocytosis	1g16	

of the sugar transferases from the *median*-GOLGI is fused with a protein bearing an ER retention signal, this enzyme will accumulate in the ER instead of being transported to the GOLGI. But it will also retain other transferases which have not been manipulated, and in such a mutant cell the GOLGI apparatus will be very small or even nonexistent. Formation of such complexes is made easier by the fact that most GOLGI-proteins have a very similar structure, an N-terminal cytoplasmic domain, a single transmembrane domain (where complex formation occurs), and a C-terminal catalytic domain facing the lumen of the cisterna.

Another hypothesis starts with the observation that the concentration of cholesterol in the membrane increases from ER over Golgi to plasma membrane. This leads to an increase in the thickness of the membranes in the same order. The transmembrane domains of Golgi proteins are slightly shorter than their counterparts in plasma membrane proteins. If such a protein were to be brought into the plasma membrane, hydrophilic side-chains from the cytoplasmic or external domains would be drawn into the hydrophobic membrane environment, which is energetically unfavourable.

17.5 Other Vesicular Transport Pathways

17.5.1 Transport of Newly Synthesised Proteins to Their Destination

Proteins destined for the cell surface or for excretion into the extracellular environment are packed into transport vesicles in the *trans*-GOLGI network (TGN). The coat protein on those vesicles has not been identified so far; initial evidence for the involvement of clathrin has turned out to be spurious.

It is important to note that some of the vesicles fuse with the cell membrane in a nonregulated, automatic fashion, whereas others fuse only after a specific stimulus has been received (hormones, digestive enzymes, neurotransmitter). The mechanism for protein sorting into these different classes of vesicles is not clear. However, proteins destined for regulated secretion appear to form aggregates with two proteins (chromogranin B and secretogranin II) under the low *p*H (6.5), and high $[Ca^{2+}]$ conditions of the GOLGI-complex.

17.5.1.1 Protein Secretion

Vesicles for regulated exocytosis shrink up to 200-fold after they have been formed; for this reason they show a dark core in the EM. This core is of irregular shape and can therefore be distinguished from the quasi-crystalline core of peroxisomes. During formation of the mature exocytotic vesicle cleavage of their cargo may occur to convert inactive preproteins to the final form. The signal for fusion between exocytotic vesicles and the cell membrane is usually an increase in (local) $[Ca^{2+}]$.

Vesicles are transported from the Golgi to the cell membrane along microtubuli. They may cover considerable distances, for example, in nerve cells. This transport is inhibited by **Nocodazole**.

17.5.1.2 Plasma Membrane Proteins

Polar cells with distinct apical and basolateral domains sort their exocytosed proteins into different vesicles for the two domains. The signals responsible for sorting are not known. A particularly interesting case is Na^+/K^+-ATPase (normally in the basolateral domain of animal cells) and the homologous H^+/K^+-ATPase (in the apical membrane of stomach). Responsible for their sorting are the noncatalytic β-subunits. Expressing Na^+/K^+-ATPase α- (catalytic) subunit in the absence of the β-subunit leads to protein destruction by the ER-quality control system, as the protein cannot fold. If the Na^+/K^+-ATPase α-subunit is expressed together with H^+/K^+-ATPase β-subunit, the chimaera is found at the apical side of the cell. Similarly, the H^+/K^+-ATPase α-subunit can be directed to the basolateral side by expressing it together with the Na^+/K^+-ATPase β-subunit.

Misdirected proteins can be sorted between apical and basolateral domain by transcytosis. Only the basolateral membrane has an ankyrin/spectrin network attached to it; the cytosolic domain of proteins belonging there interacts with this network, locking them to the cytoskeleton. Tight junctions (see Fig. 14.10 on page 335) prevent direct exchange of proteins between the apical and basolateral domains.

17.5.1.3 Lysosomal Proteins

Lysosomal proteins are initially produced and glycosylated like those destined for export. However, in the *cis*-GOLGI one or two outer mannose residues of the high-mannose glycoproteins are phosphorylated by first adding GlcNAc-phosphate (from UDP-GlcNAc) to them and then removing the GlcNAc. The enzyme responsible for this reaction, GlcNAc-phosphotransferase, recognises lysosomal proteins by specific internal signal sequences.

The mannose-6-phosphate residues thus created tightly bind to the luminal domain of M6P-receptor in the *trans*-GOLGI-network. This receptor gets enriched in clathrin-coated transport vesicles which bind to the **late endosome**.

In the slightly acidic environment of the endosome, the M6P-residues dissociate from their receptor, which is recycled back to the GOLGI-apparatus by transport (recycling) vesicles, via the cell membrane. The M6P-residues are cleaved off the lysosomal proteins by a specific phosphatase, thus preventing them from rebinding to the M6P-receptor.

From the late endosome the lysosomal enzymes move to the lysosome. Once the lysosomal proteins arrive in the lysosome, they undergo conformational changes in the acidic environment (*p*H 5.0) of the lysosome and by proteolytic removal of

presequences. This is required for them to become enzymatically active. The reason is obvious: the hydrolytic activity of lysosomal enzymes, turned loose in any other place, could do considerable damage to the cell.

Lysosomal Storage Diseases
Inherited inability to produce GlcNAc-phosphotransferase (see Fig. 17.11) leads to **I-cell disease**, because the lysosome cannot be supplied with the hydrolytic enzymes it requires for function. Instead, the proteins normally destined for transport into the lysosome are secreted. As a result glycolipids and other materials accumulate in the lysosome as inclusions. Interestingly, the lysosomes in the hepatocytes of people with I-cell disease have a normal complement of lysosomal enzymes. Apparently in hepatocytes there must be a second pathway for sorting proteins into the lysosome; the mechanism is not known, however.

There are other "lysosomal storage diseases" caused by the inability to produce one of the hydrolases of the lysosome. The corresponding substrate will accumulate inside the lysosome in large amounts. For example, some glycogen is continuously turned over in lysosomes; the function of this pathway is unknown. In POMPE disease (glycogen storage disease type II) the required enzyme, α-1,4-glucosidase, is defective, leading to the accumulation of glycogen in vacuoles inside the cells, for example, in liver, heart, and muscle. The resulting cardiomegaly leads to early death.

The degradation of proteoglycans also occurs in the lysosome and requires a chain of enzymes. If any of these enzymes is defective, **mucopolysaccharidoses** result.

The same enzymes may also be involved in the breakdown of glycoproteins and glycolipids, so that **glycoprotein storage diseases** and **sphingolipidoses** also result.

Depending on which of the enzymes is affected, the symptoms can range from bone and connective tissue deformation over corneal clouding and blindness to severe mental deficiency. Many of these diseases lead to an early death. Treatment is by

Enzyme replacement Some Man-6P receptor is found on the plasma membrane. After infusion of the missing enzyme (produced by genetic engineering with the Man-6P) it will bind to these receptors and be internalised by endocytosis. Thus it will end up in the lysosome. In theory this should not help against CNS problems, as proteins are unable to cross the blood-brain barrier. However, transcytosis across endothelial cells allows successful treatment if the enzyme dose is high enough. Cost of production of the various enzymes for a small number of patients is very high.

(continued)

Substrate depletion With suitable inhibitors the production of the accumulating substance can be reduced, thereby also reducing the amount of accumulating waste product. Glycosylation inhibitors (see Fig. 17.12), for example, are successfully used in sphingolipidoses.

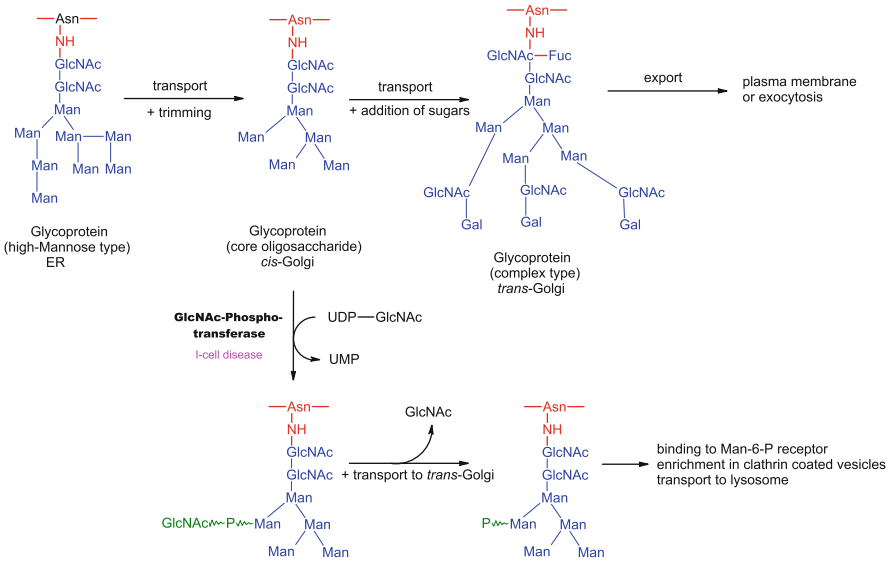


Fig. 17.11 Addition of a phosphate group to a Man residue in the N-linked sugar tree of lysosomal proteins. This Man-residue is recognised by a receptor. Lack of the Man-P transferase leads to I-cell disease, where the lysosomes are nonfunctional, leading to the accumulation of waste material in inclusion bodies that eventually fill the cell

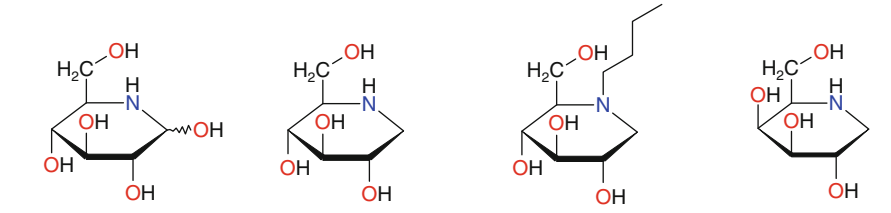


Fig. 17.12 Inhibitors of protein glycosylation that may be used to treat viral infections or sphingolipidoses. The ring oxygen in these sugar analogues is replaced by nitrogen

17.5.2 Proteins Taken Up by the Cell

17.5.2.1 Phagocytosis

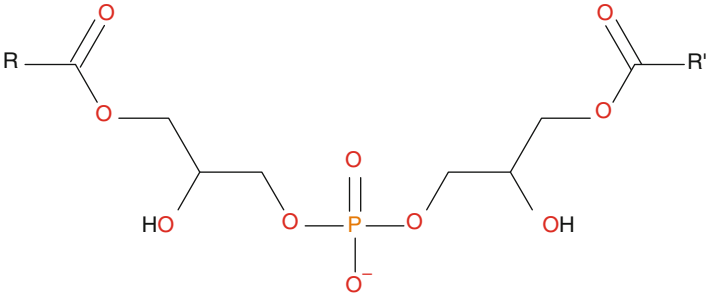
Endocytosis as described above (also known as **pinocytosis**, Gr. cell drinking) works only for small particles ($\leq$ 150 nm); bigger particles are taken up by phagocytosis (Gr. cell eating). Although most cells in our bodies are capable of pinocytosis, phagocytosis is carried out by specialised cells, in particular macrophages. Phagosomes are not coated, and their size is determined by the size of the prey.

17.5.2.2 Endosome and Lysosome

Regardless of the pathway used, material endocytosed from the plasma membrane ends up in the early endosome (EE). The EE collects material for about 10 min and lipids and fluids are removed as the early matures to the late endosome (LE). The EE also communicates with the TGN through vesicles that carry endosomal and lysosomal enzymes. In plants, the early endosome's function is done entirely by the TGN. EEs are found in the periphery of the cell and move along microtubules.

In the TEM EEs (then called **multivesicular bodies**) are seen to contain intraluminal vesicles (ILVs) that develop from plaques in the endosomal membrane and collect ubiquitinated membrane proteins. The ILVs have several functions:

- Remove endocytosed signalling receptors from contact with the cytosol and hence prevent further signalling activity.
- Collect lipids and transmembrane proteins to be degraded in the lysosome and make them easily accessible to lysosomal hydrolases.



bis(monoacylglycero)phosphate

- the ILV membrane contains the negatively charged phospholipid **bis(monoacylglycero) phosphate**, which serves as binding site for lysosomal hydrolases that carry positive charges.
- Cholesterol is initially enriched in ILVs, but then is collected and transported back to the ER membrane. Sphingolipids are degraded to ceramide at the same time. If cholesterol removal cannot occur, NIEMANN-PICK-disease type C results (OMIM #257220 and #607625, note that this is distinct from types A and B caused by defective sphingomyelinase).

Cytosolic proteins with the KFERQ-motive are taken up into the EE by vesicle formation (**microautophagy**); this requires help by Hsc70. Fusion of ILVs with the LE membrane releases their content into the cytosol (**backfusion**). In addition, EE may fuse with the plasma membrane, releasing their ILVs as **exosomes** into the body fluid.

The number of ILVs increases as an EE matures into a LE; the LE is localised perinuclear and formed by fusion of EEs. It is no longer capable of exchanging material with the plasma membrane by recycling vesicles. During this process Rab5 is replaced by Rab7 and Rab9, and phosphatidylinositol-3-phosphate is converted to phosphatidylinositol-3,5-bisphosphate. Both the Rab-proteins and the phospholipids recruit different proteins to the vesicle membrane. When the LEs fuse with the lysosome they contain only lysosomal hydrolases and material to be degraded. This sorting between recyclable and degradable material involves dynein- and kinesin-generated oscillatory movement that presumably generates pulling forces.

The result of the fusion between LE and lysosome is the endolysosome, a hybrid organelle that contains both lysosomal and endosomal markers. It eventually matures into lysosomes. The lysosomal membrane is resistant to the action of hydrolases because of special glycoproteins including **lysosome-associated membrane glycoproteins (LAMP)s**. Lysosomes also fuse with **autophagosomes**, which contain cytosolic material—up to entire organelles—to be degraded.

17.6 Exercises

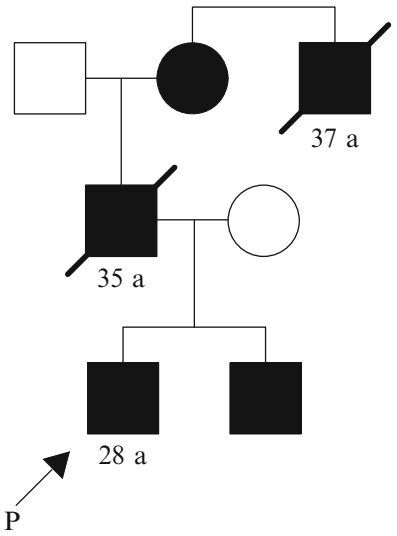
17.6.1 Problems

17.1. A 28-year-old man was brought to the emergency department with chest pain and dyspnea[1]. There was no relevant medical history. Examination revealed a BMI of 22 kg/m^2, tendon xanthomas on hands and feet, corneal arcus (cloudy rim around the cornea) and tachycardia. Echokardiogram and electrokardiogram (EKG) were consistent with an acute myocardial infarct (AMI). Laboratory tests revealed elevated total and LDL cholesterol (16.8 and 14.6 mM, respectively; normal less than 7.5 and 4.9 mM). Several family members also had xanthomas, some had died from an AMI in their 30s.

What is the underlying cause of the disease?

A Chylomicrons are not absorbed by muscle.
B Muscular cholesterol synthesis is not down-regulated by plasma [cholesterol].
C LDL is not taken up in the liver.
D LDL particles contain excess cholesterol.
E HDL cannot transport cholesterol to peripheral tissues.

[1] A full account of this case may be found in [1].



17.2. Pedigree of the patient in the case above. Filling indicates the presence of xanthomas, age of myocardial infarct (if any) is noted.

What is the pattern of inheritance of this disease?

A autosomal recessive
B autosomal dominant
C X-linked
D Y-linked
E Mitochondrial

17.6.2 Solutions

17.1

C This is a case of familial hypercholesterolæmia. The LDL-receptor is defect; mutations may affect transcription, translation, maturation, LDL-binding, sorting into clathrin-coated vesicles or recycling of the stripped receptor to the plasma membrane. In any case, LDL particles cycle through the body about twice as long as normal (four rather than two days), resulting in an increased plasma [LDL]. In addition, because the liver does not receive cholesterol from the blood, cholesterol synthesis is not blocked at the HMG-CoA synthase level. Thus, the liver's cholesterol production is abnormally high, further increasing plasma cholesterol levels. Apart from xanthoma and corneal arcus, which are more cosmetic problems, this leads to a significant increase of the risk for AMI.

Familial hypercholesterolæmia is the most common genetic condition, about 1:500 persons is affected, amongst survivors of an AMI this increases to 1:20.

17.2

B This is a typical case of autosomal dominant inheritance, all generations and all sexes are affected, transmission is through both male and female line. This mode of inheritance at least in part explains the high incidence of the disease, as the relative frequency of affected persons increases with the relative frequency of the defective allele (here 1:500). If the inheritance pattern were autosomal recessive, two carriers would have to mate and the disease frequency would increase with the square of the relative frequency of the defective allele, that is $1 : 500^2 = 1 : 250\,000$.

References

1. Al Montasir, A. and Sadik, M. Acute myocardial infarction in a 28 year man with familial hypercholesterolemia. Indian J. Med. Sci. **66**(3), 78–81 (2012). doi: 10.1534/genetics.112.142810
2. C.K. Barlowe, E.A. Miller, Secretory protein biogenesis and traffic in the early secretory pathway. Genetics **193**, 383–410 (2013). doi: 10.1534/genetics.112.142810
3. E. Buxbaum, P.G. Woodman, Selective action of uncoating ATPase towards clathrin coated vesicles from brain. J. Cell Sci. **108**, 1295–1306 (1995)
4. E. Buxbaum, P.G. Woodman, Binding of ATP and nucleotide analogues to Hsc70. Biochem. J. **318**, 923–929 (1996a)
5. E. Buxbaum, P.G. Woodman, The speed of partial reactions of the uncoating ATPase Hsc70 depends on the source of coated vesicles. J. Cell Sci. **109**, 705–711 (1996b)
6. A. El-Sayed, H. Harashima, Endocytosis of gene delivery vectors: From clathrin-dependent to lipid raft-mediated endocytosis. Mol. Therapy **21**(6), 1118–1130 (2013). URL http://www.nature.com/mt/journal/v21/n6/pdf/mt201354a.pdf
7. A.D. Gillon, C.F. Latham, E.A. Miller, Vesicle-mediated er export of proteins and lipids. Biochim. Biophys. Acta **1821**(8), 1040–1049 (2012). doi: 10.1016/j.bbalip.2012.01.005
8. J. Huotari, A. Helenius, Endosome maturation. EMBO J. **30**, 3481–3500 (2011). doi: 10.1038/emboj.2011.286
9. T. Kirchhausen, Imaging endocytic clathrin structures in living cells. Trends Cell Biol. **19**(11), 596–605 (2009). doi: 10.1016/j.tcb.2009.09.002
10. P.S. McPherson, E.P. Garcia, V.I. Slepnev, C. David, X. Zhang, D. Grabs, W.S. Sossin, R. Bauerfeind, Y. Nemoto, P. De Camilli, A presynaptic inositol-5-phosphatase. Nature **379**, 353–357 (1996). doi: 10.1038/379353a0
11. S.G. Miller, L. Carnell, H.H. Moore, Post-golgi membrane traffic: brefeldin a inhibits export from distal golgi compartments to the cell surface but not recycling. J. Cell Biol. **118**(2), 267–283 (1992). doi: 10.1083/jcb.118.2.267
12. Y. Noda, K. Yoda, Molecular mechanism of the localization of membrane proteins in the yeast Golgi compartments. Biosci. Biotechnol. Biochem. **77**(3), 435–445 (2013). doi: 10.1271/bbb.120982
13. V. Popoff, F. Adolf, B. Brügger, F. Wieland, Copi budding within the golgi stack. Cold Spring Harb. Perspect. Biol. **3**(11), (2011) doi: 10.1101/cshperspect.a005231
14. Y. Saheki, P. De Camilli, Synaptic vesicle endocytosis. Cold Spring Harb. Persp. Biol. **4**(9), (2012). doi: 10.1101/cshperspect.a005645

15. S. Sever, J. Chang, C. Gu, Dynamin rings: Not just for fission. Traffic **14**(12), 1194–1199 (2013). doi: 10.1111/tra.12116
16. S.M. Stagg, P. LaPointe, A. Razvi, C. Gürkan, C.S. Potter, B. Carragher, W.E. Balch, Structural basis for cargo regulation of copii coat assembly. Cell **134**(3), 474–484 (2008). doi: 10.1016/j.cell.2008.06.024

Chapter 18
Transport of Solutes Across Membranes

Abstract Cells and their environment are separated by the plasma membrane, which is permeable only for small hydrophobic molecules. Larger or hydrophilic molecules (or even ions) require protein transporters to get across. These can work as

Primary active transporters that hydrolyse energy-rich molecules to pump ions or molecules actively across a membrane against a concentration gradient. Substrate is usually ATP, but some transporters use other energy-rich molecules such as phospho-*enol*pyruvate, PP_i, or GTP. There are three main groups of ion pumps:

- **F-type** occur in the mitochondrial and plastid membrane; they work as ATP-synthases that convert the chemiosmotic energy of an ion gradient into chemical energy of a phosphodiester bond. Archaea have a related ATP synthase, called A-type. Proton pumps related to F-type occur in vacuoles and other organelles ("V-type"); they are responsible for the low *p*H inside these compartments. Transported ion is usually H^+, but Na^+ may also be used.
- **P-type** energise the plasma membrane by pumping ions against a concentration gradient. Bacteria, yeasts, and plants use a H^+-ATPase, and animals a Na^+/K^+-ATPase. The γ-phosphate of ATP is transferred onto an Asp residue of the enzyme, forming an acylphosphate. This is then hydrolysed by water.
- **ABC-type** use the energy of ATP hydrolysis to pump nutrients into, or waste products out of a cell. Some members of the family have lost the ability to pump substrates actively; they act as regulated channels.

Secondary active transporters use the ion gradient produced by the pumps to transport molecules across the membrane against a concentration gradient. Ions and substrate can go in the same (symport) or in opposite direction (antiport).

Passive channels or pores that facilitate the diffusion of molecules down a concentration gradient, but unlike passive diffusion transport are selective and saturable.

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7_18

Cells are surrounded by a cell membrane, which consists of a double layer of lipid molecules. Only small lipophilic (fat-loving) substances (gases, ethanol) can pass this membrane unaided, most biologically relevant molecules are hydrophilic (water-loving); they cannot pass the cell membrane freely. Water too can penetrate the membrane only slowly. This allows the cell to actively maintain an internal composition that is different from the outside environment, one of the characteristics of life.

Even in equilibrium ion concentrations on both sides of the plasma membrane are not equal, because proteins (most of which are charged at physiological *p*H) cannot diffuse across. Because the number of positive and negative ion species must be equal on both sides of the membrane, this also influences the movement of small ions, resulting in an electrochemical potential named DONNAN-potential after its discoverer.

18.1 Passive Diffusion

Hydrophobic molecules can enter a cell by first partitioning into the extracellular leaflet of the plasma membrane, flipping to the cytosolic leaflet, and then partitioning into the cytosol. Of these steps, flipping across the membrane is the slowest, because of the high viscosity of phospholipids (about 1000 times the viscosity of water).

Partitioning is determined by the relative equilibrium concentration of a substance in water $[S]_a$ and in phospholipid $[S]_l$ (in the lab, octanol is often used as a model for phospholipids). This ratio is called **partitioning coefficient** $K_p = \frac{[S]_a}{[S]_l}$.

The rate of transport of a substance across a membrane (dn/dt, in mol/s) depends on the area of membrane (A), the partitioning coefficient (K_p), the **diffusion coefficient** inside the membrane (D, depends on the viscosity of the lipid and the size and shape of the substrate), the thickness of the membrane (d, 2.5–4.0 nm for a phospholipid bilayer) and the concentration difference across the membrane according to FICK's law:

$$\frac{dn}{dt} = \frac{A}{d} \times K_p \times D \times ([S]_i - [S]_o) \tag{18.1}$$

Most biologically relevant substances have similar diffusion, but quite different partitioning coefficients. Thus in first approximation passive transport of a substance depends on its partitioning coefficient.

As apparent from Eq. (18.1), the rate of passive diffusion through the membrane increases linearly with the concentration difference. If, on the other hand, the transport is mediated by transport proteins, the rate increases hyperbolically with concentration difference. Only a limited number of such proteins are present in the membrane, and each transport process needs a finite amount of time. Thus transport proteins follow HENRI–MICHAELIS–MENTEN-kinetics and passive diffusion and transport can be distinguished simply from their kinetic behaviour.

18.2 Transporters

In general, cell content is higher in K^+ and Mg^{2+} and lower in Na^+, Ca^{2+} and Cl^- than the interstitial fluid (or the outside in the case of a single-cell organism). Also, the cytosol is reducing.

However, passive leakage through the membrane is not zero and over time those gradients would collapse if they were not actively maintained. At the same time, nutrients need to enter the cell, and waste products need to leave. For these purposes, cell membranes contain proteins which allow the selective passage of molecules and ions through "channels". Polarised cells (e.g., in epithelia) need to have different sets of transporters on the apical and basolateral surface in order to achieve substrate transport across the cell.

Transporters come in 3 basic flavours (Fig. 18.1):

Primary active transporters ("pumps") use the energy obtained from the hydrolysis of molecules such as ATP, GTP, pyrophosphate, or phospho*enol*pyruvate to actively pump molecules across the membrane against the concentration gradient. In animals, the most important pump is the Na^+/K^+-ATPase, which hydrolyses 1 molecule of ATP to ADP and P_i to transport 3 Na^+ ions out of, and 2 K^+ into the cell. In plants, bacteria, and fungi a similar function is performed by a H^+-ATPase, which pumps hydrogen ions out of the cell. Some

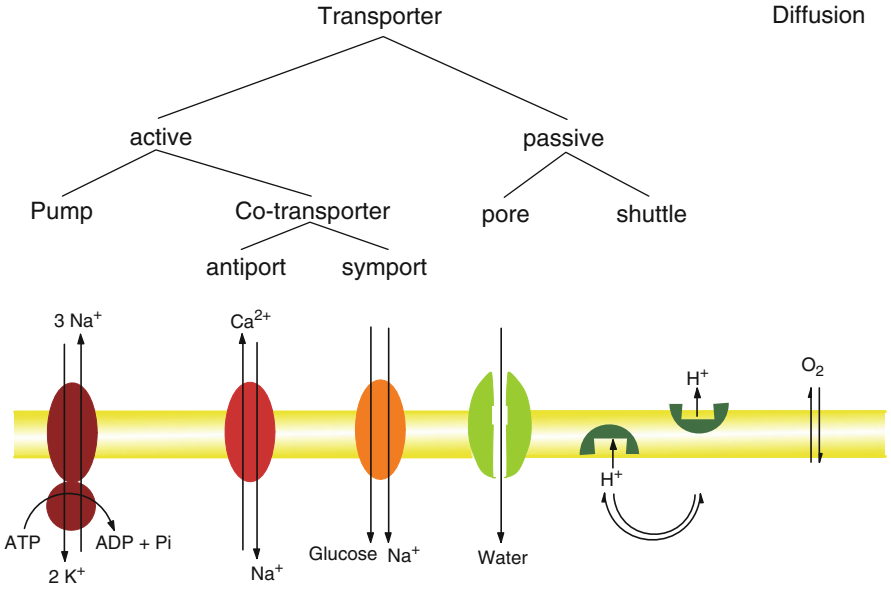


Fig. 18.1 The basic mechanisms of transmembrane transport. Primary and secondary active transport can work against a concentration gradient across the membrane; passive transporters merely speed up the equilibration of gradients

bacteria can use Na^+-ions instead of protons. Under the right experimental conditions, ATP-driven pumps can run in reverse and use the energy of an ion gradient to convert ADP and P_i into ATP (ATP-synthase).

Secondary active transporters (cotransporters) use the energy of the gradients established by the pumps to drive transport of nutrients, hormones, and waste products into or out of the cell. We distinguish

Antiporters where the transport of the substrate occurs in the opposite direction from the coupling ion

Symporters where both substrate and coupling ions flow in the same direction.

Secondary active transport can work against concentration gradients just like primary active transport.

Passive transporters speed up the equilibration of concentration gradients across the membrane by facilitated diffusion. Transport is specific and saturable. They can work either

as gated channels where the substrate flows through a pore in the protein which controls the kind of substrate that can pass, or

as shuttles which bind the substrate on one side of the membrane, then flip around to release the substrate on the other side

18.2.1 Primary Active Transporters (Pumps)

Most pumps use the hydrolysis of ATP to ADP and P_i as their energy source ($\Delta G'^0 = -30.5$ kJ/mol). However, there are some important exceptions. In the plant tonoplast membrane, a H^+-pyrophosphatase splits pyrophosphate (PP_i) and uses the energy to pump protons into the vacuole. Because the pyrophosphate is a metabolic waste product that needs to be split into 2 phosphate ions anyway, the plant can conserve energy ($\Delta G'^0 = -33.5$ kJ/mol) by using it as fuel.

Phospho*enol*pyruvate is a super-energy-rich compound ($\Delta G'^0 = -61.9$ kJ/mol) produced during the glycolysis. It can be used to produce ATP from ADP, but some of the energy stored in PEP is lost as heat. Using PEP directly to drive transport can reduce this loss. Transporters able to do that are found in some bacteria.

Transport ATPases can be grouped into five families, based on subunit composition, protein sequence, reaction mechanism, and inhibitor profiles.

P-type (E_1E_2) ATPases are found in the plasma membrane and in the endoplasmic reticulum. These enzymes occur in two principal conformations, called E^1 (with high affinity for ATP) and E^2 with low affinity. The phosphate group of ATP is transferred onto an aspartic acid residue of the enzyme, forming a phospho-intermediate. **Vanadate** is able to replace the phosphate in the E-P·ADP state, inhibiting the enzyme with very high affinity.

Historical Note
In the early years of Na^+/K^+-ATPase research very different kinetic results were obtained by groups working in America and Europe. By exchanging samples of all reagents between the labs the difference was traced to the ATP preparations used. The American supplier (Sigma) isolated ATP from horse muscle, the European supplier (Boehringer) from yeast. Finally, CANTLEY *et al.* were able to isolate a high-affinity inhibitor from muscle ATP, which was identified as vanadate [8]. Today vanadate-sensitivity of an enzyme is taken as an indicator for the formation of a covalent enzyme-P_i intermediate in the enzyme's reaction cycle.

F_1F_o-ATPases are found in the membranes of bacteria, mitochondria and plastids. This enzyme uses a proton (in some bacteria also Na^+) gradient across the membrane to drive the formation of ATP from its precursors ADP and P_i; that is, it acts as an ATP-synthase rather than an ATP-hydrolase. The enzyme consists of a transmembrane part (called F_o, as it binds the inhibitor oligomycin) which is driven by the ion gradient like a turbine by water and powers ATP-synthesis in the F_1-head. Each of these parts consists of several protein subunits.

A_1A_o-ATPases are very much like F_1F_o-ATPases, but are found in archaebacteria. The protein-sequences of their subunits are different enough from F_1F_o-ATPases that A_1A_o-ATPases are singled out.

V-type ATPases pump protons into the organelles of eukaryotes, such as vacuoles (name!), endosomes, lysosomes, or exocytotic vesicles, but not the ER. They also occur in the plasma membrane of some acid-secreting cells such as osteoclasts (cells that degrade bone). V-type ATPases are inhibited by **bafilomycin A** and by **nitrate**. Structurally, V-type ATPases are similar to F-type, but a considerable evolutionary distance of the sequence results in different behaviour toward their respective inhibitors.

ATP binding cassette (ABC)-type ATPases form the largest family of active transporters, with some 500 identified members. Originally, these transporters consisted of 2 transmembrane subunits, 2 cytosolic subunits which bind and hydrolyse ATP, and a regulatory subunit. Nutrient importers have an additional extracellular binding protein that binds the substrate with high affinity and makes it available for transport. However, in the course of the evolution of some ABC-transporters these subunits have fused. No specific inhibitors of this family are known. They are inhibited by vanadate, even though apparently they do not form a phospho-intermediate. ABC-transporters pump nutrients into the cells, waste products and toxins out, and metabolites from cytosol to organelles or *vice versa*. A few ABC-proteins have lost ATPase function altogether and now act in regulation. Mutations in some ABC-ATPases cause some important inherited diseases such as cystic fibrosis.

18.2.1.1 P-Type (E^1E^2) ATPases

P-ATPases occur in cell membranes and in the endoplasmic reticulum.

In the sarcoplasmic reticulum (ER of muscle cells) a **Ca^{2+}-ATPase** (EC 3.6.3.8, Fig. 18.2) is responsible for the accumulation of Ca^{2+}; upon activation of the muscle this Ca^{2+} is released into the cytosol and causes contraction. Two different forms of this sarcoplasmic Ca^{2+}-ATPase exist, found in slow-twitch and fast-twitch fibres, respectively.

One of the **aminophospholipid transporter** (EC 3.6.3.1) , responsible for maintaining the difference in composition between inner and outer leaflets of the cell membrane also is a P-type ATPase.

Copper is taken up in the intestine and excreted into bile by two different P-type ATPases (EC 3.6.3.4; note that although these enzymes are

(continued)

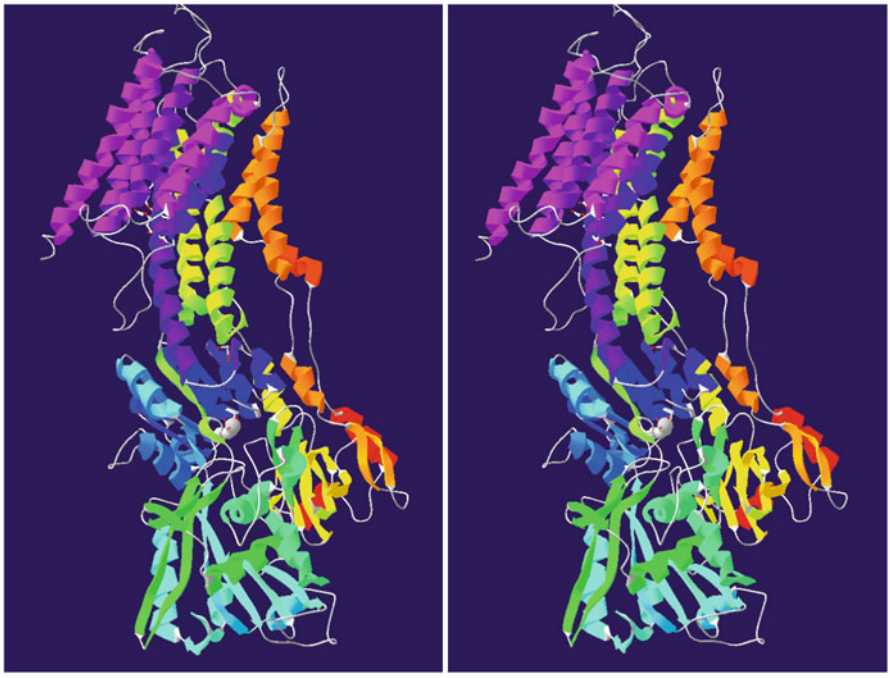


Fig. 18.2 Stereo view of the crystal structure of sarcoplasmic Ca^{2+}-ATPase from rabbit muscle at 2.4 Å resolution (PDB-code 3n5k). The structure is stabilised by the inhibitor thapsigargin. The enzyme contains the Mg^{2+} complex of $[AlF_4]^-$, an analogue of phosphate. Thus the enzyme is in the E^2P state

physiologically different and encoded on different genes, they perform the same reaction and hence have the same EC number). Deficiency in these transporters causes MENKE- (OMIM #309400) and WILSON-disease (OMIM #277900), respectively.

Some bacteria have P-type transporters that export **toxic heavy metals** including Cd or Hg from their cytosol (EC 3.6.3.5, 3.6.3.3). Others take up nutrients such as the **KDP**-system in *E. coli*, which transports potassium (EC 3.6.3.12).

In the acinar cells of the stomach a H^+/K^+-ATPase (EC 3.6.3.10) pumps **protons** into the stomach lumen, causing the low *p*H of stomach juice.

The stomach H/K-ATPase used to incite considerable pharmacological interest, as over-acidification of the stomach causes ulcer, a potentially life-threatening disease. Blockers of H^+/K^+-ATPase were used to treat this condition, until it turned out that the primary cause of over-acidification is an infection of the stomach with the bacterium *Helicobacter pylori* that can be treated with antibiotics. This discovery (for which B.J. MARSHALL and J.R. WARREN shared the NOBEL-Price for Physiology or Medicine in 2005) was originally heavily disputed by the pharmaceutical industry, because H^+/K^+-ATPase blockers needed to be taken for the rest of the patient's life, whereas antibiotics treatment takes only about two weeks; a considerable loss of revenue. Today H^+/K^+-ATPase blockers are still used to treat psychogenic (stress-related) over-acidification of the stomach; of course a change in the patient's life style would be the more appropriate response.

The Membrane Potential

The plasma membrane of all eukaryotes is energised by P-type ATPases, Na^+/K^+-ATPase in animals (EC 3.6.3.9), and H^+-ATPase in plants and fungi. Na^+/K^+-ATPase pumps 3 Na^+ out of, and 2 K^+ ions into the cell for each molecule of ATP hydrolysed. Because more positive charges are pumped out than in Na^+/K^+-ATPase is **electrogenic**. K^+ can leave the cell again through a leakage channel of high conductivity (see Fig. 18.3).

The reversal potential across the cell membrane is calculated for each ion according to the NERNST equation. For example:

$$E_{\mathrm{eq,K}} = \frac{RT}{zF} \ln\left(\frac{[K_o^+]}{[K_i^+]}\right) = \frac{8.3143\,\mathrm{J\,mol^{-1}\,K^{-1}} \times 310\,\mathrm{K}}{1 \times 96\,485\,\mathrm{J\,V^{-1}\,mol^{-1}}} \ln\left(\frac{4\,\mathrm{mM}}{139\,\mathrm{mM}}\right)$$
$$= 26.7\,\mathrm{mV} \times \ln(0.0288) = -94.7\,\mathrm{mV} \qquad (18.2)$$

Ion	Cytosol	ECF	E_{eq}
K^+	139	4	−94.7
Na^+	12	145	66.5
Mg^{2+}	0.8	1.5	8.4
Ca^{2+}	0.2×10^{-3}	1.5	119.1
H^+	7×10^{-5} (pH=7.2)	4×10^{-5} (pH=7.4)	−14.9
Cl^-	4	116	−89.9
HCO_3^-	12	29	−23.6

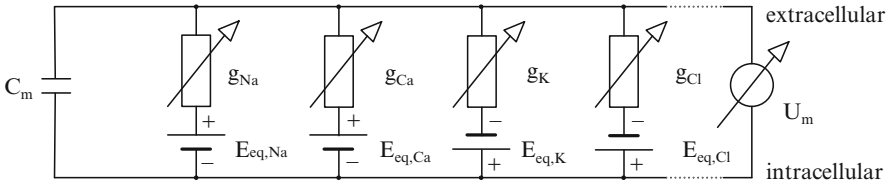


Fig. 18.3 *Top*: Concentration of important ions in the cytosol and the extracellular fluid, given in mM. The resulting reversal potential E_{eq} (see Eq. (18.2)) is given in mV. *Bottom*: The membrane can be modelled by an electrical circuit. Several combinations of voltage sources and resistor are connected in parallel to the membrane capacity ($2\,\mu F/cm^2$). The conductivity of the lipid bilayer for ions is effectively zero, thus the conductivity of the membrane is determined by the conductivities g of the ion channels. Note that the polarity of voltage sources depends on the charge of the ion and on the direction of the concentration gradient

To take all relevant ions into account, the GOLDMAN-(HODGKIN–KATZ-)equation is used:

$$E_m = \frac{RT}{F} \ln \left(\frac{g_{Na}[Na_o^+] + g_K[K_o^+] + g_{Cl}[Cl_i^-] + \ldots}{g_{Na}[Na_i^+] + g_K[K_i^+] + g_{Cl}[Cl_o^-] + \ldots} \right) \tag{18.3}$$

where g_{ion} is the conductivity for that ion (measured in $C\,s^{-1}\,V^{-1}$). The membrane potential is changed not so much by changing ion concentrations, but by changing the conductivity of the channels. For example, during an action potential g_K becomes small, but g_{Na} increases. Thus, E_m drops from $\approx E_{eq,K}$ toward $E_{eq,Na}$. The resulting change in $[Na^+]$ is relatively small: for a cell of 5 µm radius ($V = 524\,\mu m^3, A = 314\,\mu m^2$) a change of $\Delta E_m = 100$ mV requires transport of 1.9×10^6 of the 3.8×10^9 Na^+-ions present (0.05 %). Because g_K is considerably larger than any other ion conductance the membrane's resting potential is near the reversal potential of K^+.

Most secondary active transporters in eukaryotes are powered by the Na^+-gradient.

Reaction Mechanism of P-Type ATPases

We have some knowledge about the reaction mechanism of P-type ATPases; most of our understanding was gained from work on Na^+/K^+-ATPase (see Fig. 18.4).

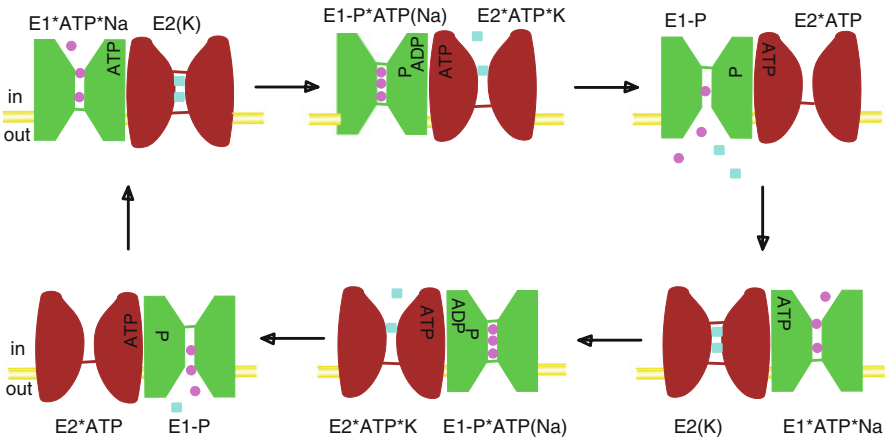


Fig. 18.4 Reaction cycle of Na^+/K^+-ATPase as an example of P-type ATPases. Note the collaboration of 2 protomers, one in E^1, the other in E^2-conformation. Some investigators claim that P-type ATPases work as even larger oligoprotomers

Na^+/K^+-ATPase in its E^1-conformation binds 3 cytosolic Na^+ ions and ATP with high affinity. ATP is split, and the terminal phosphate group is transferred onto an aspartic acid residue of the enzyme.

When substrates are transported, they need to pass a **channel** through the enzyme. This channel must never be open at both ends at the same time, as this would result in free movement of substrate and the breakdown of any concentration gradient. Thus there need to be gates at both ends of the channel, and at least one of them must always be closed (similar to an airlock). This implies that there must be a state where the substrate is inside the channel and both gates are closed, so that the substrate is not accessible from the outside. This is called **occlusion**.

The enzyme releases ADP while delivering the Na^+ ions to the extracellular space. It converts to the E^2-form, which binds 2 extracellular K^+ ions. These are occluded while the phosphate group is transferred to water and released. Low-affinity ATP-binding causes release of the K^+-ions to the cytosol and conversion of the enzyme to the E_1-form, which closes the cycle.

In actual fact there are at least 2 molecules of Na^+/K^+-ATPase cooperating in this cycle: while one is in the E^1-conformation, the other is in E^2 and *vice versa* [7]. Thus the energy of ATP hydrolysis from the E^1-protomer can be used to release K ions from the occluded state, as originally suggested in [22]. Some investigators even claim that 4 or more subunits have to cooperate; this is a topic of ongoing research.

Na^+/K^+-ATPase and Cardiotonic Drugs

Na^+/K^+-ATPase α-subunit occurs in our bodies in four **isoforms**, which are expressed in different tissues and have slightly different affinities for

(continued)

cardiotonic steroids. Most cells express the α_1-isoform, whereas heart and arterial wall muscle cells also express the α_2-isoform. In addition, there are α_3 and α_4, about whose function we know relatively little. These combine with three isoforms of β.

Na^+/K^+-ATPase is the receptor for the cardiotonic steroids such as ouabain (Figs. 18.5 and 18.6), which are produced in the adrenal glands in response to ACTH and angiotensin II and involved in the regulation of blood pressure. The Na^+-gradient built by Na^+/K^+-ATPase across the cell membrane is used to drive a Na^+/Ca^{2+}-exchanger, which reduces the cytosolic Ca^{2+}-concentration. According to the BLAUSTEIN-hypothesis ([2], still found in many pharmacology textbooks) inhibition of Na^+/K^+-ATPase by **ouabain** results in reduced Ca^{2+} removal by Na^+/Ca^{2+}-exchanger (NCX) and increased cytosolic Ca^{2+}-concentration. This will activate muscle contraction in the heart and in arterial walls, thus increasing pumped blood volume (**positive inotropic effect**) and blood pressure. Also, the heart's beating frequency is reduced.

However, at physiologically or pharmacologically relevant concentrations of ouabain ($\approx$ 1 nM), inhibition of Na^+/K^+-ATPase—specifically the α2 isoform in heart and arterial smooth muscle (K_d = 32 nM)—is small, only a few %. This has no noticeable effect on cytosolic [Ca^{2+}]. Today, we know that there is not only the Na^+/K^+-ATPase in the plasma membrane's glycerophospholipid domains—which is active as ATP-dependent ion pump—but also a second population in caveolae, which acts as an ouabain receptor and stimulates the EGF-receptor/Src pathway [3, 12, 21, 23].

Ouabain itself is not absorbed in the intestine, and thus of little pharmacological use. However, some plants (most notably the foxglove *Digitalis ssp.*, see Fig. 18.5) produce related compounds which can be used to increase pumped blood volume in elderly patients with congestive heart failure. Foxglove has been in use for this purpose for some 250 years [25] and has prolonged the life of many a person. Today, of course, preparations of purified digitalis-glycosides are used, which allow more precise dosing. This is important as the therapeutic index (difference between effective and lethal dose) of digitalis-glycoside is very small: 1 nM is helpful, 3 nM can be fatal.

On the other hand, excessive secretion of ouabain into the plasma leads to primary high blood pressure, a serious condition that can result in heart failure. This can be reproduced in animal experiments by chronic injection of ouabain. Interestingly, therapy with digoxin does not lead to high blood pressure; the reason for that difference is not understood.

Most P-type ATPases consist of a single subunit (α-subunit) of about 112 kDa, but Na^+/K^+-ATPase in the animal cell membrane and the H^+/K^+-ATPase in the stomach have a second, noncatalytic β-subunit of 47 kDa; the KDP-system in *E. coli* has three subunits.

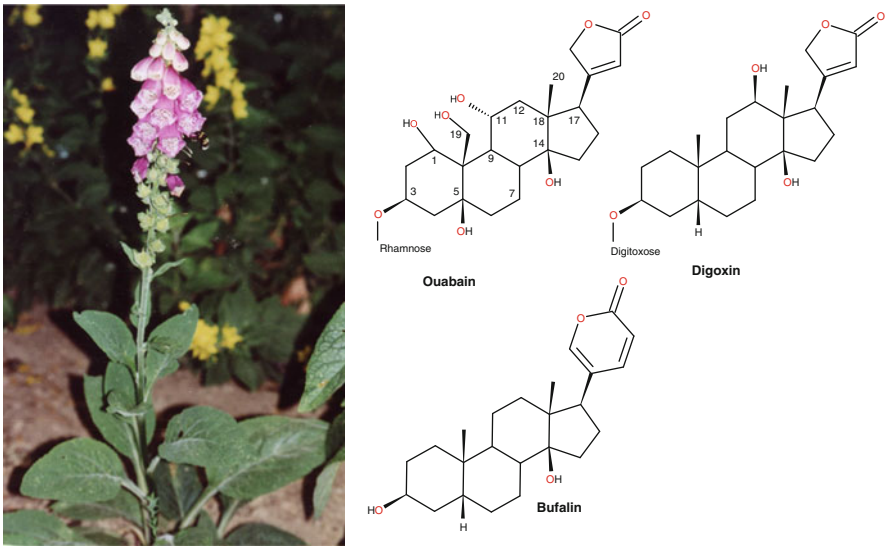


Fig. 18.5 The foxglove (here red foxglove, *Digitalis purpurea*) has been used for centuries to increase heart output in elderly patients. It contains digoxin, which acts (at least in part; see text for further discussion) like the hormone ouabain which regulates the activity of the heart. Both are cardenolides, glycosides of modified steroids; the sugar moiety increases solubility, but has little effect on interaction with Na^+/K^+-ATPase. Bufalin is an example for a bufadienolide, cardiac glycosides found in amphibian skin. These have a 6- rather than 5-membered ring in position 17

The β-subunit acts as a scaffold during folding of the catalytic α-subunit. If either Na^+/K^+-ATPase or H^+/K^+ are expressed in genetically engineered cells without a β-subunit, the proteins are made on the ribosomes, but cannot fold correctly in the ER and are destroyed by the quality control system of the cell. The β-subunit also acts as a sorting signal. If Na^+/K^+-ATPase α-subunit is expressed in cells together with the H^+/K^+-ATPase β-subunit, functional Na^+/K^+-ATPase enzyme is found on the apical, rather than the basolateral side of the cell. Similarly, if H^+/K^+-ATPase α-subunit is expressed together with Na^+/K^+-ATPase β-, we get functional H^+/K^+-ATPase, but at the basolateral rather than the apical side of the cell. In addition, the β-subunit reaches far into the extracellular space and makes contact with β-subunits from neighbouring cells, acting as a cell adhesion molecule [9].

18.2.1.2 F_1F_o-ATPases

The F_1F_o-ATPase is a splendid molecular machine. (P. BOYER)

F_1F_o-ATPases (EC 3.6.3.14, F-type ATPases for short) are found in the plasma membrane of bacteria and in the organelles derived from them: mitochondria and plastids. They act as **ATP-synthases**, using the chemosmotic energy of a proton gradient across the membrane to drive ATP-synthesis (MITCHELL-hypothesis, NOBEL-Price 1978). Remember that reactions are always written in the direction

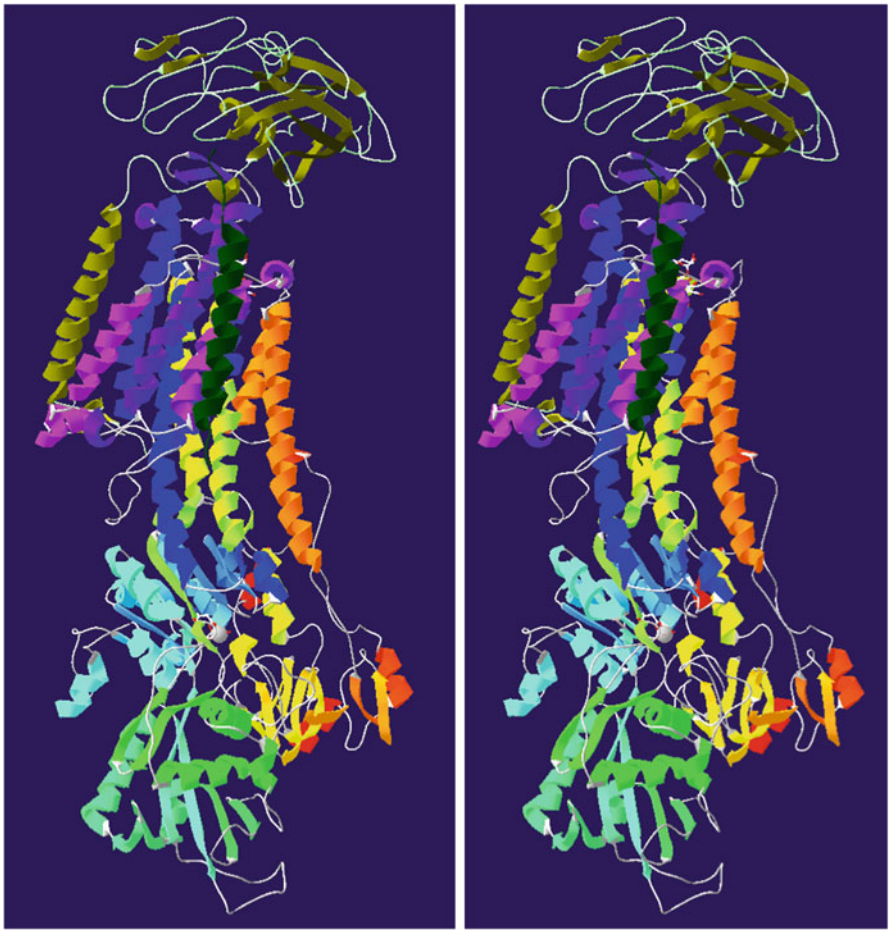


Fig. 18.6 X-ray crystallographic structure of Na/K-ATPase (PDB-code 4hyt). The enzyme is in the E-P conformation; the acylphosphate is shown as ball-and-stick, and the Mg^{2+} space filling. Orientation of the molecule is cytosolic bottom, extracellular top. Inserted from the extracellular surface is one molecule of the inhibitor ouabain (ball-and-stick). The β-subunit is shown in olive (light N-terminal to dark C-terminal); the dark green molecule is the γ-subunit. The molecule is shown in roughly the same orientation as the Ca-ATPase in Fig. 18.2, note the strong similarities

of negative ΔG^0, thus ATP is the substrate and ADP + P_i are products, even if the reaction occurs in the opposite direction under physiological conditions. For a historical review on F- (and V-) ATPases see [14]

In the mitochondria food is oxydised to carbon dioxyde and activated hydrogen ($NADH + H^+$ and $FADH_2$, respectively). The activated hydrogen is then oxydised to water; the resulting energy is used to transport protons across the inner mitochondrial membrane into the intermembrane space:

$$\mathrm{NADH} + \frac{1}{2}\mathrm{O_2} + 11\mathrm{H}_i^+ \rightarrow \mathrm{NAD}^+ + \mathrm{H_2O} + 10\mathrm{H}_o^+ \tag{18.4}$$

In plastids and photosynthetic bacteria the energy of light is used to split water:

$$\mathrm{H_2O} \rightarrow \frac{1}{2}\mathrm{O_2} + 2\mathrm{H}_i^+ + 2\mathrm{e}^- \tag{18.5}$$

The electrons are used to reduce NAD^+ to $NADH + H^+$ for sugar synthesis; during this process further protons are pumped from the stroma into the thylakoid lumen. So 9–12 protons assemble inside the thylakoid lumen for every molecule of water split, and can be used to drive ATP-synthesis.

The difference between mitochondria and plastids is that in mitochondria downhill proton flow is from the intermembrane space into the matrix (outside to inside), whereas in plastids the flow is from thylakoid lumen out into the stroma. Accordingly, orientation of the F_1F_o-ATPases is reversed.

F-ATPases consist of a water-soluble head (the F_1-part) and a transmembrane part (called F_o, because it binds the inhibitor **oligomycin**). F_1 can be isolated and has ATPase-, but no ATP-synthase activity. F_o alone is a unregulated proton channel through the membrane; it cannot bind nucleotides. Only when both parts are brought back together can ATP-synthesis (or ATP-driven proton transport against a concentration gradient under certain experimental conditions) occur (see Fig. 18.7).

The human body produces and uses about 40 kg of ATP each day, each molecule of ATP is converted to ADP and back about 1000 times a day (about once every 9 s). Transmembrane movement of 3–5 protons is required for the production of one molecule of ATP, depending on the number of c-subunits. One additional proton is required for phosphate import into the mitochondrial matrix.

ATP synthesis occurs in the nucleotide binding sites of the β-subunits. The α-subunits have nucleotide binding sites as well, but these have regulatory function and do not directly participate in ATP synthesis.

For many years people have worked at solving the reaction mechanism of F-ATPases. By now we have some understanding how this molecular machine works, and it is probably the most unusual and exciting scientific discovery of recent years (NOBEL-price 1978 to P. MITCHELL and 1998 to J.E. WALKER and P. BOYER, together with J. SKOU for the discovery of Na^+/K^+-ATPase).

The key finding was that during ATP synthesis the substrates are hidden into a solvent-inaccessible pocket of the F_1-head; inside this pocket formation of ATP from ADP and P_i is almost spontaneous (K_{eq} for ATP hydrolysis is 2.4, rather than 1×10^5 as in solution!). This is caused by the tight binding of ATP to the enzyme, K_d for ATP is 1 nM in the catalytic site, for ADP only 10 μM. The difference in binding energy of 40 kJ/mol drives the formation of ATP. The energy of the proton flow through the membrane is required not for ATP-synthesis, but for opening the pocket and releasing the bound ATP. This is achieved by reducing the dissociation constant for ATP from 1 nM in the catalytic to 1 μM and 30 μM, respectively, in the two other sites.

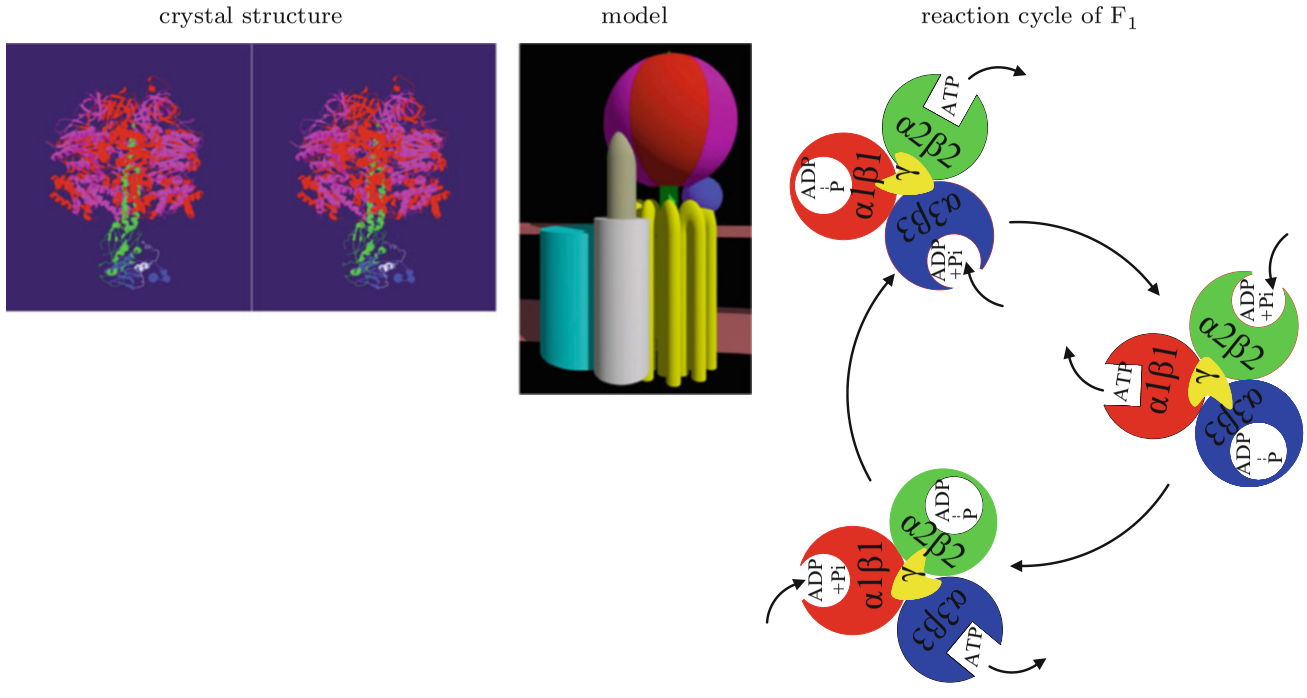


Fig. 18.7 (**a**): F_1-ATPase from bovine mitochondria (PDB-code 2wss). (**b**): Schematic diagram of the entire F_1-F_o-ATPase. α- and β-subunits (3 each, *red and magenta*) form the bulk of the F_1-head, the γ-, δ-, and ϵ-subunits (*green, blue, and grey*) connect this head to the hollow transmembrane cylinder formed by subunit c (*yellow*). During turnover the γ-subunit rotates around its long axis. The proton channel is formed by subunits a (*cyan*) and c (*yellow*); the subunits b (*white*) and δ prevent rotation of the F_1-head. (**c**): Reaction cycle of the F_1-part of ATP-synthase. Rotation of the γ-subunit (driven by proton flow through the F_o-part) makes each of the αβ-protomers assume 3 different conformations in sequence: an open, substrate binding conformation (β_{DP}), a closed conformation where ATP-synthesis occurs (β_{TP}), and an open, product-releasing conformation (β_E)

We have X-ray crystal structures of the F_1-head, in the presence of ATP-analogues and ADP, which give us a good idea how ATP-synthesis and -hydrolysis happen. The conversion between the open and closed conformations of the β-subunit nucleotide binding site is driven by rotation of the γ-subunit around its long axis, which brings it into contact with each of the β-subunits in turn. This results in conformational changes, which affect the affinity for ATP (see Figs. 18.7c and 18.8).

Rotation of the γ-subunit (together with ϵ) during ATP-turnover has been demonstrated directly by attaching a fluorescently labelled actin-filament to its lower end. Each molecule of ATP hydrolysed in F_1 resulted in a turn of the γ-subunit by 120°, and the torque generated (40 pN/nm) is consistent with the energy released by ATP-hydrolysis. By the same method it can be shown that the ring formed by the c-subunits also rotates in the membrane, driving the γ-subunit. With gold particles (which are smaller and experience less drag during rotation) a rotation speed of 134 min^{-1} was observed (at 23 °C, 2 mM ATP), which agrees with the rate of ATP hydrolysis. It appeared as if the 120° rotation for each step could be separated into two steps of 90 and 30 degrees, respectively, with a short dwell time in between. These steps may correspond to substrate binding and catalysis.

The number of subunits in the transmembrane rotor of the F_o-part (see Fig. 18.7) varies from species to species between 8 and 15; each subunit binds one proton per turn, and each turn produces 3 molecules of ATP. Thus the number of protons required to produce one molecule of ATP depends on the number of subunits in the rotor [11], a simple way for evolution to adjust the ATP-synthase to different electrochemical potentials across the membrane. Note, however, that the number of subunits in the rotor is constant for a given species and does not depend on environmental conditions, even in organisms like bacteria, where the environmental conditions can vary significantly.

How does proton flow drive the rotor (subunits c)? It appears that these subunits have a normally protonated Asp-residue, which gets deprotonated in that subunit which faces a positively charged Arg-residue of the stator (subunit b). The released proton moves downhill of the proton gradient, that is, into the mitochondrial matrix (H^+_m). Salt bridge formation between the Arg and Asp residues keeps the c-subunit which faces the stator in a fixed position, while the rotor continues to move. Thus a considerable mechanical tension is built up, similar to the spring in a mechanical clock (see Fig. 18.9). When a proton from the intermembrane space (H^+_i) binds to the Asp-residue, making it uncharged and breaking the salt bridge to the stator's Arg, this tension is released by driving the rotor forward. Thus the rotor/stator assembly of the ATP-synthase resembles the turbine of a water-driven power plant, with the F_1-part being the generator.

Na^+-Powered F_1F_o-ATPases

As discussed, F_1F_o-ATPases are usually powered by protons. In some organisms, however, Na^+ is used instead (EC3.6.3.15), either facultatively or even obligatory. This occurs, for example, in *Vibrio alginolyticus*, a bacterium that lives in algal mats.

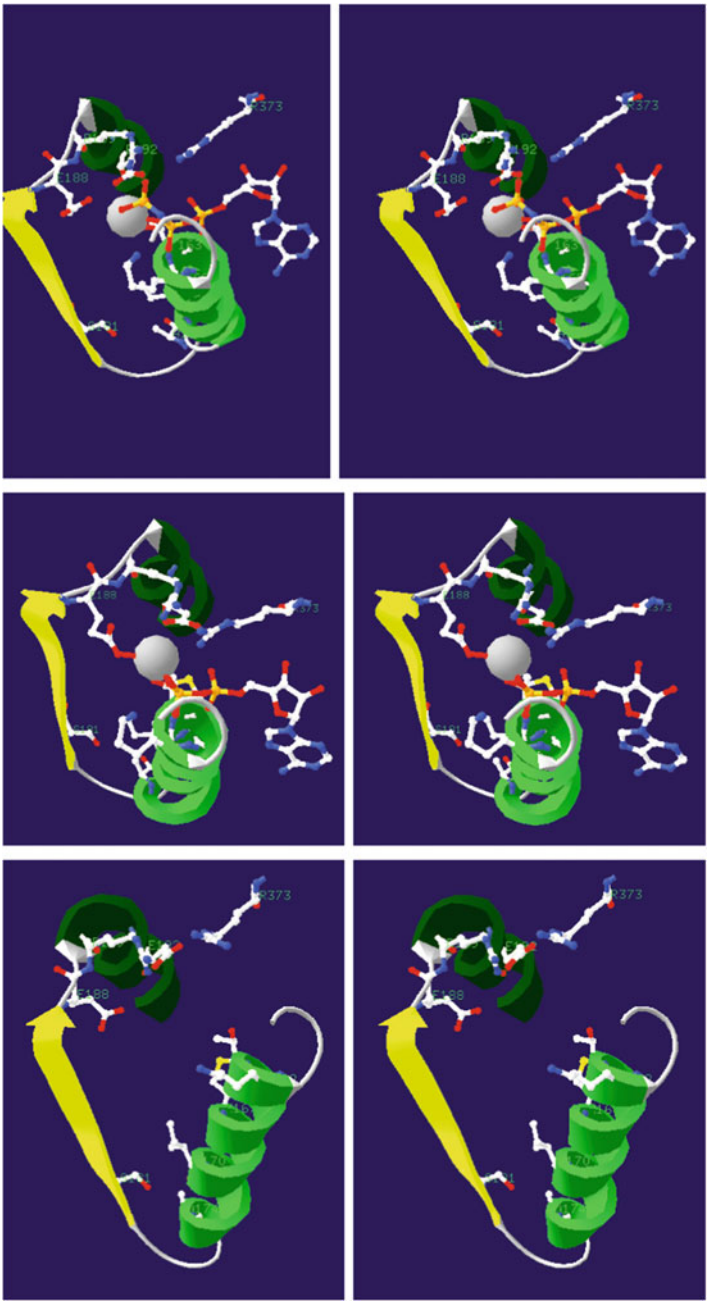


Fig. 18.8 The three states of the nucleotide binding site ("hinge domain") in bovine mitochondrial ATP synthase (PDB-code 2wss), from top to bottom β_{TP} (with AMP-PNP, note the nitrogen between β- and γ-phosphate), β_{DP} (with ADP) and β_E (empty). Shown is the β-subunit from A158 to L196 and R373 from the α-subunit. K155 and R182 bind the γ-phosphate, T156 coordinates the Mg^{2+}, and E181 binds the water required for hydrolysis. E185 and α R376 transmit conformational changes between subunits. Hydrophobic amino acids in the light-green α-helix stabilise the hinge [14]

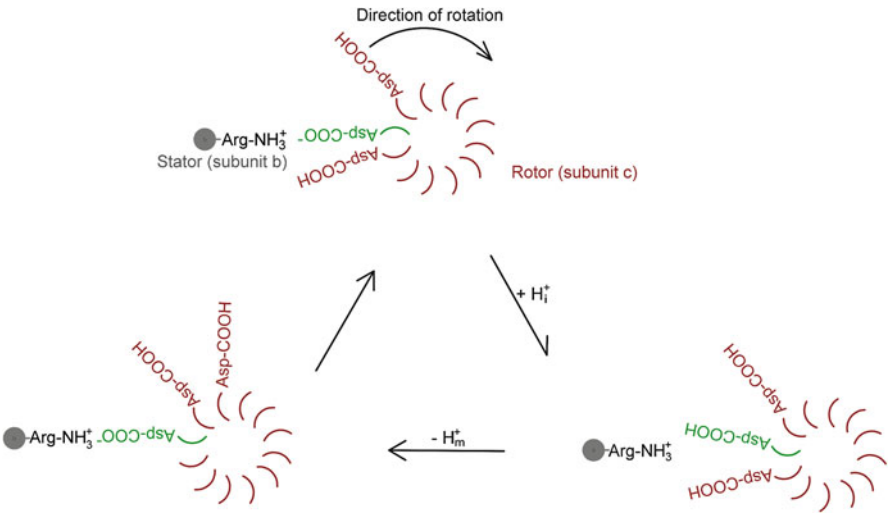


Fig. 18.9 The oldest wheel in the world: how the proton gradient across the mitochondrial membrane drives the rotor of the F_o-subunit of the ATP-synthase. For details see text

As these algae remove carbon dioxyde from the water for photosynthesis, the water becomes alkaline. If the bacterium would export protons during endoxydation, it would effectively try (and fail) to neutralise the medium instead of establishing a useful *p*H-gradient across its cell membrane. Under these circumstances, *Vibrio* will use Na^+ instead, and the resulting Na^+-gradient is used to power ATP synthesis. This is a nice adaption to a specialised environment.

18.2.1.3 V-Type ATPases

Like F-ATPases, V-ATPases consist of a cytosolic head and a transmembrane domain, called V_1 (subunit composition $A_3B_3CDE_3FG_3H_{1-2}$) and V_o (c, c", d, e and in yeast also c'). The A_3B_3-complex has ATPase activity and is linked by the DFd-axle to the proteolipid ring made of c and c" (and c', where present). Subunits C, E, G, H, a, and e form the stator. Some of the subunits have organ-specific isoforms; nothing is known about functional differences between them.

In addition to acidifying GOLGI-apparatus, endosomes, vacuoles, and lysosomes V-type ATPases in the plasma membrane are required in

intercalated cells of the kidney move protons between blood and urine, and hence regulate blood *p*H
clear cells of epididymis create proper storage conditions for spermatozoa
osteoclasts of bone create acidic environment for bone resorption
interdental cells of inner ear endolymph formation
epithelium of nose acidification required for olfaction

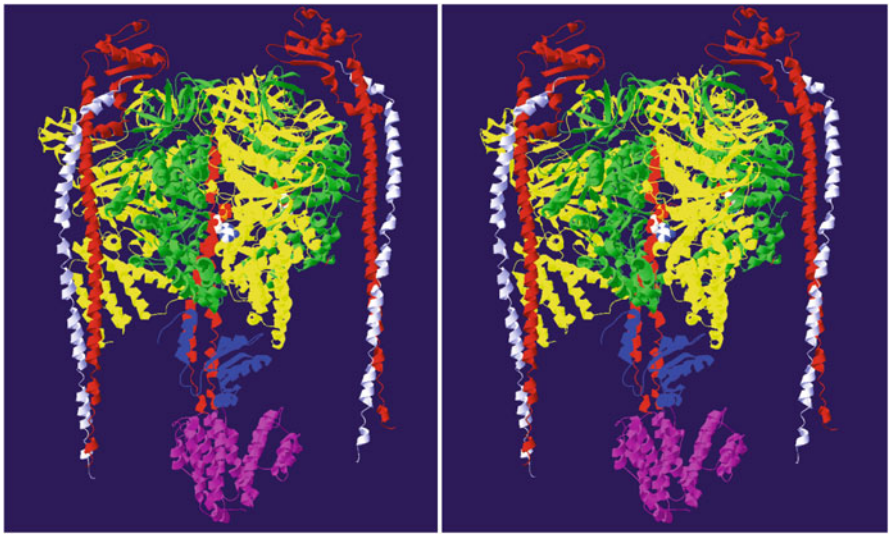


Fig. 18.10 Cryo-electron microscopic image of A_1-ATPase at 9.7 Å resolution (PDB-code 3j0j). Subunits A yellow, B green, C purple, D red, E brown, F blue, G brown. A-ATPases resemble the V-ATPases

In all these functions the V-type ATPase is tightly controlled [4, 24] by cycling between plasma membrane and vesicles (apparently regulated by AMPK and PKA). V_1 dissociation from V_o occurs under low glucose conditions in yeast; the isolated subunits show no ATPase or proton channel activity, respectively. V-ATPase A-subunits have a highly conserved 90 amino acid sequence not present in F-ATPases, the "nonhomologous region". This motive, when expressed separately, binds to the V_0-complex in a [Glc]-dependent manner.

Compartments acidified by V-ATPase are entry points for some enveloped virus (*e.g.*, Ebola and influenza) and toxins (*e.g.*, diphtheria and anthrax). The function of lysosomes and endosomes—and hence immune response maturation—depends on their low *p*H. Defects in tissue-specific isoforms of plasma membrane V-ATPase leads to inherited disease such as distal renal tubule acidosis (OMIM #602722), osteopetrosis (OMIM #259700), or male sterility (epididymis) [24]. During metastasis, V-ATPases create an acidic extracellular environment that activates cathepsins.

18.2.1.4 ABC-ATPases

A good overview over the field of ABC-transporters is [16]. Physiologically ABC-transporters always work as ATPase, but they can work either as importers for nutrients (e.g., trace elements, sugars, vitamins, or amino acids), or as exporters for waste products (*e.g.* bile transporters in the liver), xenobiotics (multidrug resistance transporters), and hormones (e.g., the transporter for the yeast pheromone α-factor). 15 ABC-transporters are encoded in intracellular parasites such as *Mycoplasma*; 70 ABC-transporters (about 5 % of the genes) are encoded by the genome of the free-living *E. coli*; in the more complex human genome 50 out of 30 000 genes (0.2 %) encode ABC-transporters. None of the mammalian ABC-transporters is involved in the uptake of nutrients into the cytosol, one of their major roles in bacteria.

Some members of the ABC-transporter family have lost their transport capacity and act as receptors, which regulate the activity of other membrane proteins, including CFTR and SUR. Others have lost their transmembrane domains and act as soluble enzymes.

Archetypical ABC-transporters consist of 2 transmembrane proteins (TMD), 2 cytosolic proteins with ATPase-activity (nucleotide binding domain, NBD), sometimes a cytosolic regulatory protein; and, in the case of bacterial importers, a high-affinity periplasmic binding protein (see Fig. 18.11). During evolution, some of these proteins may have fused, however, in the most extreme cases (such as the multiple drug resistance ATPase Mdr1) all subunits have fused into a single multidomain protein.

As far as X-ray crystallographic studies have revealed so far, the NBD-domain is an L-shaped molecule with the WALKER-B motive (a sequence characteristic for ATP-utilising enzymes) in the hinge-region. ATP-binding occurs across both

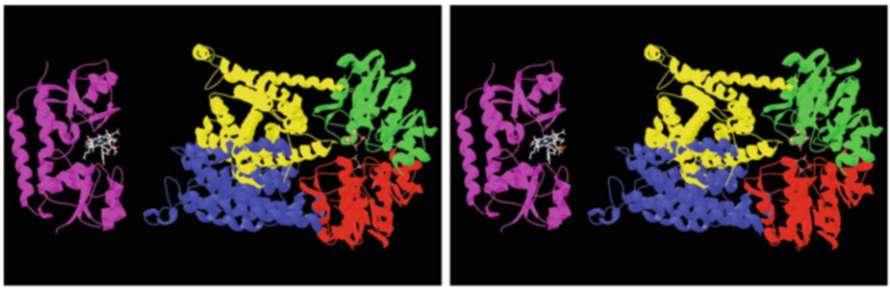


Fig. 18.11 Stereo view of the bacterial vitamin B_{12} importer (Btu for B Twelve Uptake, EC 3.6.3.33, PDB-code 1l7v). Btu is an ABC-type ATPase whose crystal structure has been solved to a resolution of 3.2 Å. Note the 2 permeases (BtuC, *blue and yellow*) with their transmembrane α-helices and the 2 cytosolic ATPases (BtuD, *red and green*). The enzyme was crystallised as complex with **tetravanadate ions**, which mark the ATP-binding site. The **periplasmic binding protein** (BtuF, *magenta*) binds the substrate with high affinity and makes it available for transport. Unfortunately, only BtuC and BtuD have been crystallised together; BtuF was crystallised separately. Thus we do not know how binding of BtuF allows transport

arms of the L and results in conformational changes in line with the induced-fit hypothesis. The adenine ring stacks with a Tyr-residue of the proteins; this contributes significantly to the binding energy of ATP. ATP-hydrolysis leads to a 15° change in the angle between the two arms of the L, bringing the phosphate-binding Q-loop out of reach of the β-phosphate of the ADP. Absence of any nucleotide leads to destabilisation of the NBD-structure. Very little is known about how these conformational changes during the ATPase-cycle are transmitted to the transmembrane domain for coupling with transport.

Our knowledge about the reaction mechanism of ABC-transporters is still very sketchy (see Fig. 18.12). In particular, it is not yet clear whether the oligomer described above is indeed the catalytically active unit, or whether several such oligomers have to work together.

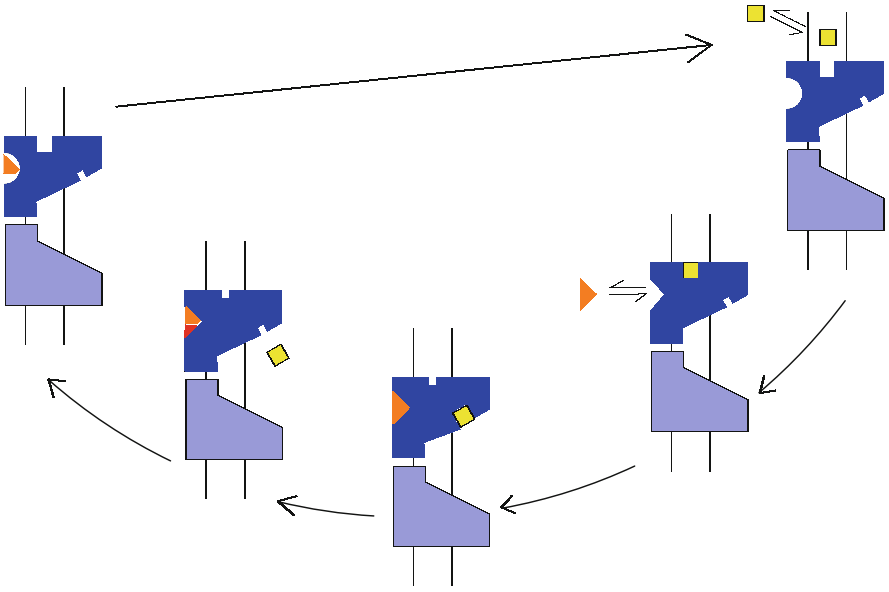


Fig. 18.12 Reaction mechanism of Mdr1 as an example for ABC-type exporters. The substrate of Mdr1 are hydrophobic xenobiotics (*yellow square*), which partition into the cytoplasmic leaflet of the plasma membrane. From there they are bound to the substrate binding site of Mdr1. This changes the conformation of the protein and allows binding of ATP (*orange triangle*). The free energy of ATP-binding is used to translocate the substrate across the membrane (Mdr1 with the nonhydrolysable ATP-analogue. AMP-PNP has an outward facing, low-affinity substrate binding site). Dissociation of the substrate leads to splitting of ATP; the phosphate (*red small triangle*) leaves the nucleotide binding site (the resulting Enzyme-ADP complex can bind vanadate, which is a phosphate analogue). ADP-release then closes the cycle. Note, however, that the Mdr1 molecule consists of 2 almost identical halves (*blue and cyan*); it is currently not known how they cooperate in the transport cycle. Indeed, cooperation of 4 ATP-binding sites can be demonstrated by enzyme kinetics [6], and the distance of fluorescent antibodies against Mdr1 changes during its catalytic cycle [20]. This would indicate that two Mdr1 molecules, with 4 ATP-binding sites, cooperate during turnover

ABC-Transporter in Humans

In humans ABC-transporters are designated with the letters ABC followed by a further letter and a number, like ABCA1. Several of them are involved in inherited diseases. The following list gives an overview of the human ABC-transporters:

ABCA All members of the **ABCA**-family are full transporters (both TMDs and NBDs fused into a single polypeptide). They are involved in the transport of lipophilic substances.

ABCA1 (ABC1, TDG, HDLDT1, CERP) is encoded on chromosome 9q31.1 and expressed in all cells. Cells synthesise phospholipids, especially Phosphatidyl choline (PC), and take up cholesterol from the blood by receptor-mediated endocytosis. Both are incorporated into the plasma membrane, cholesterol mostly into special regions of the membrane, the **rafts**. The external leaflet of the membrane contains preferentially PC and sphingomyelin, the inner leaflet Phosphatidyl serine (PS) and Phosphatidyl ethanolamine (PE). Some of the lipids in the outer membrane are recycled to the liver as High density lipoprotein (HDL). The ABCA1-encoded protein serves as a binding site for Apolipoprotein A-I (apoA-I) and catalyses the ATP-hydrolysis dependent transfer of PC from the membrane to the apoA-I by an unknown mechanism, followed by transfer of cholesterol from the caveolae to the apoA-I/PC-complex.

Apart from its role in lipid metabolism ABCA1 is also required for the uptake of apoptotic cell fragments by macrophages, possibly by catalysing the outward-flip of PS in the plasma membrane, which serves as an "eat me"-signal to macrophages.

Deficiency in ABCA1-expression leads to Tangier-disease (OMIM #205400 (analphalipoproteinæmia), characterised by the deposition of cholesterol esters in peripheral cells (**foam cells**) and absence of HDL in the blood. Lymph nodes, spleen, and tonsils are enlarged, also the liver. Corneas become cloudy, and peripheral neuropathy is also observed. The disease is named after an island in the Chesapeake Bay, where the first family suffering from it was described.

Genetic variability in the noncoding region of ABCA1 may affect the severity of atherosclerosis, without necessarily influencing serum lipid levels. The reason is unclear.

ABCA2 (ABC2) is encoded on chromosome 9q34.4 and expressed in the lysosomal membranes of brain, kidney, lung, and heart. It is closely related to ABCA1; perhaps it performs similar functions in lipid metabolism. A drug-resistant cell line with ABCA2-overexpression has been described.

ABCA3 (ABC3) is encoded on chromosome 16p13.3 and expressed in **lamellar bodies of type II cells** in the **lung**, which produce (and reabsorb) lipid-rich surfactants. Presumably the ABCA3-product is involved in this transport.

ABCA4 (ABCR, RmP, ABC10, STGD1) is encoded on chromosome 1p21.3. It transports N-retinylidien-PE (conjugated visual yellow) from retinal rod outer segments into the cytosol, acting as a lipid floppase (flippases catalyse transport from the cytosolic to the extracellular leaflet (flip); transport in the reverse direction is called a flop). There the compound is converted back to visual purple for reuptake by the rods.

If the visual yellow cannot be regenerated in this fashion, accumulation of this compound and its degradation products leads to **retinitis pigmentosa** (homozygotes), STARGARDT-disease (heterozygotes, OMIM #248200) or **age-related macular degeneration** beyond the age of 60 (partially functional transporter).

ABCA5 is encoded on chromosome 17q24.3 and expressed in muscle, heart, testes, and hair follicles. Defect causes congenital generalised hypertrichosis (OMIM #135400). The pathomechanism is unknown.

ABCA6 is encoded on chromosome 17q24.3 and expressed in the liver. Its function is unknown.

ABCA7 is encoded on chromosome 19p13.3 and expressed in spleen, thymus, and bone marrow. Its function is unknown; it may be involved in lipid homeostasis.

ABCA8 is encoded on chromosome 17q24.3 and expressed in ovary, but also in testes and the CNS. It is assumed to be a xenobiotic exporter.

ABCA9 is encoded on chromosome 17q24.2 and may be required for the maturation of monocytes into macrophages.

ABCA10 is encoded on chromosome 17q24.3 and expressed in muscle and heart. Its function is unknown.

ABCA12 is encoded on chromosome 2q34 and expressed in skin. The protein transports glucosylceramide (see Fig. 18.13) from keratinocytes. Defects cause ichthyosis (OMIM #601227 for amino acid substitutions and #242500 for nonsense-mutations).

ABCA13 is encoded on chromosome 7p12.3; it is ubiquitously expressed. Mutations in *ABCA13* have been linked to depression, schizophrenia, and bipolar disorder; the mechanism and, indeed, the function of ABCA13 are unknown.

ABCB is the only family of human ABC-proteins which contains both full and half-transporters. Half-transporters can form either homo- or hetero-dimers and are found in organelles, whereas the full transporters are found in the plasma membrane.

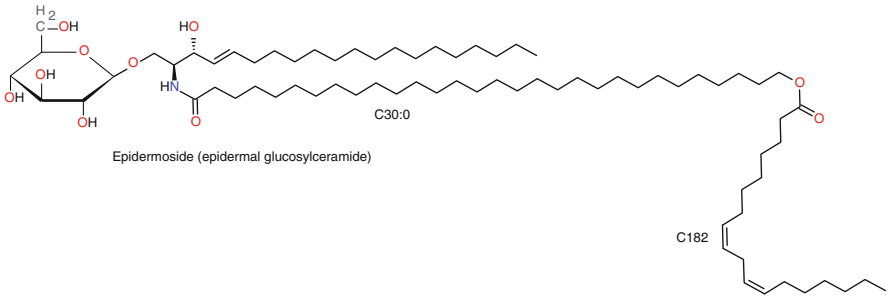


Fig. 18.13 Glucosylceramids serve functions that we only begin to understand [17]. Epidermosid is found in mammalian skin and transported into extracellular lipid in the *stratum corneum* by ABCA12. It is required for water barrier function

ABCB1 (EC 3.6.3.44) is encoded on chromosome 7q21.12. The **multiple drug resistance transporter Mdr1**, also known as **permeability glycoprotein (Pgp)** is a full transporter involved in the protection of our body against xenobiotics, which may be present in our food. It is expressed in intestine, kidney, liver, and the cells of the blood-brain and blood-testes barriers. Mice have 2 different genes for ABCB1, *Abcb1a* and *Abcb1b*; double knockouts are viable but sensitive to neurotoxins such as Ivermectine (acaricide, insecticide). The same, incidentally, is true for collie dogs, which have a mutation in *Mdr1*. This was the first indication for the importance of Mdr1 for the blood-brain barrier.

Unlike most other enzymes, Mdr1 is not specific for a single (or at least a small group of) substrate, but transports a wide variety of compounds, which show cooperativity in stimulating Mdr1 ATPase activity [5]. They all seem to be amphophilic molecules, which partition between the cytosol and the cytoplasmic leaflet of the membrane. The binding site of Mdr1 seems to open to the membrane, rather than to the cytosol. Thus cellular metabolites (which are usually hydrophilic and do not partition much into the membrane) do not act as Mdr1 substrates. Molecules bound at the cytoplasmic leaflet are then transported across the membrane and released into the extracellular medium (**not** the extracellular leaflet of the membrane!) under expenditure of ATP. This is called the "hydrophobic vacuum cleaner model of Mdr1" (see Fig. 18.12).

Unfortunately, during chemotherapy of cancer patients, some cancer cells which by chance express a higher concentration of Mdr1 in their membrane are selected for (recall that cancer cells are aneuploid). Thus the cancer will initially vanish, but reappear after a few years, when

(continued)

those few cells that survived chemotherapy have multiplied. Such cells can express impressive amounts of Mdr1, which may account for 10 % of all the membrane proteins.

Because of the broad spectrum of drugs that is transported by Mdr1, these cells are resistant not only to the drug originally used to treat the tumour, but against many alternative drugs as well. This so-called multidrug resistance can be fatal for the patient.

Polymorphisms in the ABCB1-locus have been linked to variations in the pharmakokinetics of drugs such as digoxin.

ABCB2/ABCB3 (EC 3.6.3.43) are found on chromosome 6p21 and encode for the Tap1/Tap2 half-transporters, which are responsible for the transport of immunogenic peptides from the cytosol into the ER, where they bind to MHC-I (see Fig. 11.19 on page 260). The peptides transported are preferentially 8–16 amino acids long (but up to 40 is possible) and have the sequence [K/N/R]R[W/Y]...[F/L/R/Y]. Experimental evidence indicates that ATP-hydrolysis in one and ADP-release in the other NBD are associated with peptide transport, and the reverse with regeneration of the cytosolic peptide binding site. Thus two molecules of ATP would be hydrolysed for each molecule of peptide transported. Even if this were true for Tap1/Tap2 we do not know whether it might be true for all ABC-proteins.

Congenital defects in either Tap1 or Tap2 result in **bare lymphocyte syndrome** (OMIM #604571), that is, the inability to make MHC-I. In virus-infected and in tumor cells the expression of Tap1/Tap2 is sometimes downregulated to prevent detection by the immune system. Thus Tap1/Tap2 expression in tumour biopsies can be used for grading and corresponds to the prognosis. Mutations in ABCB2 are also implicated in BEHÇET disease (OMIM %109650), a rare autoimmune disorder.

ABCB4 (Mdr3) is a full transporter encoded on chromosome 7q21.12 and highly homologous with the Mdr1 multiple drug resistance transporter. However, its main function appears to be the flipping of phosphatidyl choline (PC) from the inner to the outer leaflet of the canalicular membrane. From there PC is extracted by cholate micelles.

If Mdr3 is not present, the PC-free micelles can attack the membranes of the cells lining the canaliculi and the bile duct, resulting in **progressive familial intrahepatic cholestasis type 3 (PFIC-3, OMIM #602347).**

(continued)

The resulting nonsuppurative inflammatory cholangitis leads to liver cirrhosis and possibly liver tumours. Females with +/- genotype may develop the disease during pregnancy. Some mutations in ABCB4 have been linked to gallstones.

ABCB6 is encoded on chromosome 2q35. It forms homodimers required for mitochondrial porphyrin import.

ABCB7 is encoded on chromosome Xq21-q22. It forms homo-dimers in the inner mitochondrial membrane and exports FeS-clusters.

The mitochondrial matrix contains an **iron-sulphur cluster assembly machine (ISU)** of about 10 proteins, which have not been fully characterised yet. Fe^{2+} is taken up into the matrix by a membrane-potential dependent secondary transporter; S^0 is synthesised inside the mitochondrium from cysteine. On ISU first [2Fe-2S] and then [4Fe-4S]-clusters are assembled in an NAD(P)H-dependent reaction. They can be used directly for transfer to mitochondrial apo-proteins, but must be exported also to the cytosol for union with (Hsc70/Hsp40-stabilised) apo-proteins there. Export possibly requires chelation (to glutathione?) (see Fig. 18.14).

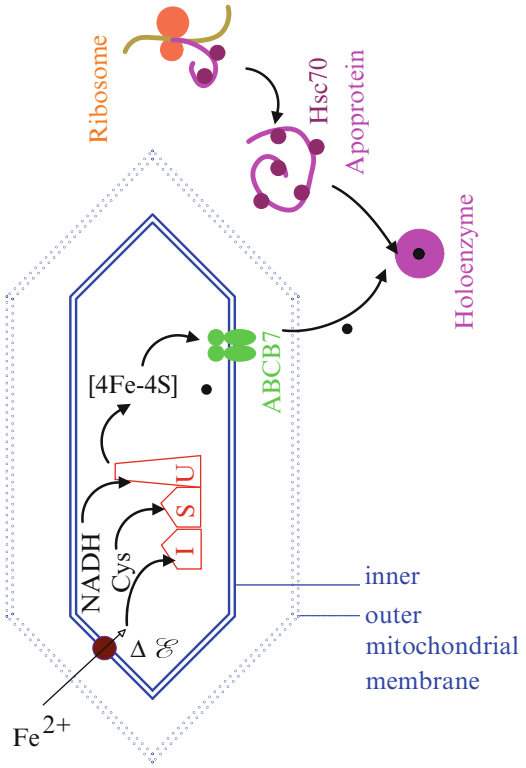


Fig. 18.14 Assembly of FeS-cluster-dependent proteins. The clusters are synthesised in the mitochondrial matrix. The clusters not required inside the mitochondrium are exported through the ABCB6/ABCB7-transmembrane ATPase into the intermembrane space; from there they move to the cytosol by passive diffusion (the outer mitochondrial membrane is leaky enough to allow that). In the cytosol the clusters are united with apo-enzymes which were kept in a folding-competent state by Hsc70 and its cochaperone Hsp40

Mutation in ABCB7 results in **X-linked sideroblastic anaemia/-cerebellar ataxia** (OMIM #301310), where the iron-laden mitochondria of bone marrow cells form sideroblasts (high iron bodies that can be stained by the prussian blue reaction) located around the nucleus.

ABCB8/ABCB10 (EC 3.6.3.43) are encoded on chromosomes 7q36.1 and 1q42.3, respectively. The transporter is responsible for the export of peptides from the mitochondrial matrix to the intermembrane space.

ABCB9 is encoded on chromosome 12q24.3 and expressed in the lysosomal membrane; its function is unknown.

ABCB11 (sister of Pgp (Spgp)) is encoded on chromosome 2q24.3. It acts as a bile salt exporter in the apical (canalicular) membrane of hepatocytes. The protein is a full transporter with considerable sequence homology to ABCB1.

Mutations result in **progressive familial intrahepatic cholestasis type 2 (PFIC-2, OMIM #601847)**.

ABCC genes encode for full transporters of hydrophobic substances. Because many drugs are hydrophobic, ABCC-encoded transporters may be involved in multiple drug resistance, hence they are called **Multidrug resistance related protein (Mrp)**. In addition, the ABCC-family of genes also encodes for 3 proteins which are not pumps, but ion channels (CFTR, ABCC7) or regulators of ion channels (Sur1 and -2, ABCC8 and ABCC9).

ABCC1 (Mrp1) is encoded on chromosome 6p13.12. It's a transporter for anionic xenobiotics conjugated to glutathione. In some cases transported substrate and glutathione need not be actually linked by a chemical bond, but only present at the same time (aflatoxin B_1, and nitrosamines from tobacco smoke). Arsenic and antimony can also be transported as complex with glutathione. In addition, Mrp1 is involved in the IgE-mediated efflux of leukotriene C_4 (LTC_4) from mast cells, important for the regulation of inflammation.

ABCC2 (Canalicular multiple organic anion transporter (cMOAT), Mrp2) is encoded on chromosome 10q24.2. The protein is located in the apical membrane of hepatocytes as well as in kidney, intestine, gall bladder, and lung. Substrates for cMOAT are amphophilic substances conjugated to glutathione or glucuronic acid. One of the most important functions of cMOAT in the liver is the excretion of bilirubin into the gall (see Fig. 18.16).

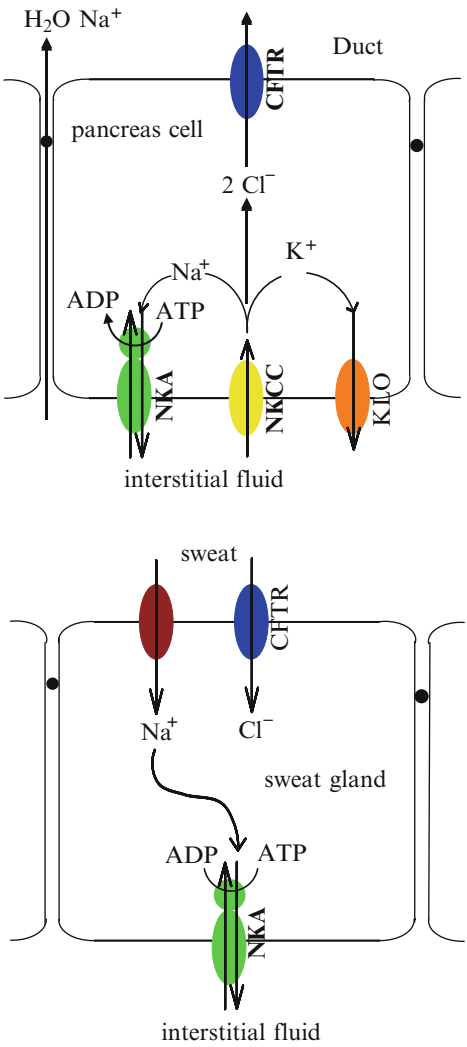


Fig. 18.15 Function of CFTR in the pancreas and lung (*top*) and in the sweat gland (*bottom*). For details see text

Deficiency in cMOAT causes DUBIN-JOHNSON-syndrome (OMIM #237500): chronic jaundice, high bilirubin-levels in serum, excretion of bilirubin-derivatives in urine, hepatomegaly, and melanin deposition in liver. This is a relatively mild disease that does not require treatment.

ABCC3 (Mrp3) is encoded on chromosome 17q21.33. The protein is expressed in liver, intestine, adrenal gland, and gall bladder. Substrates are

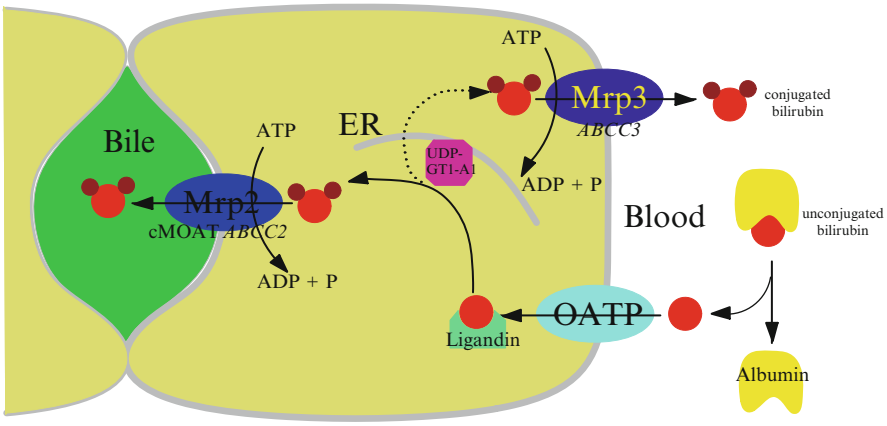


Fig. 18.16 Transport of bilirubin in liver cells. "Unconjugated bilirubin" is almost insoluble in water as intramolecular hydrogen bonds make its polar groups unavailable for interaction with water. In plasma it is bound to albumin. It is taken up into the hepatocyte by the organic anion transport protein (OATP) and immediately bound to ligandin. This complex moves to the ER, where two glucuronic acid residues are added. This conjugated bilirubin can leave the hepatocyte at the canalicular membrane into bile, using ABCC2. Alternatively, it can leave at the basal side into blood, using ABCC3

glucuronlylated or sulphated bile salts and acidic drugs. This transporter is involved in the cholehepatic and enterohepatic cycling of bile salts; it also exports drugs modified in the liver into the bloodstream for excretion with urine. High expression in the adrenal gland may indicate involvement in steroid hormone transport.

ABCC4 (Mrp4) is encoded on chromosome 13q32.1. The protein transports nucleoside monophosphates (both pyrimidine and purine), cAMP, and cGMP. The exact physiological role is unclear. Interestingly both Mrp4 and Mrp5 transporters are inhibited by phosphodiesterase blockers such as Sildenafil (Viagra®).

Overexpression of Mrp4 causes resistance against some nucleoside analogues used in antiviral and anticancer therapy.

ABCC5 (Mrp5), is encoded on chromosome 3q27.1. It transports hyaluronan, a GlcNAc-GlcA copolymer with a molecular mass of 4 MDa that is synthesised by a protein on the cytosolic side of the plasma membrane, from the cytosol to the extracellular space. Excess production of hyaluronan is involved in arthrosis and tumor metastasis; inhibitors for Mrp5 are currently under development.

ABCC6 is encoded on chromosome 16p13.12. It's a transporter in the basolateral membrane; the physiological substrate is unknown, possibly glutathione conjugates or an unknown hormone.

Deficiency causes a rare autosomal recessive disorder, **pseudoxanthoma elasticum** (OMIM #264800). Patients suffer from calcification of the elastic fibres in multiple organs, resulting in loss of skin elasticity, arterial insufficiency, and retinal hæmorrhage. **Generalised arterial calcification of infancy** (OMIM #614473), caused by calcification of the elastic lamina is probably only a more severe clinical presentation of the same disease.

ABCC7 (Cystic fibrosis transconductance regulator (CFTR), EC 3.6.3.49) is encoded on chromosome 7q31.31. It is a nonrectifying anion channel of low conductivity (7–10 pS) (Fig. 18.15). CFTR is unspecific, but because Cl^- is the most common anion inside our bodies CFTR is essentially a chloride-channel (bicarbonate may also play a role). In exocrine glands Na^+, K^+ and 2 Cl^- pass together from the interstitial fluid into the cytosol through the **NKCC**-transporter in the basolateral membrane. The sodium is removed from the cytosol by Na^+/K^+-ATPase (NKA), and the potassium by the potassium leakage channel (KLC), both also in the basolateral membrane (see Fig. 18.15). As a result high [Cl^-] are built up in the cytosol, which leave the cell passively through CFTR at the apical membrane. Water and sodium follow passively by paracellular transport. Cl^--transport is the rate-limiting step in the secretion of glandular products.

In sweat glands Na^+-ions are reabsorbed by an apical Na^+-channel and the basolateral Na^+/K^+-ATPase. Passive flow of Cl^- through CFTR into the cytosol is required for electroneutrality. In addition, uptake of chloride through CFTR seems to be required as counterion for proton transport into the lysosome. If not enough Cl^- is available, lysosomal activity in respiratory epithelia is impaired, resulting in the accumulation of ceramid and apoptosis. CFTR has 9 sites for regulatory phosphorylation by Protein kinase A (PKA), and 9 for phosphorylation by PKC. Their phosphorylation changes the secondary structure of CFTR, increasing the probability of the open (conducting) state. Rapid downregulation of chloride flow is possible by a PP2C-like phosphatase. Additionally, P_{open} can also be increased by binding of ATP to CFTR. ATP binding to the first NBD is fairly stable; in the second NBD ATP is hydrolysed quickly, and ADP and P_i are released. The reason for this difference is not understood.

In cystic fibrosis (CLARKE-HADFIELD-syndrome, mucoviscidosis), an autosomal recessive inherited disease, CFTR is mutated. There are several possible mutations; the most frequent (70 %) is the deletion of phenylalanine 508 (ΔF^{508}-CFTR). This slows down protein folding in the ER so much that most CFTR molecules are degraded and do not reach the cell surface. Those molecules that do are, interestingly, fully functional despite the mutation, but the CFTR concentration in the cell membrane is too low.

Other mutations are found enriched in specific populations such as W1282X in Ashkenazi Jews and 1677delTA in Georgia, Bulgaria, and Turkey.

Patients with cystic fibrosis make a very viscid mucus, which easily becomes infected by bacteria. These recurrent lung infections are often fatal. Pancreas juice and bile secretion are also affected. The reabsorbtion of NaCl from sweat is prevented, resulting in salty sweat, which is a diagnostic marker for the disease. With proper care, patients can reach an almost normal life span.

Because heterozygotes for the CFTR-mutations lose less water in their intestinal secretions, they may be protected from the diarrhoea following infection with intestinal parasites such as *Vibrio cholerae*, *Shigella ssp*, or *Entamoeba histolytica*. This selective advantage may account for the relatively high frequency of mutated CFTR in the general population. $CFTR^-$ knock-out mice indeed do not develop diarrhoea after infection with *Vibrio cholerae*. CFTR is also the receptor for the toxin of *Salmonella typhi*.

ABCC8 (sulphonylurea receptor Sur1) is encoded on chromosome 11p15.1. Four molecules of Sur1 interact with four molecules of Kir6.x to form an ATP-dependent potassium channel (K_{ATP}) in the plasma membrane. The complex is assembled in the ER; only complete molecules can reach the plasma membrane. Kir6.x can bind ATP and ADP, but not AMP, without Mg^{2+} (very unusual!) and is closed by it. However, binding of MgATP to Sur1 opens the channel, thus the regulation of K_{ATP} by ATP is complex. Sur1 hydrolyses ATP.

In the β-cells of the pancreas high blood glucose levels lead to glucose uptake by the GluT2 transporter and to increased glucose metabolism. This results in a change in the ATP/ADP-ratio (in normal cytoplasm [ATP] ≈ 1–5 mM, [ADP] ≈ 100 µM); a decrease in [ADP] leads to dissociation of ADP from Sur1 and allows ATP-rebinding. That in turn closes the K_{ATP}-channel, depolarising the membrane and opening voltage-dependent Ca^{2+}-channels. The resulting influx of Ca^{2+} causes release of insulin from the β-cell.

In α-cells glucagon release is inhibited and in *δ*-cells somatostatin release is stimulated by a similar mechanism. Sur1 is also found in GABAergic cells of the *substantia nigra*.

A defective Sur1 protein results in **Persistent hyperinsulinæmic hypoglycæmia of infancy (PHHI)** (OMIM #256450). If the low blood glucose levels are not treated, severe neurological damage results, possibly leading to death. Treatment is either with K_{ATP}-channel openers such as diazoxyde or by subtotal pancreatectomy. PHHI is rare in normal populations; however, in inbred populations its frequency can rise to 1:2500. Usually inheritance is recessive, but dominant mutations have been described. PHHI may also result from mutations in Kir6.2, glucokinase or glutamate dehydrogenase. Hypoglycæmia may be triggered by a diet high in Leu (OMIM #240800). In other cases, ABCC8 mutations result in pancreatic agenesis and hence permanent neonatal diabetes mellitus (OMIM #606176).

A mutation in Kir6.2 (E23K) has been implicated in **type 2 diabetes mellitus** (OMIM #125853). At least in mice, this mutation can cause diabetes only in association with **obesity**. The increased blood lipid levels in obesity change the ATP-sensitivity of the K_{ATP}-channel, resulting in hyperinsulinæmia followed by apoptosis of the β-cells. Inhibitors of the K_{ATP}-channel (imidazolines, sulphonylureas, and benzamido-derivatives) are used to treat diabetes, as they can be given orally. K_{ATP}-openers including chromakalim have little clinical use at present.

ABCC9 (Sur2) is encoded on chromosome 12p12.1. The protein has similar function as Sur1. It is found together with Kir6.2 in dopaminergic SN neurons and seems to protect against seizures in hypoxia. Sur2 is apparently also involved in the regulation of blood glucose levels by the ventromedial hypothalamus.

ABCC10, ABCC11, ABCC12 encode for Mrp7–9, proteins of unknown function and low expression in all tested tissues. No associated disorders have been described.

ABCD proteins (EC 3.6.3.47) are closely related half-transporters, which form homo- or hetero-dimers. The functional consequences of partner choice have not been investigated yet. Also the reasons for tissue specific expression of ABCD-isoforms are unknown. All ABCD-transporters transport long chain (14–24 carbon atoms) and very long chain (> 24 carbon atoms) branched or straight fatty acids (or their CoA-derivatives) into the peroxisomes for β-oxydation. Short- and medium-chain fatty acids are catabolised in mitochondria. The VLCFA-CoA-synthase is located at the matrix leaflet of the peroxisome and associates with ABCD1.

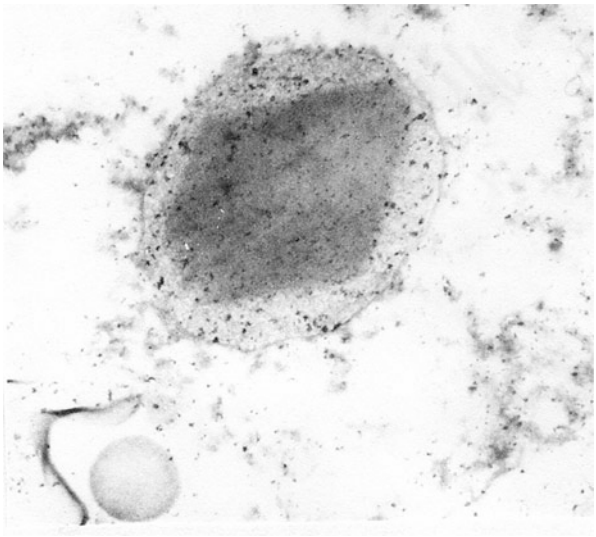


Fig. 18.17 TEM image of a peroxisome in tomato fruit, magnification 17 800. During preparation for TEM proteins in the peroxisome form a crystalloid body

The peroxisome (see Fig. 18.17) can be formed from other peroxisomes by enlargement and fission, or by *de novo* synthesis from the ER. Peroxisomal proteins are expressed in a coordinated manner under the control of the **peroxisomal proliferator activation receptor** α (**PPAR**α); they have C-terminal—SKL and/or N-terminal RKX_5[Q/H]L—import sequences. Diseases related to the peroxisome can be caused by general defects of peroxisome formation (such as ZELLWEGER-syndrome) or by specific defects of a single enzyme, for which adrenoleukodystrophy is a typical example.

ABCD1 (ALD, ALDP), encoded on Xq28, is mutated in **adrenoleukodystrophy**, an X-linked recessive inherited disease. Activity of the synthase is also reduced. Failure to metabolise LCFA and VLCFA results in their accumulation in plasma and tissue (brain, adrenals). Myelin degeneration in the white matter of the brain leads to mental deficiency. The film "*Lorenzo's oil*" is about adrenoleukodystrophy, and probably the only film Hollywood has ever made on lipid metabolism. The mixture of erucic (C22:1ω9) and oleic (C18:1ω9) acid reported in the film does normalise VLCFA in plasma by inhibiting VLCFA synthesis, but fails to slow the progression of the disease significantly.

In yeast PXA-1 and -2 form hetero-dimers which are required for growth on oleic (C18:1) but not octanoic (C8:0) acid as the sole carbon source. β-oxydation of oleic acid is possible in the peroxisomes of such cells if their membrane is permeabilised with detergent.

ABCE proteins have no transmembrane domain at all; the two nucleotide binding domains are fused and act as a soluble enzyme.

> ABCE1, encoded on chromosome 4q31.31, inhibits RNaseL and thereby prevents interferon-dependent protection against virus infections. Expression of its product was recently found necessary for HIV capsid formation.

ABCF These proteins too lack TMDs. ABCF1 is activated by TNF-α and may regulate a kinase responsible for phosphorylating eIF-2α in ribosomes. The functions of ABCF2 and -3 are unknown.

ABCG -encoded proteins are half-transporters [18]. In the fused proteins the NBD is N-terminal, and the TMD C-terminal ("reverse transporters").

ABCG1 (White), encoded on chromosome 21q22.3, exports cholesterol to lipoproteins in macrophages. Defects may be involved in foam-cell formation and hence atherosclerosis. In *Drosophila* ABCG1 is required for eye pigment transport; if the protein is nonfunctional, eyes remain white.

ABCG2 (ABCP, MRX1, BCRP1, CDw338) is encoded on chromosome 4q22 and expressed in the trophoblast and other tissues. It has been found in drug-resistant cell lines; expression of ABCG2 in cells leads to drug resistance, so presumably it forms homo-dimers. Its natural substrate in the kidney is urate; the Q141K mutation increases the risk for gout.

ABCG3 expression has so far been demonstrated only in the hæmatopoetic stem cells of rodents. The function of the gene product is unknown. The sequence of the NBD is unusual; the protein may not be able to bind or hydrolyse ATP.

ABCG4 is highly homologous to ABCG1, its function is unknown.

ABCG5/ABCG8 both encoded on chromosome 2p2.1, form a transporter for plant sterols (phytosterols, in particular sitosterol, 24-ethyl cholesterol) in the apical membrane of liver and intestinal cells. In the intestine, transport of phytosterols out of the cell reduces their uptake into the body; in the liver the transporter increases the rate of excretion into bile. Thus the concentration of phytosterols in our blood is kept low.

> **Sitosterolæmia** (OMIM #210250) results from a genetic defect in either half of the transporter. The elevated blood level of phytosterols leads to atherosclerosis, hæmolytic episodes, arthritis, and tendinous xanthomas (yellowish nodules of lipid-laden histiocytes forming in tendons and fascia). It is unclear whether homo-dimers of ABCG5- or

(continued)

ABCG8-product exist and may have specificity for other sterols of plant or fish origin.

Plant sterols compete with the uptake of cholesterol from the intestine and hence lower blood [cholesterol]. Average uptake in the United States is 80 mg/d, but 345 mg/d in vegetarians. Their absorption is less effective than that of cholesterol (7–22 % vs. 45–55 %). However, there is little epidemiological evidence for beneficial effects of increased phytosterol levels against atherosclerosis or cancer.

18.2.2 Secondary Active Transporter (Cotransporter)

Secondary active transport uses the energy stored in an electrochemical gradient across a membrane for one substrate to transport another substrate against a concentration gradient. As mentioned above, the plasma membrane of animals is energised by a sodium, and that of bacteria, fungi, and plants usually by a proton gradient. Internal membranes are also energised by proton gradients. As these ions "fall downhill", they can "lift uphill" other substrates, for example, nutrients that have to be transported into, or waste products and secreted compounds, that have to be moved out of the cell.

The electrochemical gradient of an ion contains free energy according to the NERNST equation:

$$\Delta G = R \times T \times \ln\left(\frac{[Ion]_i}{[Ion]_o}\right) + z \times F \times \Delta\mathfrak{E} \tag{18.6}$$

with R = universal gas constant (8.28 J/(mol K)), T = absolute temperature (310 K for mammals), z = number of charges per ion, F = FARADAY-constant (96 102 J/(mol V)) and $\Delta\mathfrak{E}$ = electrical potential across the membrane (≈ -70 mV for most animal cells). In mammals, $[Na^+]_i \approx 12$ mM, $[Na^+]_o \approx 145$ mM. Thus moving Na^+ downhill the concentration gradient across a mammalian cell membrane provides ≈ -6.4 kJ/mol + -6.7kJ/mol $= -13.1$ kJ/mol in free energy, with the chemical and the electrical gradient contributing about half each. For comparison, the free energy of ATP-hydrolysis under the conditions of the cell is about -52 kJ/mol (standard free energy only -32 kJ/mol).

The concentration gradient for an uncharged substrate such as glucose, which can be created using Na^+-coupled flow is

$$\frac{[Glc]_i}{[Glc]_o} = \exp\left(\frac{n \times \Delta G}{R \times T}\right) \tag{18.7}$$

$$= \exp\left(\frac{2 \times 13\,100\,\mathrm{J\,mol\,K}}{8.28 \times 310\,\mathrm{J\,mol\,K}}\right) \tag{18.8}$$

$$\approx 27\,600 \tag{18.9}$$

Many cotransporters belong to the **Major facilitator superfamily (MFS)**. These consist of a single polypeptide chain of around 50 kDa with 12 transmembrane helices. They are connected by long cytoplasmic and short extracellular loops (see Fig. 18.18). They usually bind their substrates with µM dissociation constant; transport is reversible under suitable experimental conditions.

Crystal Structure and Reaction Mechanism of LacY

E. coli lactose permease (LacY, see Fig. 18.18) is an example for MFS-proteins [1]. The molecule consists of two domains (helices 1–6 and 7–12), which are folding homologues although sequence homology is low. A mutant protein (C154G) was used for crystallisation, as this mutant is locked in the inward-facing conformation. This reduction in conformational freedom makes crystallisation possible; at the resolution of 3.6 Å major structural features can be recognised but finer details (e.g., the position of bound water molecules involved in indirect hydrogen bonds) are not visible.

From previous biochemical studies it was known that Glu^{126} and Arg^{144} are involved in substrate binding, Arg^{302} and Glu^{325} in proton translocation, and Glu^{269} and His^{322} in coupling between proton and sugar flow.

In the crystal, hydrophobic interaction between Trp^{151} and the pyranosyl ring is apparent which orients the sugar molecule in the binding site. Met^{23} forms a VAN DER WAALS-bond with C6, Arg^{144} a hydrogen bond to O3 of the pyranosyl-group. Glu^{126} may indirectly form a hydrogen bond via a water molecule, however, at the resolution of 3.6 Å this cannot be verified. Lys^{358} forms a hydrogen bond with O4', and Asp^{237} forms a hydrogen bond with Lys^{358} and possibly an indirect hydrogen bond with O4' via a water molecule. Glu^{359} may also form such an indirect hydrogen bond. The proton binding site is formed by Tyr^{236}, Asp^{240}, Arg^{302}, Lys^{319}, His^{322}, and Glu^{325}. Interaction between His^{322} and Glu^{269} link sugar and proton binding.

Mutations in Cys^{148} and Ala^{122} are known to interfere with substrate binding; in the crystal structure they are close enough to the sugar that steric hindrance can explain this result.

As far as can be deduced from biochemical and crystallographic data, LacY in the outward-facing conformation rapidly binds a proton to Glu^{269}/His^{322}, followed by sugar binding to Arg^{144}/Glu^{126}. This leads to a conformational change in the protein which closes the extracellular and opens the intracellular binding site. During this transition, a salt bridge is formed between Glu^{269} and Arg^{144}, and the bridge between Glu^{269} and His^{322}

(continued)

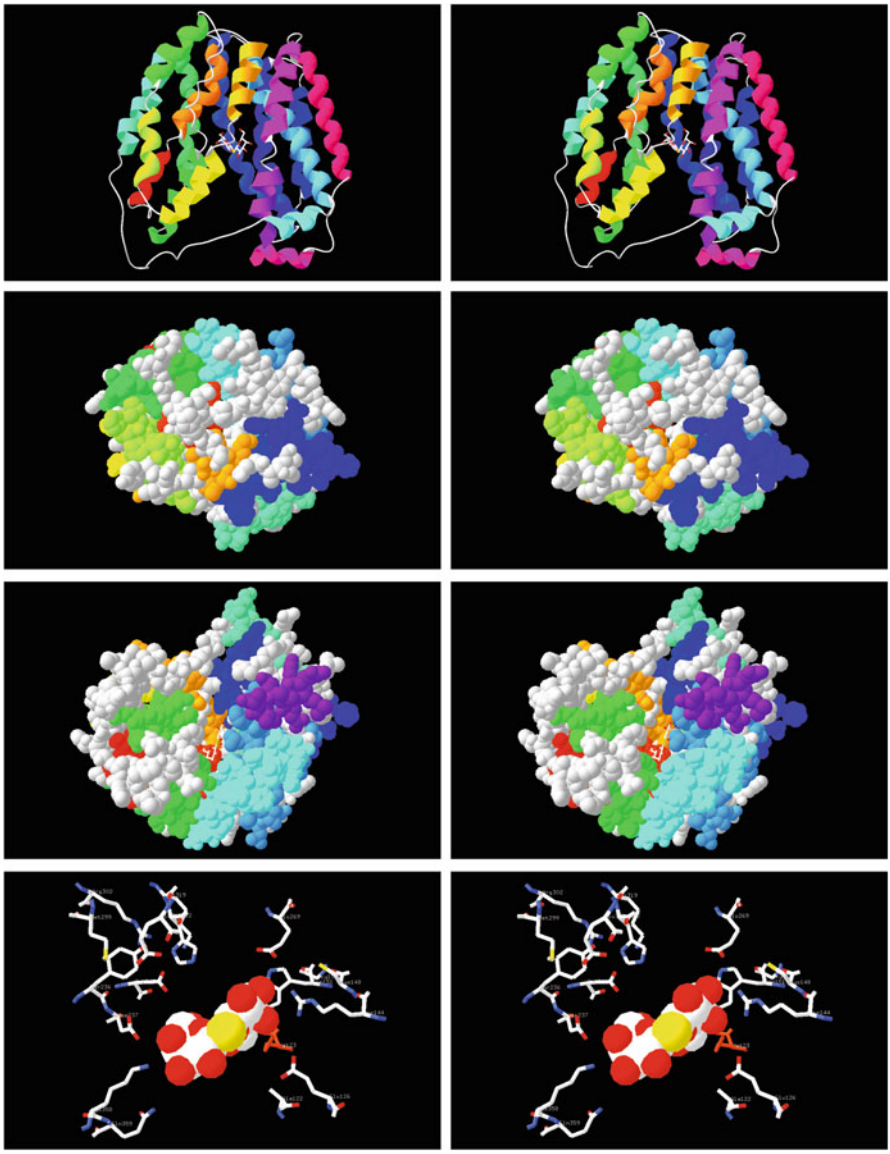


Fig. 18.18 The lactose permease LacY from *E. coli* (PDB-code 1PV7). β-D-galactopyranosyl-1-thio-β-D-galactopyranoside was used as substrate analogue. *Top*: view from the membrane plane (extracellular pointing up), *second*: view from the top onto the closed extracellular binding site, *third*: view from the bottom onto the open intracellular binding site *bottom*: position of important residues with respect to the substrate. For details see text

is broken, leading to proton transfer to Glu^{235}. Lactose is released to the cytoplasm; the resulting conformational change reduces the pK_a of Glu^{235} by interaction with Arg^{302}, Met^{299}, and Tyr^{236}, followed by proton release to the cytoplasm. The cycle is closed by a return of LacY to the outward-facing conformation. According to this interpretation, the reaction mechanism of LacY would be ordered bi-bi.

Some bacteria express MFS-transporters which use external binding proteins to catch substrates, just as ABC-transporters do. These binding proteins bind to a 25 kDa integral membrane protein with 4 transmembrane helices that is closely associated with an MFS transporter. Contrary to normal MFS-transporters, these systems, called **Tripartite ATP-independent periplasmic transporters** (**TRAP-T**), catalyse a unidirectional transport. In some cases, the two membrane proteins have fused during evolution.

18.2.2.1 Symporter

Na^+-linked symporters transport nutrients including amino acids and glucose into many animal cells.

For example, the 2 Na^+/1 glucose symporter has two domains: a C-terminal glucose channel with 9 transmembrane α-helices and a N-terminal domain which couples Na^+ and glucose flow.

Such a transporter (located in the apical membrane) can very efficiently scavenge glucose from the intestinal lumen into epithelial cells (see Fig. 18.21). The basolateral membrane of these epithelial cells contains Na^+/K^+-ATPase to provide the low Na^+-concentration inside the cell, and a glucose uniporter that allows the glucose molecules to move from the cell into the bloodstream by facilitated diffusion (blood glucose concentration is lower than that in the epithelial cell).

18.2.2.2 Antiporter

The Na^+/Ca^{2+}-Exchanger

A typical example for an antiporter is the Na^+/Ca^{2+}-exchanger in cardiac myocytes, which expels a Ca^{2+} in exchange for 3 Na^+-ions entering the cell. The energy of the combined downhill flow of 3 Na^+-ions plus the movement of one net positive charge into the cell (which has a negative potential compared to the extracellular medium) allows the maintenance of a steep Ca^{2+}-concentration gradient between the extra- and intracellular environment (2 mM vs 200 nM).

The Anion Exchanger AE1

When the erythrocyte passes through the capillaries in peripheral tissues, oxygen is released from hæmoglobin and diffuses into the tissue (small gas molecules can move across the membrane without transporter!). At the same time, carbon dioxyde enters the erythrocyte. The carbonic anhydrase inside the erythrocyte catalyses the reaction

$$H_2O + CO_2 \rightleftharpoons HCO_3^- + H^+ \tag{18.10}$$

The proton binds to a His-residue of hæmoglobin, lowering its oxygen affinity and aiding in oxygen release (BOHR-effect; see page 175). The bicarbonate ion is transported out of the erythrocyte in exchange for a chloride ion, thus the concentration of bicarbonate in the erythrocyte is low and the conversion of CO_2 into HCO_3^+ and H^+ is almost complete.

In the lungs, the opposite process occurs: bicarbonate enters the erythrocyte through AE1 in exchange for a chloride ion and is converted to CO_2 by carbonic anhydrase. The carbon dioxyde diffuses out of the erythrocyte and into the lungs. This process abstracts the bound proton from hæmoglobin, increasing oxygen affinity again for complete loading.

About 80 % of all carbon dioxyde generated in our body is transported as bicarbonate, and two thirds of that in the plasma. About 5×10^9 bicarbonate ions pass through AE1 transporters back and forth during each passage of an erythrocyte through the body. AE1 is one of the most abundant proteins in erythrocytes and forms a prominent band if erythrocyte membrane proteins are subjected to electrophoresis. Early workers in the field simply numbered the bands visible after electrophoresis, so AE1 is sometimes called "**band 3 protein**".

18.2.3 Facilitated Diffusion

18.2.3.1 Channels

Potassium Leakage Channels

We have seen that the Na^+/K^+-ATPase in the animal cell membrane removes Na^+-ions from, and pumps K-ions into the cell. The Na^+-ions are used for coupled transport, but K^+ flows back through special leakage channels, which are normally open. This high conductance results in an electrical potential difference across the cell membrane (see Fig. 18.3 on page 428), with the inside of the cell negative. The growing potential difference will resist the outflow of K^+, until the outflow of K^+ powered by the concentration gradient is balanced by the inflow powered by the electrical potential. This steady state is achieved when the voltage across the cell membrane is approximately 70 mV. This is slightly lower than the reversal potential of K^+ (−95 mV for 139 mM inside, 4 mM outside); the remainder is contributed by other ions.

Thermogenin in Brown Adipose Tissue

Shortly after birth a proton channel, the 30 kDa protein **thermogenin**, dissipates proton gradients generated by lipid oxydation in the mitochondria of **brown adipose tissue**. The energy stored in fat is converted to heat instead of ATP; this is used to keep the body temperature of the newborn up until the body can produce heat by other means. In some hibernating animals the same mechanism is used; in some cases thermogenin is induced even in muscle cell mitochondria. Even as adults we still have some brown adipose tissue on our back between the shoulders. This is lost during our first diet, decreasing the basal metabolic rate and making it that much harder to lower or even just maintain weight!

18.2.3.2 Shuttles (Uniporters)

Shuttles bind their substrates on one side of the membrane and release it on the other.

Uncoupler

The simplest shuttles are the so-called "uncouplers" , lipophilic acids that dissolve in the membrane of mitochondria, bind protons on the acidic side, and release them on the basic. Thus the proton gradient across the mitochondrial membrane is dissipated (the stored energy is converted into heat) and ATP production ceases.

Uncouplers such as DNP (Fig. 18.19) are used as rat poison and sometimes for slimming. Both the hyperthermia induced and the loss of ATP-synthesis make this highly dangerous; the difference between minimal effective and fatal dose is very small.

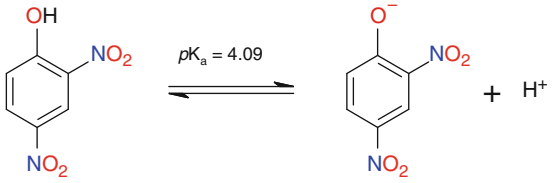


Fig. 18.19 2,4-Dinitrophenol is a typical uncoupler. The hydrophobic molecule can insert into the mitochondrial membrane and shuttle protons from the acidic to the basic side, short-circuiting the proton gradient. DNP has been used as a slimming aid, but may lead to blindness whose onset can be several months after a patient has stopped taking this substance

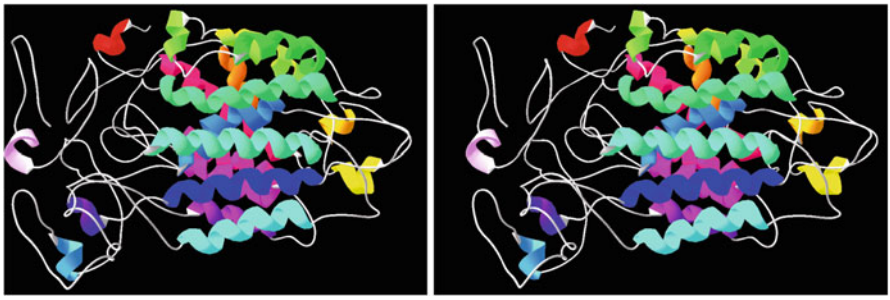


Fig. 18.20 Stereo view of the crystal structure of GluT1 (PDB-code 1JA5)

Glucose Transporters

Glucose is one of the most important energy sources for animal cells; some cells cannot use other sources such as lipids or ketone bodies. Normal blood glucose concentration is about 5 mM; from the bloodstream glucose is transported into the cell by the glucose transporters because glucose is too polar to cross a membrane on its own. There are 14 known isoforms of GluT (SLC2A) in mammals [19]; the four most important are:

GluT1 (Fig. 18.20) is expressed in fetal tissues, erythrocytes, and endothelia (blood-brain barrier!), in small amounts in most tissues. In addition to glucose this protein also transports ascorbate. It acts as receptor for HTLV. Mutation leads to familial neuroglycopenia (DE VIVO disease, OMIM #606777) [15].

GluT2 is a high capacity, bidirectional transporter found in liver, pancreatic β-cells, renal tubules, and small intestine. Defect in FANCONI-BICKEL syndrome (OMIM #227810).

GluT3 is found in neurons and placenta, and some also in fetus, sperm, and white blood cells. There is very high affinity and transport capacity for glucose.

GluT4 is found in striated muscle and adipose tissue. It is insulin-regulated (transport from microsomes to plasma membrane) and stimulated by muscle contraction.

Glucose uptake from intestine and primary urine occurs by Na^+/Glc symporters; SLC5A1 uses two Na^+ per glucose, SLC5A2 one. Free-living bacteria use primary active glucose-pumps, often fuelled with phospho-*enol*pyruvate.

Inside the cell, glucose is immediately phosphorylated to glucose-6-phosphate, thus the glucose concentration inside the cell is virtually zero. GluT1 has a MICHAELIS-constant of 1.5 mM for D-glucose and will be running at 78 % of maximal velocity at normal blood glucose concentrations. K_m for L-glucose is 3000 mM, 20 mM for D-mannose, and 30 mM for D-galactose; GluT1 is very specific for D-glucose and even the change of configuration on a single carbon atom has considerable effects on substrate binding.

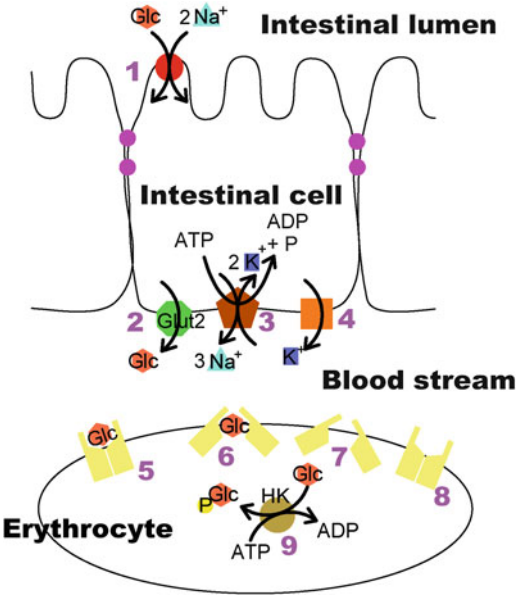


Fig. 18.21 Transport of glucose from the intestinal lumen to the consuming cells, here assumed to be an erythrocyte. Some histological features not required for the discussion were left out. For details see text

Cooperation of Transporters in the Body

Different transporters in different cells of the body work together. Cells of the intestinal epithelium have a Glucose/2Na^+-symporter (see Fig. 18.21, red) that transports glucose from the intestinal lumen into the enterocyte. Coupling with downhill Na^+-transport results in cellular concentrations of glucose which can be almost 30 000-fold higher than in the intestinal lumen (**1**; see page 454 for details). Because of the high glucose concentration inside the enterocyte, glucose will move into the bloodstream passively through the GluT2 uniporter (green, **2**). In the blood, glucose concentration is maintained at 4–5 mM by the liver. Because GluT2 has a K_m for glucose of 66 mM, blood glucose will not be transported into the enterocyte at an appreciable rate, even if the glucose concentration in the enterocyte drops between meals. The energy for glucose accumulation inside the enterocyte is provided by a sodium gradient, which is produced by Na^+/K^+-ATPase (brown, **3**); the Glc/2Na^+-symporter together with Na^+/K^+-ATPase also ensure efficient uptake of Na^+ from the intestine. A potassium leakage channel (orange, **4**) provides a return path for the K^+-ions transported by Na^+/K^+-ATPase and maintains a stable membrane potential. Direct contact between bloodstream and intestinal content is prevented by tight junctions (magenta). The erythrocyte membrane contains GluT1 (cream, **5–8**), which transports glucose into the cell with a K_m of 1.5 mM, thus it works at three fourths of its maximal activity at normal blood glucose concentrations. Inside the cell the glucose is immediately phosphorylated by hexokinase (gold, **9**), which has a K_m for glucose of 0.1 mM and thus prevents the accumulation of appreciable concentrations of glucose inside

the erythrocyte. GluT1 works by binding glucose on an external binding site (**5**), then changing its conformation so that the binding site is exposed to the cytosol (**6**). Dissociation of glucose results in an empty transporter (**7**), which changes its conformation again, bringing its binding site to the outside ready for renewed glucose binding (**8**). The last step is rate-limiting; if erythrocytes are suspended in a solution of radioactive glucose, the rate of uptake increases with the concentration of nonradioactive glucose inside the erythrocyte, because the conformational change is faster in the glucose-loaded than in the empty transporter.

Water Channels: Aquaporins

Water as a polar molecule cannot simply cross the biological membrane. Although some water is always taken in the hydration sphere of transported solutes (ions, glucose), most water moves through specialised water pores, aquaporins (Fig. 18.22).

Of the 150 aquaporin types characterised, 10 (Aqp0 to Aqp9) have been found in humans. They occur in a variety of tissues, in highest concentration in the kidney.

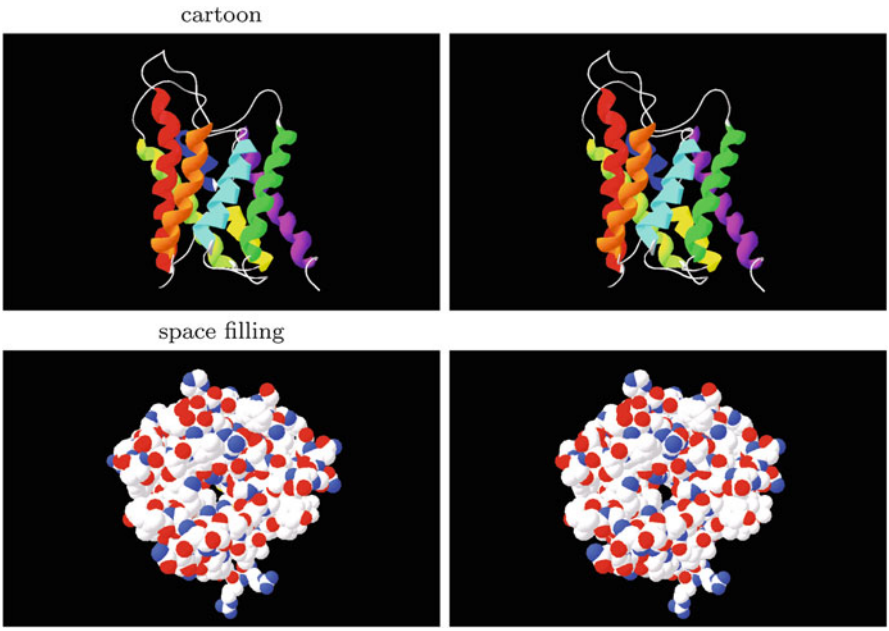


Fig. 18.22 (Top): Stereo cartoon of cattle aquaporin structure at 2.2 Å resolution (PDB-code 1J4N), viewed from the membrane plane. The transmembrane helices are clearly visible. (Bottom): Stereo view of the crystal structure of aquaporin, viewed from the extracellular side. Note the small channel through the protein, which allows passage of water, but not larger molecules (including ions with their coordinated water)

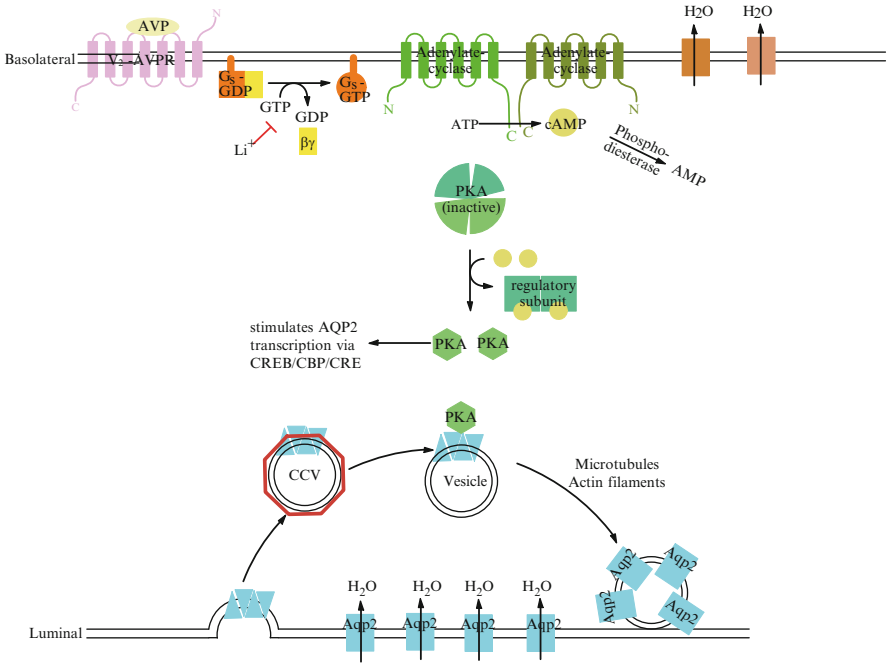


Fig. 18.23 Regulation of aquaporin activity in the kidney by Arg-vasopressin (aka antidiuretic hormone (ADH)). The vasopressin receptor is coupled to adenylate cyclase (EC 4.6.1.1) by G_s (EC 3.6.5.1). This leads to production of cAMP which activates Protein kinase A (PKA) by removing the regulatory subunit. Phosphorylation by PKA (EC 2.7.11.11) leads to a conformational change in Aqp2, which moves from intracellular storage vesicles to the plasma membrane. Once Aqp2 is no longer needed it is taken up into vesicles again by clathrin-mediated endocytosis. The water removed from the primary urine by Aqp2 in the luminal membrane leaves the cell via Aqp3 and -4 in the basolateral membrane. Li^+ interferes with activation of G_s; this explains some of the side effects of this drug

Aquaporins cycle between intracellular storage vesicles and the cell membrane (see Fig. 18.23). Aquaporins are thus involved in regulating urine volume.

Severity of the consequences of aquaporin mutations depends on the type of aquaporin affected. For example, in COLTON-null disease Aqp1 is missing, but this results only in a slightly reduced capacity to concentrate urine, patients can still have a normal life.

Problems in Aqp2 lead to **diabetes insipidus** (OMIM #125800), because water cannot be reabsorbed in the tubuli. Mutations of the AQP2 gene (on chromosome 12q13) tend to prevent the cycling of Aqp2 between storage vesicles and cell membrane. Although the functional unit of aquaporins is the

(continued)

monomer, the transported unit is the tetramer. Of the several known mutations of Aqp2, some have dominant negative inheritance, because a single mutated molecule can prevent the transport of the entire tetramer. Correct diagnosis of diabetes insipidus in dehydrated patients in an emergency room is essential, as rehydration with i.v. saline leads to hypernatræmia, possibly resulting in severe brain damage. If rehydration cannot be performed orally, i.v. glucose may be used.

Cataract and neuronal loss of hearing also may be caused by nonfunctional aquaporins.

After brain injury excess activity of aquaporins leads to brain swelling, which can be fatal. It is hoped that a good understanding of aquaporin structure and function will result in the development of specific blockers, which can prevent this.

Each aquaporin molecule can pass about 3×10^9 water molecules per second. A $1\,m^3$ membrane with densely packed aquaporin molecules would need less than 1 s to desalt 1 l of water by ultrafiltration. Thus aquaporins are also of technical interest.

Water molecules have a size of about 2.8 Å, and the channel of aquaporins is at the narrowest part about 3 Å in diameter. This is, however, not sufficient to ensure the selectivity of water transport because an unbroken chain of water molecules, with hydrogen bonding between them, could lead to proton transport by hydrogen bond exchange (GROTTHUSS-mechanism: a proton being bound at one end of the chain and a different hydrogen being released as proton on the other end). This "proton wire" would be physiologically dangerous and is not observed experimentally. The crystal structure of Aqp1 has solved this riddle: at the narrowest point of the channel two Asn-residues (N76 and N192) form hydrogen bonds with a single water molecule, which thus can no longer form hydrogen bonds with other water molecules. This interruption of the water chain in the channel prevents proton leakage (see, e.g., https://www.youtube.com/watch?v=GSi5-y6NHjY).

Computer Simulation of Enzyme Structure and Reaction Mechanism

The phenomenal increase in computer power over the last 30 years has made it possible to simulate the behaviour of molecules numerically. For this purpose atoms are placed in virtual space, guided for example by the data from X-ray crystallography or NMR-studies. Then there exist forces between these atoms, for example, from stretching, compression, or torsion of covalent bonds, electrostatic interaction between charged or polarised groups, VAN DER WAALS-attraction and so on. PAULI-repulsion prevents atoms from getting too close to each other. These forces can be modelled as mechanical

(continued)

springs between the particles; the resulting movements can then be calculated by numerical integration of NEWTON's laws of motion (see page 189 for a discussion of numerical integration). This is called **molecular dynamics**.

This way it is possible to refine models of protein structure, if their resolution is not high enough. The virtual molecule is exposed to a high temperature, with rapid movement of the atoms. As the temperature is reduced, movement becomes slower and atoms take the position of minimal energy (**simulated annealing**).

It is also possible to simulate the behaviour of molecules during a reaction, for example, the behaviour of water molecules during their passage through aquaporin molecules [10, 13]. The aquaporin molecule was embedded into a lipid membrane, which was immersed in water. The model totalled more than 100 000 atoms. Simulation of this system took several months of calculation time on a supercomputer, resulting in a film that showed the passage of water molecules through the aquaporin. Speed of passage agreed well with measured values, thus the model is probably correct.

The "dance of water molecules" through the aquaporin channel is made possible by the fact that whenever hydrogen bonds need to be broken to allow passage through the narrow channel they are replaced by other hydrogen bonds. Thus the activation energy for water transport is low and the speed can be high. In addition, the aquaporin molecule contains two potential barriers which prevent the passage of both positively and negatively charged ions (see https://www.youtube.com/watch?v=7EGPtMqZ7pY for a cartoon).

Most residues in the channel wall are hydrophobic to allow water to move through the channel quickly, however, there are four water binding pockets with hydrophilic residues as well. These lower the energy required for breaking the hydrogen bonds with other water molecules, which is necessary to transport water molecules in single file.

18.3 Exercises

18.3.1 Problems

18.1. Connect each physiological transport reaction to the type of transporter involved:

1)	Removal of hydrophobic drugs from the cells of the blood-brain barrier	A)	P-type ATPase
2)	Ca^{2+}-transport into the sarcoplasmic reticulum	B)	ABC-type ATPase
3)	transport of N-retinylidene-PE from the retinal rod outer segments	C)	Antiporter
4)	dissipation of proton gradients in brown adipose tissue mitochondria	D)	Symporter
		E)	Channel

18.2. Assume the concentrations of ATP, ADP, and P_i in the cytosol to be 5.0 mM, 0.9 mM, and 5.0 mM, respectively. The free energy of ATP hydrolysis is $\Delta G'^0$ = −30.5 kJ/mol. Calculate the actual ΔG under cellular conditions.

18.3. Using the actual ΔG from 18.2 and assuming that hydrolysis of one ATP moves three H^+, calculate the maximal *p*H-difference between cytosol and lysosome that can be generated by the V-ATPase if the membrane potential is ΔE = 115 mV (lysosome positive).

18.3.2 Solutions

18.1 Removal of hydrophobic xenobiotics is the job of Mdr1, an ABC-type ATPase. Ca^{2+}-transport into the sarcoplasmic reticulum is achieved by Ca^{2+}-ATPase, a P-type ATPase. The transport of N-retinylidene-PE (product of the conversion of visual purple into visual yellow) is the responsibility of ABCA4p. In brown adipose tissue the proton gradient across the mitochondrial membrane is dissipated by thermogenin, a channel.

18.2

$$\Delta G = \Delta G'^0 + RT \ln \left(\frac{0.9 \times 10^{-3}\,\mathrm{M} \times 5.0 \times 10^{-3}\,\mathrm{M}}{5.0 \times 10^{-3}\,\mathrm{M}} \right) \tag{18.11}$$

$$= -30.5\,\mathrm{kJ/mol} + 2.55\,\mathrm{kJ/mol} \ln(9 \times 10^{-4}\,\mathrm{M}) \tag{18.12}$$

$$= -30.5\,\mathrm{kJ/mol} + 2.55\,\mathrm{kJ/mol} \times -7.01 = -30.5\,\mathrm{kJ/mol} - 17.9\,\mathrm{kJ/mol} \tag{18.13}$$

$$= -48.4\,\mathrm{kJ/mol} \tag{18.14}$$

which is considerably more than under standard conditions.

18.2 If hydrolysis of ATP yields ΔG = −48.4 kJ/mol and for the entire reaction ΔG = 0 kJ/mol (equilibrium), then the transport reaction can cost +48.4 kJ/mol:

$$\Delta G = RT \ln\left(\frac{c_l}{c_c}\right) + zF\Delta E \tag{18.15}$$

$$48.4\,\text{kJ/mol} = 2.55\,\text{kJ/mol} \times \ln\left(\frac{c_l}{c_c}\right) + 3{\times}96.484\,61\,\text{kJ V}^{-1}\,\text{mol}^{-1} \times 115 \times 10^{-3}\,\text{V} \tag{18.16}$$

$$= 2.55\,\text{kJ/mol} \times \ln\left(\frac{c_l}{c_c}\right) + 33.29\,\text{kJ/mol} \tag{18.17}$$

$$\left(\frac{c_l}{c_c}\right) = \exp\left(\frac{48.4\,\text{kJ/mol} - 33.29\,\text{kJ/mol}}{2.55\,\text{kJ/mol}}\right) = 375 \tag{18.18}$$

$$\Delta p\text{H} = \log_{10}(375) = 2.57 \tag{18.19}$$

As the cytosolic *p*H is 7.2 we would expect a lysosomal *p*H of 4.6.

References

1. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of *Escherichia coli*. Science **301**, 610–615 (2003). doi: 10.1126/science.1088196
2. M.P. Blaustein, Sodium ions, calcium ions, blood pressure regulation, and hypertension: a reassessment and a hypothesis. Am. J. Physiol. **232** (5), C165–C173 (1977). URL http://ajpcell.physiology.org/content/ajpcell/232/5/C165.full.pdf
3. M.P. Blaustein, J.M. Hamlyn, Signaling mechanisms that link salt retention to hypertension: Endogenous ouabain, the Na^+ pump, the Na^+/Ca^{2+} exchanger and TRPC proteins. Biochim. Biophys. Acta **1802** (12), 1219–1229 (2010). doi: 10.1016/j.bbadis.2010.02.011
4. S. Breton, D. Brown, Regulation of luminal acidification by the V-ATPase. Physiol. **28**(5), 318–329 (2013). doi: 10.1152/physiol.00007.2013
5. E. Buxbaum, Co-operative binding sites for transported substrates in the multiple drug resistance transporter Mdr1. Eur. J. Biochem. **265**, 64–70 (1999a). doi: 10.1046/j.1432-1327.1999.00644.x
6. E. Buxbaum, Co-operating ATP sites in the multiple drug resistance transporter Mdr1. Eur. J. Biochem. **265**, 54–63 (1999b). doi: 10.1046/j.1432-1327.1999.00643.x
7. E. Buxbaum, W. Schoner, Phosphate binding and ATP binding sites coexist in Na^+/K^+-ATPase, as demonstrated by the inactivating $MgPO_4$ complex analogue $Co(NH_3)_4PO_4$. Eur. J. Biochem. **195**, 407–419 (1991). doi: 10.1111/j.1432-1033.1991.tb15720.x
8. L.C. Cantley, L. Josephson, R. Warner, M. Yanagisawa, C. Lechene, G. Guidotti, Vanadate is a potent (Na, K)-ATPase inhibitor found in ATP derived from muscle. J. Biol. Chem. **252**(21), 7421–7423 (1977). URL http://www.jbc.org/content/252/21/7421.full.pdf+html
9. M. Cereijido, R.G. Contreras, L. Shoshani, I. Larre, The Na^+-K^+-ATPase as self-adhesion molecule and hormone receptor. Am. J. Physiol. Cell Physiol. **302**(3), C473–C481 (2012). doi: 10.1152/ajpcell.00083.2011
10. B.L. deGroot, H. Grubmüller, Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF. Science **294**, 2353–2357 (2001). doi: 10.1126/science.1066115
11. S.J. Ferguson, ATP synthase: From sequence to ring size to the P/O ratio. Proc. Natl. Acad. Sci. USA **107**(39), 16755–16756 (2010). doi: 10.1073/pnas.1012260107

12. H. Fuerstenwerth, On the differences between ouabain and digitalis glycosides. Am. J. Therapeut. **21**(1), 35–42 (2014). doi: 10.1097/MJT.0b013e318217a609
13. Y. Fujiyoshi, K. Mitsoka, B.L. de Groot, A. Philippsen, H. Grubmüller, P. Agre, A. Engel, Structure and function of water channels. Curr. Opin. Struct. Biol. **12**(4), 509–515 (2002). doi: 10.1016/S0959-440X(02)00355-X
14. M. Futai, M. Nakanishi-Matsui, H. Okamoto, M. Sekiya, R.K. Nakamoto, Rotational catalysis in proton pumping ATPases: From E. coli F-ATPase to mammalian V-ATPase. Biochim. Biophys. Acta. **1817**(10), 1711–1721 (2012). doi: 10.1016/j.bbabio.2012.03.015
15. V. De Giorgis, P. Veggiotti, Glut1 deficiency syndrome 2013: Current state of the art. Seizure **22**(10), 803–811 (2013). doi: 10.1016/j.seizure.2013.07.003
16. B. Holland, *ABC Proteins – From Bacteria to Man* (Academic Press, Amsterdam, 2003). ISBN 978-0-1235-2551-2.
17. Y. Ishibashi, A. Kohyama-Koganeya, Y. Hirabayashi, New insights on glucosylated lipids: Metabolism and functions. Biochim. Biophys. Acta **1831**(9), 1475–1485 (2013). doi: 10.1016/j.bbalip.2013.06.001
18. I.D. Kerr, A.J. Haider, I.C. Gelissen, The ABCG family of membrane-associated transporters: you don't have to be big to be mighty. Brit. J. Pharmacol. **164**(7), 1767–1779 (2011). doi: 10.1111/j.1476-5381.2010.01177.x
19. M. Mueckler, B. Thorens, The slc2 (GLUT) family of membrane transporters. Mol. Aspects Med. **34**(2–3), 121–138 (2013). doi: 10.1016/j.mam.2012.07.001
20. Nagy, H. and Goda, K. and Szakács, G. and Arceci, R. and Váradi, A. and Sarkadi, B. and Mechetner, E. and Szabó, G. Fluorescence resonance energy transfer studies of function-related changes in oligomerization and conformational state of Pgp. 2nd Advanced Lecture Course "ATP-Binding Cassette (ABC) Proteins: From Multidrug resistance to Genetic Disease. FEBS, Gosau, 99 (1999)
21. D. Pavlovic, The role of cardiotonic steroids in the pathogenesis of cardiomyopathy in chronic kidney disease. Nephron Clin. Pract. **128**, 11–21 (2014). doi: 10.1159/000363301
22. K.R.H. Repke, R. Schön, Flip-flop model of (NaK)-ATPase function. Acta Biol. Med. Germ. **31**(4), K19–K30 (1973)
23. W. Schoner, G. Scheiner-Bobis, Role of endogenous cardiotonic steroids in sodium homeostasis. Nephrol. Dial. Transplant. **23**(9), 2723–2729 (2008). doi: 10.1093/ndt/gfn325
24. M. Toei, R. Saum, M. Forgac, Regulation and isoform function of the V-ATPases. Biochem **49**(23), 4715–4723 (2010). doi: 10.1021/bi100397s
25. W. Withering, *An Account of the Foxglove, and some of its Medical Uses: With Practical Remarks on Dropsy, and Other Diseases* (G.G.J. and J. Robinson, London, 1785). URL http://www.munseys.com/diskfive/foxg.pdf

Appendix A: Short Biographies of Scientists Mentioned in This Book

ALPER, TIKVAH South Africa, 1909–1995. Born as the fourth daughter of an immigrant from Russia, she obtained a scholarship to study physics at Capetown University, where she received her M.A. at the age of 20. She then moved to Germany, where she worked on the δ-rays from α-particles with L. MEITNER but failed to obtain a PhD because growing anti-Semitism in Germany forced her to return to South Africa (she was awarded a DSc by London University in 1969). She married the bacteriologist MAX STERN in 1932. Their son JONATHAN was born deaf, so she moved to St. Louis (MO) to receive training as a teacher for the deaf, in which capacity she worked until she became a lecturer at the University of Witwatersrand and later head of the biophysics unit at the National Physics Laboratory. In 1951 she signed a circular against growing apartheid and was forced to emigrate to England, where she obtained a grant at the MRC Experimental Radiopathology Research Unit at Hammersmith Hospital. In 1953 she became an official staff member, and in 1962 director of this institution, from which she retired in 1973. She was the first to demonstrate that the causative agent for prion diseases does not contain DNA.

ANFINSEN, CHRISTIAN BOEHMER United States, 1916–1995, He received his BA in 1937 (Swathmore College), his MS in organic chemistry in 1939 (University of Pennsylvania), and his PhD in 1943 from Harvard Medical School. He served in various positions at Harvard Medical School, John-Hopkins University, and the National Institutes of Health. He received the Nobel Prize in Chemistry in 1972 for his discovery that proteins fold spontaneously (at least in principle) and that their entire three-dimensional structure is encoded in their amino acid sequence.

BEHRING, EMIL ADOLF VON German, 1854–1917. As the eldest of 13 children of a schoolmaster he had no funds to study, thus he entered the Army Medical College in Berlin in 1874, obtained his medical degree in 1878, and passed his licensing exam in 1880. He was then sent to Posen (now in Poland), where he worked as an army physician and at the same time continued scientific studies at the Chemical Department of the Experimental Station, working on

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

septic diseases. He found that iodoform (CHI_3) could neutralise bacterial toxins (published 1882). Behring was then sent to C. BINZ (Bonn) for further training in pharmacology. In 1888 he was ordered back to Berlin, where he became an assistant to R. KOCH at the Institute for Hygiene and later the Institute for Infectious Diseases (now known as the ROBERT-KOCH-Institute), working in a team that included amongst others P. EHRLICH. In 1894 BEHRING became professor of hygiene at the University of Halle, moving to Marburg the following year. He is remembered mostly for his work on diphtheria, tuberculosis, and tetanus. He found (together with S. KITASATO) that the supernatants of cultures of diphtheria and tetanus bacteria, when injected into animals in carefully determined doses, led to the production of *antitoxins* (which now we call antibodies) in their blood, which were able to neutralise the deadly toxins made by these bacteria. The antitoxins produced in one animal could be used to save other animals infected with the bacteria (passive immunisation). BEHRING and F. WERNICKE demonstrated that toxin-antitoxin complexes could also be used for active immunisation of animals, at much lower risk to their well-being than the pure toxins. Following a suggestion of T. SMITH such complexes were used for immunisation campaigns in man. Production of these complexes was done in the **Behringwerke**, a world-renowned company specialising in vaccines (founded in 1914). BEHRING became *Geheimer Medizinalrat* in 1895, and was raised to nobility in 1901, the year he received the NOBELPrize for Physiology and Medicine for his ground-breaking and life-saving work.

BERZELIUS, JÖNS JAKOB Swedish, 1779–1848, studied medicine at the University of Uppsala and wrote a thesis on the effect of electricity on humans. He became professor of medicine and pharmacy at the Medical College of Stockholm in 1807. He is considered one of the greatest chemists of all times: he discovered the law of constant proportions; the elements Ce (1804), Se (1817), and Th (1828); and was involved in the discovery of Li, V, and several rare earth metals. He invented chemical formulas (1814) and was the first to understand ionic compounds and coined the name "amino acid".

BLOBEL, GÜNTHER German, born 1936. He fled the Soviet Occupation Zone of Germany in 1954 and studied medicine at the University of Frankfurt/Main, Kiel, Munich, and Tübingen, graduating in 1960. In 1962 he moved to Montreal, where he obtained a PhD in 1967 under V. R. POTTER. He did his postdoctoral studies under G. PALADE at Rockefeller University, becoming assistant professor in 1969, associate professor in 1973, and full professor in 1976. He received the Nobel Prize in Physiology and Medicine in 1999.

BOHR, CHRISTIAN Danish, 1855–1911, studied medicine in Copenhagen, and received his MD in 1886. He discovered the effect of *p*H-reduction and carbon dioxyde binding on the oxygen affinity of hæmoglobin. He was also the father of the physicist NIELS and the mathematician HARALD BOHR.

BOYER, PAUL D. United States, born 1918. He received his BA in chemistry from Brigham Young University, and his PhD in biochemistry from the University of Wisconsin in 1943. His postdoctoral training was at Stanford University, and he became assistant professor of chemistry at the University of Minnesota in 1945,

and in 1963 at UCLA. In 1971 he discovered that the energy of the proton flux is not used to make ATP, but to release it from the ATP-synthase. In the following years he found the cooperation between the three catalytic binding sites of ATP-synthase and the concept of rotational synthesis. He received the Nobel Prize in 1997 for his studies on ATP synthase, together with J. E. WALKER AND J. C. SKOU.

BUCHNER, EDUARD German, 1860–1917, studied chemistry and botany in Munich and Erlangen, and received his PhD in 1888, becoming a professor of chemistry 1895. He received the Nobel Prize in Chemistry in 1907 for his discovery (made together with his brother HANS, a well-known bacteriologist) of noncellular fermentation in yeast extracts.

BUCHNER, HANS German, 1850–1902. He received his MD in 1874, and became professor of hygiene in 1892. He recognised the importance of serum proteins for immune defence and invented techniques for studying anaerobic bacteria. Together with his brother EDUARD he discovered noncellular fermentation by yeast extracts. He died before his brother received the NOBEL Prize for this discovery.

BURK, DEAN United States, 1904–1988

CECH, THOMAS ROBERT United States, born 1947. He studied at the University of California in Berkeley (doctorate 1975), then went to MIT (1975–1977) and the University of Colorado (1978; since 1983 as professor of chemistry and biochemistry as well as for cell and developmental biology). Since 1987 he has been research professor at the American Cancer Society. He discovered that the splicing of mRNA is done by the mRNA itself, in other words, that some RNA molecules can act as enzymes. For this discovery he received the NOBEL Prize for Chemistry in 1989.

CHANCE, BRITTON United States, 1913–2010. He received his BA (1935), MA (1936), and PhD in physical chemistry (1940) from the University of Pennsylvania, and his second PhD in physiology from Cambridge University in 1942. He became professor of biophysics at the University of Pennsylvania in 1949. He invented methods to investigate the initial (pre-steady-state) phase of enzyme reactions, and proved the existence of the ES-complex by measuring the absorption spectrum of the hæme group in peroxydase upon binding of H_2O_2. He also pioneered the numerical simulation of metabolic pathways. In addition to his scientific interests, he won an Olympic gold medal in sailing in 1952.

CLAUDE, ALBERT Belgian, 1899–1983. The early death of his mother from breast cancer left a lasting impression on the seven-year-old boy, even though he later claimed that it did not cause him to study medicine. He obtained a medical degree from the Université de Liège in 1928, then went to the *Institut für Krebsforschung* and the *Kaiser-Wilhelm-Institute Dalem* in Berlin. In 1929 he moved to the Rockefeller Institute, where he spent the rest of his career, although he held additional posts at the University of Brussels and the Catholic University of Louvain. One of the fathers of cell biology, he started to use the electron

microscope to study biological questions. He identified the mitochondrium as the site of oxydation. In 1974 he received the Nobel Prize for Physiology and Medicine, together with C. DE DUVE & G. E. PALADE.

CLELAND, WILLIAM WALACE "MO" United States, 1930–2013. After receiving his AB (Oberlin College) in 1950, his MS and PhD (University of Wisconsin-Madison) in 1953 and 1955, he did his postdoctoral work at the University of Chicago, becoming assistant professor at UW-Madison in 1959, associate professor in 1962, full professor in 1966, and then M. J. JOHNSON professor of biochemistry in 1978 and STEENBOCK professor of chemical science in 1982. He has been a member of the National Academy of Science since 1985. He has made significant contributions to the study of enzyme reaction mechanism by kinetic methods, pioneered the use of isotope effects on enzyme activity and invented DTT as reagent to reduce protein disulphide bonds.

CORNISH-BOWDEN, ATHELSTAN JOHN English, born 1943. He received his BA in 1965, MA in 1967, and PhD in 1967.

CREUTZFELDT, HANS GERHARD German, 1885–1964, A neuropathologist, he studied at the Universities of Jena and Rostock, MD 1909. He worked as a ship's doctor before continuing his scientific career in Hamburg, Frankfurt/M, Breslau, Berlin, Munich, and Kiel (habilitation 1920, 1925 extraordinary, 1938 ordinary professor). Under the Nazi regime he rescued some people from their euthanasia and "racial hygiene" programs, but as expert witness he was involved in some executions. The British occupation forces dismissed him as director of the University of Kiel because they did not agree with his plans for rebuilding the University. From 1953 he worked at the MAX-PLANK-Institute in Munich until his death.

DAVY, HUMPHRY English, 1778–1829. He was a chemist who showed in 1816 that platinum wire could catalyse the reaction of alcohol vapour with air oxygen, without being used up.

DONNAN, FREDERICK GEORGE British, 1870–1956. He was born in Colombo (Ceylon, now Sri Lanka) as the son of Northern Irish parents. At age three he returned to Ireland. He studied science at Queens College, Belfast (BSc 1893, BA 1894), Leipzig (PhD 1896), Berlin, London, and the Royal University of Ireland in Dublin (MA 1897), working amongst others with VAN'T HOFF, OSWALD, and RAMSAY. He taught chemistry in various positions in London, Dublin, and Liverpool until he retired in 1937. He helped emigrant scientists from Germany during the Nazi rule and contributed to the production of ammonia and nitric acid during the Second World War as a consultant to the Ministry of Munitions. His most significant scientific contribution was the theory of diffusion equilibria where some ionic species can cross a semipermeable membrane whereas others can't (1911). The resulting electrochemical potential (found across the membrane of every living cell) bears his name. He is also considered one of the fathers of colloid chemistry and made significant contributions to our understanding of detergent action.

DUCLAUX, PIERRE ÉMILE French, 1840–1904. He was director of the Pasteur Institute from 1895 to 1904.

DUCHENNE DE BOULOGNE, GUILLAUME BENJAMIN AMAND French, 1806–1875. He studied medicine in Paris before becoming a practitioner in his hometown of Boulogne (1831). Two years later he invented electrotherapy and electrodiagnostics, which he continued to elaborate for the rest of his life. He described several diseases: DUCHENNE muscular dystrophy, *tabes dorsalis*, tabetic locomotor ataxia, and progressive bulbar paralysis, and also worked on the consequences of lead poisoning and acute poliomyelitis.

DE DUVE, CHRISTIAN Belgian, born 1917. He studied medicine at the Catholic University of Louvain from 1934, and received his MD in 1941. He became army doctor, was captured and fled from the German prisoners camp, returning to Louvain where he worked as an intern and at the same time studied chemistry inasmuch as the war prevented him from continuing his research on insulin action. He undertook further studies in 1946–1947 at the Medical Nobel Institute in Stockholm with H. THEORELL, Washington University with the CORIS, and St. Louis with E. SUTHERLAND. He returned to Louvain in 1947 as teacher of physiological chemistry, becoming full professor in 1951. During the following years he developed techniques for the isolation of cell organelles, leading to the discovery of the lysosome and the peroxisome. These techniques are still in routine use today. He was awarded the NOBEL Prize for Physiology and Medicine in 1974, together with A. CLAUDE & G. E. PALADE.

EDMAN, PEHR VICTOR Swedish, 1914–1977. He studied medicine at the Karolinska-Institute (MD in 1947 after an interruption by World War II, which he spent in the Medical Corps). His postdoctoral training was at the Rockefeller Institute of Medical Research in Princeton, then becoming assistant professor at the University of Lund, where he published his first paper on the determination of protein sequences (now called EDMAN-degradation) in 1950. He became director at the St. Vincent's School of Medical Research in Australia in 1957, where he developed the first automated protein sequencer in 1967 (with G. BEGG). After 1972 EDMAN continued his work at the MAX PLANCK-Institut for Biochemistry in Martinsried (near Munich, Germany), getting ever longer sequences from ever smaller samples.

ENGELHART, JOHANN FRIEDRICH PHILIPP German, 1797–1857. He obtained his DPhil at the University of Göttingen in 1825 with a thesis on hæmoglobin, showing that the iron content of hæmoglobin was identical in all species. This thesis represents the birth of macromolecular chemistry. He became professor of chemistry at the *Gewerbeschule zu Nürnberg* in 1829. He also rediscovered an ancient method to stain glass red with CuO.

EULER, LEONHARD Swiss, 1707–1783. He was born into a family of distinguished scholars. He studied under BERNOULLI at the University of Basel (1720) where he obtained a magister in 1723. He moved to St. Petersburg (1727), Berlin (1741), and back to St. Petersburg (1766). He discovered fundamental rules of calculus and was one of the first to apply maths to physical, economic, and musical problems. Euler is considered one of the greatest scientists of all times, despite the fact that he had severe eye problems and became blind in 1771.

FARADAY, MICHAEL British, 1791–1867, started his professional life at age 14 as a bookbinder's apprentice, becoming an assistant to HUMPHREY DAVY later. In 1831 he started with a series of experiments on electromagnetism; that led him to the invention of the dynamo and electric motor. He discovered benzene, the laws of electrochemistry, and invented oxydation numbers and the FARADAY-cage. In 1845 he discovered that magnetic fields could turn the direction of polarised light (FARADAY-effect).

FENN, JOHN BENNETT United States, 1917–2010. He trained at Berea College and Allied Schools in Berea, Kentucky (high school diploma 1932, and graduated in chemistry in 1938. Then he moved to Yale, working as an assistant while studying physical chemistry, obtaining a PhD in 1941. After several industrial positions he became an academic staff member at Princeton University working on combustion and molecular beams; later he moved to Yale and Richmond. He received the NOBEL Prize in Chemistry 2002 for the development of electrospray ionisation (with KOICHI TANAKA and KURT WÜTHRICH).

FICK, ADOLF EUGEN German, 1829–1901. He discovered the diffusion law named after him (in 1855), described the FICK-principle (1870, cardiac output = oxygen consumption/arteriovenous oxygen difference), and invented the contact lens (1887).

FISCHER, HERMANN EMIL German, 1852–1919. He studied in Bonn and Straßburg (now Strasbourg), receiving his PhD in 1874, habilitation 1878, professor of chemistry in Munich 1879, Erlangen 1881, and Würzburg in 1888. He worked on dyes and hydrazines, and described the structures and metabolism of purines. Between 1882 and 1906 he worked on the stereochemistry of sugars and glucosides. He separated and identified the amino acids in proteins (1899–1908) and discovered the peptide bond. In 1902 he was awarded the Nobel Prize in Chemistry. The German Chemical Society has named its medal after him.

GERSTMANN, JOSEPH Austrian, 1887–1969. He was a neuropsychiatrist who studied medicine in Vienna, graduating in 1912. As Jews he and his wife had to flee to the United States in 1938, where he worked in Springfield, Washington and New York.

GIBBS, JOSIAH WILLARD United States, 1839–1903. He studied at Yale (PhD 1863) and became professor of mathematical physics there in 1871. His work on the driving force behind chemical reactions (thermodynamics and statistical mechanics) initially received little attention because he published it in obscure places.

GOLGI, CAMILLO Italian, 1843–1926. He studied medicine at the University of Pavia, graduating in 1865. He continued to work at the Hospital St. Matteo in Pavia, specialising in neurological diseases. In 1872 he became chief medical officer in the Hospital for the Chronically Sick in Abbiategrasso, where he converted a kitchen into a laboratory. He became extraordinary professor for histology at the University of Pavia and later chair of general pathology (1881). He identified the three different forms of malaria and their pathogens, and in 1890 succeeded in obtaining photographs of them. He discovered silver staining of nerve cells, the tendon sensory organ, and the GOLGI apparatus. He received

the Nobel Prize for Physiology and Medicine 1906 together with SANTIAGO RAMÓN Y CAJAL for studies on the nervous system.

GROTTHUS, CHRISTIAN JOHANN DIETRICH THEODOR VON German, 1785–1822. He studied in Leipzig, Paris, and Naples. From 1808 he lived as a private scholar on the family estate in Lithuania. He worked on electrolysis in 1806 and photochemistry in 1817.

HALDANE, JOHN BURDON SANDERSON Scottish, 1892–1964, educated in Eton and Oxford. Member of the communist party, which he left after the LYSENKO-disaster. However, he was never happy with the politics of his native country and sought refuge in India 1957, where he also died. He was the founder of population genetics, and worked on applying mathematical principles in biology, thus laying the foundations for our modern understanding of evolution. He coined the word "clone" and contributed to enzyme kinetics.

HENRI, VICTOR French/Russian. 1872–1940. His father NIKOLAI ALEXANDROVITCH KRYLOV lived with his wife SOFIA VICTOROVNA and her sister ALEXANDRA VICTOROVNA LIAPUNOV in a happy *menage a trois*. When ALEXANDRA became pregnant, the triple moved to Marseille where Victor was born. His name HENRI was assigned by the registrar, as was usual in France in cases of doubtful paternity. NIKOLAI and SOFIA then adopted the "orphan", so he could grow up with his family and have full citizenship rights, not given to illegitimate children in Russia at the time. He studied psychology at the Sorbonne and in Göttingen, Germany (there DPhil 1897). He worked at the Sorbonne on memory (with BINET), in Leipzig (from 1892) with KÜLPE and TAWNEY on touch perception and on intelligence, and in Göttingen (from 1896) with his wife CATHERINE on childhood recollections. He then studied physiology under DASTRE, obtaining a doctorate in 1903. He went to Russia in 1915, where he worked in the chemical industry with a grant from the French government (Russia and France were allies during WWI). In 1917–18 he was professor of physiology at the University of Moscow, and in 1920–30 professor of physical chemistry at the University of Zurich, Switzerland. He worked as director of the Berre refinery in Marseille 1930–31, before he became professor of physical chemistry at the University of Liège (Belgium) 1931–40. He worked on enzyme kinetics (first formulation of the HMM-equation 1902), photochemistry, and spectroscopy.

HILL**,** SIR ARCHIBALD VIVIAN English, 1886–1977. Supported by several fellowships, he studied maths and science at Trinity College, Cambridge (1905–1909). In 1909 he started experimental work in the lab of J. N. LANGLEY. His principal interest was muscle contraction and exercise physiology, a field in which he continued to work throughout his life, using calorimetric methods which he learned during his stay with BURKER and PASCHEN in Tübingen, Germany (1911). He improved these so much that in 1937 he was able to measure temperature differences of 0.000 15 °C. In addition he worked on nerve physiology, hæmoglobin (resulting in the famous HILL-equation), and antiaircraft munitions development (during World War I). He became lecturer for physical chemistry at Cambridge 1914, and professor of physiology in

Manchester 1920, followed by various positions in London. During the war years 1940–45 he served as a member of Parliament (Independent Conservative for Cambridge) and as a member of the War Cabinet Scientific Advisory Committee. Like DONNAN he was a member of the Academic Assistance Council which helped refugees from Nazi Germany to relocate in the West. He received the NOBEL Prize for Physiology and Medicine in 1922 together with his collaborator, the German biochemist O. MEYERHOFF for the discovery that muscle contraction can be fuelled by both aerobic and anaerobic metabolism.

HOFMEISTER, FRANZ Austrian, 1850–1922. He studied science and medicine under ERNST MACH in Prague and under E. HERING in Leipzig from 1870 and obtained his doctorate in 1879 under H. HUPPERT. He was the first to identify lactose in glycosuria of pregnancy (1877). He received his *Venia legendi* for medicinal chemistry 1879 with work on peptones. He went to O. SCHMIEDEBERG in Straßburg for specialist training and then returned to Prague (1883) as professor for experimental pharmacology (ordinarius from 1885). In 1896 he went to Straßburg again as successor of the late F. HOPPE-SEYLER. He was the first to suggest the nature of the acid amide (peptide) bond, which was then confirmed by E. FISCHER. He worked on the enzyme theory of intermediary metabolism, importance of vitamins for nutrition, buffering of blood, and HOFMEISTER's caotropic series of ions. Evicted from Straßburg in 1919 (after it became Strasbourg) he went to Würzburg, where he died only three years later.

HOPPE-SEYLER, ERNST FELIX IMMANUEL German, 1825–1895. Born as E. F. I. HOPPE he became an orphan early and was adopted by his brother-in-law, the priest Dr. SEYLER. He studied medicine at the Universities of Halle (1846–1847), Leipzig (1847–1850), and Berlin (1850), where he wrote his thesis on cartilage structure. After short stays in Prague and Vienna he started as a general practitioner in Berlin 1852, then became research assistant at the University of Greifswald 1854 where he was habilitated in 1855. Prosector and head of the Chemical Laboratories at the Pathological Institute at the Charité Clinic in Berlin 1856 (under VIRCHOW), he became professor of applied chemistry at the University of Tübingen 1860–1871, then professor of physiological chemistry at the University of Straßburg (now Strasbourg) from 1872 until his death. He became rector in 1873, founded the first journal for physiological chemistry (*Zeitschrift für Physiologische Chemie*, now *Biological Chemistry Hoppe-Seyler*) 1877, and inaugurated the first independent institute for physiological chemistry in Germany 1884. His research work included fermentation, lipid and bile metabolism, urine composition, and the quantification and classification of proteins. Together with his students FRIEDRICH MIESCHER and ALBRECHT KOSSEL he discovered and characterised DNA. He demonstrated reversible oxygen binding to hæmoglobin (a word he coined) by spectroscopic methods and found the pathomechanism of carbon monoxyde and hydrogen sulphide poisoning. He was the first to crystallise a protein (hæmoglobin) in pure form.

JACOB, FRANÇOIS French, born 1920. He started studying medicine at the Faculty of Paris in 1939 but his education was interrupted by the Second World War,

during which he served in the Free French Forces in North Africa and in Normandy, where he was severely wounded. He continued his studies (MD 1947, MSc 1951, PhD 1954). In 1950 he joined A. LWOFF at the Institute Pasteur, becoming laboratory director in 1956 and head of the department of cell genetics in 1960. In 1964 he became chair of the department for cell genetics at the Collège de France. His scientific work centred on the genetics of bacteria and prophages; he discovered bacterial conjugation (with E. WOLLMANN), the circularity of the bacterial chromosome, mRNA, and the concepts of operons and allostery (with J. MONOD). In addition he has worked on cell division in cultured mammalian cells (with S. BRENNER). He received the NOBEL Prize for Physiology and Medicine in 1965 together with A. LWOFF & J. MONOD.

JAKOB, ALFONS MARIA German, 1884–1931. Neurologist, was the first to describe ALPER's disease and CREUTZFELDT-JAKOB-disease.

JERNE, NIELS KAY English, 1911–1994. He became professor of biophysics in Geneva in 1960, in 1962 head of microbiology at the University of Pittsburgh, in 1966 director of the Paul-Ehrlich-Institute in Frankfurt/M, and was director of the Institute for Immunology in Basel 1969–1980. Most well known for his work on the time course of the antibody response (published in 1955), for which he received the NOBEL Prize for Physiology and Medicine in 1984, together with C. MILSTEIN & G. J. F. KÖHLER.

KENDREW, JOHN COWDERY English, 1917–1997. Studied chemistry in Clifton College (Bristol) and Trinity College (Cambridge), BA 1939. During the war he did research on radar. Returned to Cambridge 1946, PhD 1949, ScD 1962. He established the 3D-structure of myoglobin at 6 Å resolution in 1957. He received the Nobel Prize in Chemistry 1962 together with M. PERUTZ.

KIRCHHOFF, GOTTLIEB SIGISMUND CONSTANTIN German, 1764–1833. Learned pharmacy from his father, who ran a pharmacy in Teterow. Then moved to St. Peterburg (Russia) where he became an employee, and later the director of the royal pharmacy. In addition to his duties there he was an important technical chemist, working on ceramics, fire proofing of wood, and technical uses of potatoes. In the course of the latter studies he found in 1812 that dilute, boiling sulphuric acid could turn insoluble potato starch into soluble glucose, without being used up in the process. This was one of the first demonstrations of catalysis. In 1815 he reported that wheat protein had the same property.

KÖHLER, GEORGES JEAN FRANZ German, 1945–1995. Studied Biology at the University of Freiburg, receiving his degree in 1971 and his PhD in 1974. His thesis work on β-galactosidase he performed under F. MELCHERS in Basel. For postdoctoral studies he went to C. MILSTEIN in Cambridge, where both developed a method for obtaining monoclonal antibodies which would earn them the NOBEL Prize in Physiology and Medicine together with N.K. JERNE in 1984. In 1975 KÖHLER returned to Basel, where he worked at the Institute for Immunology until his untimely death in 1995.

KOSHLAND, DANIEL. E., JR. United States, 1919–2007. BS in chemistry from the University of California, Berkeley (1941), worked on plutonium purification for the Manhattan Project (1941–46), then received his PhD in organic chemistry

from the University of Chicago (1949). He was professor of biochemistry at the University of California, Berkeley (1965), and worked on enzymology (induced fit model of substrate binding) and metabolic regulation (role of protein kinases and phosphatases). He is considered one of the most influential scientists of the United States, editor of *Science* (1985–95) and of several other journals, member of the Council of the National Academy of Sciences, and chairman of the editorial board of its *Proceedings*.

KREBS, SIR HANS ADOLF German, 1900–1981. Studied medicine in Göttingen, Freiburg/Breisgau, Berlin, and Hamburg (MD 1925), completing his training with one year of chemistry in Berlin. He became assistant to O. WARBURG in Berlin-Dahlem (1926–1930), and then returned to clinical work. In 1933 he fled from Nazi Germany to England where he became demonstrator of biochemistry at the University of Cambridge in 1934. In 1935 he moved to Sheffield as lecturer, becoming full professor in 1945. In 1954 he moved to Oxford. His main work was on energy transformations in living cells, discovering what is now called the KREBS-cycle. He also discovered the urea cycle and worked on membrane transport and the synthesis of uric acid and purines. He was awarded the NOBEL Prize for Physiology and Medicine 1953 and a knighthood in 1958.

KÜHNE, WILLY German, 1837–1900. Professor of physiology in Heidelberg, he isolated trypsin in 1876. Following the vitalistic view of his time he postulated that biosynthesis can be performed only by the organised ferments inside cells, and suggested the name "enzyme" for the purified preparations in a test tube.

KUPFFER, KARL WILHELM VON German, 1829–1902. He obtained a degree from the medical school at Dorpatt (now Tartu), then worked as general practitioner until he started further training in physiology in Vienna, Berlin, and Göttingen (1856–1857). He became professor of anatomy in Kiel (1866; there he discovered the star-cells in the liver, which bear his name), Königsberg (1875), and Munich (1880, where he retired in 1901).

LINEWEAVER, HANS United States, 1908–2009. He received his BA in chemistry 1930 and MA 1933 from GEORGE-WASHINGTON University, and PhD in physical chemistry in 1936 from JOHNS-HOPKINS University. He worked in various capacities for the US Department of Agriculture for his entire career. He reinvented the double-reciprocal plot (originally proposed by W. WOOLF) while in the lab of DEAN BURK.

LWOFF, ANDRÉ MICHEL French, 1902–1994. Joined the Institute Pasteur at the age of 19, while studying medicine. He received his MD in 1927, and PhD in 1932 with work on the development cycle and nutrition of ciliates. He spent one year (1932–33) in Heidelberg with OTTO MEYERHOFF, then seven months in Cambridge. He became departmental head at the Institute Pasteur in 1938 and professor of microbiology at the Science Faculty in Paris in 1959 where he studied lysogenic bacteria and polio virus. For his work he received the NOBEL Prize in Physiology and Medicine in 1965, together with his collaborators F. JACOB & J. MONOD.

MENKES, JOHN HANS Austrian, 1928–2008. His family had to flee Vienna in 1939, first to Ireland and later to Los Angeles. He received his BA(1947) and

MA (1951) in organic chemistry at USC. Following the family tradition he studied medicine (Johns Hopkins University (MD 1952) and Harvard University), specialising in pediatric neurology. In 1951 as intern at Boston Children's Hospital he described maple syrup urine disease, managing to isolate branched chain amino acids from the urine of affected children several years later (with a budget of US$ 35.00). As resident at Johns Hopkins University in 1962 he described the disease of Cu-metabolism that is named after him. In 1966 he became the founding head of the UCLA division of pediatric neurology; later he worked at the Veterans Administration hospital in Brentwood. His *Textbook of Child Neurology* (1st edition 1974) became a standard textbook in the field. Finally he entered private practice and served as expert witness in some trials about children allegedly damaged by diphtheria-pertussis-tetanus vaccine. Apart from scientific texts, he wrote three dramas and three novels.

MENTEN, MAUDE LEONORA Canadian, 1879–1960. She was one of the first women to have a scientific career. She studied medicine at the University of Toronto (BA 1904, MB 1907, MD 1911). For her thesis work she had to go to the University of Chicago, as at that time woman were not allowed to do research in Canada. In 1912 she moved to Berlin where she worked with L. MICHAELIS, obtaining a PhD in 1916. She worked as pathologist at the University of Pittsburgh (1923–1950, rising to the rank of full professor) and as research fellow at the British Columbia Medical Research Institute (1951–1953). Apart from her work on enzyme kinetics together with L. MICHAELIS based on earlier findings of V. HENRI (resulting in the famous HENRI-MICHAELIS-MENTEN-equation 1913) she characterised bacterial toxins (*B. paratyphosus, S. scarlatinae* and *Salmonella ssp.* 1924 with HELEN MANNING), invented the azo-dye coupling reaction for alkaline phosphatase (1944 with JUNGE & GREEN, still used in histochemistry), and (in 1944 with ANDERSCH & WILSON) characterised fetal and adult hæmoglobin by electrophoresis and ultracentrifugation. She also worked on regulation of blood sugar level and kidney function. Despite suffering from arthritis she was an accomplished musician and painter; her paintings were exhibited in several galleries.

MERRIFIELD, BRUCE United States, 1921–1984. He graduated in chemistry from UCLA in 1942, and received his PhD in 1949. He moved to the Rockefeller Institute for Medical Research in New York where he developed the solid phase synthesis of peptides in 1959; for this discovery he received the Nobel Prize in Chemistry in 1984.

MICHAELIS, LEONOR German, 1875–1947. Studied medicine in Freiburg, graduating in 1897, and then moved to Berlin, where he received his doctorate the same year. He worked as assistant to Paul Ehrlich (1898–1899), Moritz Litten (1899–1902), and Ernst Victor von Leyden (1902–1906). In 1906 he started as director of the bacteriology lab in Berlin's Charitè clinic, becoming professor extraordinary at Berlin University in 1908. In 1922 he moved to the Medical School of the University of Nagoya (Japan) as professor of biochemistry, in 1926 to John Hopkins University in Baltimore (United States) as resident lecturer in medical research, and in 1929 to the Rockefeller Institute of Medical Research

in New York, where he retired in 1941. Besides his role in the formulation of the Henri-Michaelis-Menten law (1913) he discovered Janus Green as a supravital stain for mitochondria and the Michaelis-Gutman body in urinary tract infections (1902) and found that thioglycolic acid could dissolve keratin which made him the father of the permanent wave.

MILSTEIN, CÉSAR Argentinean, 1927–2002. He studied biochemistry in Buenos Aires, obtaining his degree in 1952 and his PhD in chemistry in 1957 with a thesis on the kinetics of aldehyde dehydrogenase. He then went on a fellowship to M. DIXON in Cambridge (1958–61), where he worked with F. SANGER on protein translation and obtained another PhD. After returning to Buenos Aires he worked on the mechanism of phosphoglyceromutase and alkaline phosphatase, but had to leave the country during the political unrest in 1963. He went to Cambridge again, where he worked at the MRC Laboratory of Molecular Biology under F. SANGER, on whose suggestion he shifted the focus of his research from biochemistry to immunology. In 1983 he became the head of its Protein and Nucleic Acid Chemistry Division. He is best known for the development of monoclonal antibodies (together with his postdoc, G. J. F. KÖHLER). In 1984 he was awarded the NOBEL Prize for Physiology and Medicine together with N. K. JERNE & G. J. F. KÖHLER.

MITCHELL, PETER DENIS British, 1920–1992. He studied in Cambridge, BA in 1942 (with very mixed results), and PhD 1951 with a thesis on the mode of action of penicillin. He worked as demonstrator in Cambridge; in 1955 he moved to Edinburgh University (senior lecturer in 1961, reader 1962). From 1965 on he used his own house in Cornwall as a research laboratory, embarking on a research programme that led him to formulate his chemosmotic hypothesis of oxydative phosphorylation. The scientific debate that followed is known as "the OxPhos-war" and became so heated that major journals refused to publish any more papers on the subject. This animosity was aggravated both by the initially scant evidence for the chemosmotic hypothesis and by the difficult personality of P. MITCHELL. However, this discussion was also considered so productive that P. MITCHELL was awarded the Nobel Prize in Chemistry in 1978, long before his hypothesis became widely accepted.

MONOD, JAQUES LUCIEN French, 1910–1976. He started studying science in Paris in 1928, obtained his degree in 1931, and his PhD in 1934. After a year at the California Institute of Technology in 1936 and the war he joined the Institute Pasteur as laboratory director under A. LWOFF. In 1954 he became director of the department of cell biochemistry, in 1959 professor of metabolic chemistry at the Sorbonne, in 1967 professor at the College de France and in 1971 director of the Institute Pasteur. He was very interested in quantitative description of biological processes and is most famous for his work on the regulation of gene expression and on cooperativity in enzymes. He received the NOBEL-Prize in Physiology and Medicine in 1965, together with his collaborators A. LWOFF & F. JACOB.

MULDER, GERARDUS JOHANNES Dutch, 1802–1880. He studied medicine in Utrecht and worked as professor of chemistry in Rotterdam and Utrecht. His work on the nature and composition of proteins was seminal.

NORTHROP, JOHN HOWARD United States, 1891–1987. He studied chemistry and zoology at Columbia University from 1908, BA 1912, MA 1913, and PhD 1915. He started working at the Rockefeller Institute in 1917, and became professor of bacteriology and biophysics at the University of California in 1949. He isolated and crystallised pepsin in 1929, followed by trypsin, chymotrypsin, carboxypeptidase, and pepsinogen. Later he worked on bacteria and virus (managing to crystallise the first virus in 1938), and enzyme kinetics and during World War II on the detection of nerve gases. He received the Nobel Prize in Chemistry for protein crystallisation in 1946, together with W. M. STANLEY & J. B. SUMNER.

PALADE, GEORGE EMIL Romanian, 1912–2008. Studied medicine at the Carol Davila School of Medicine in Bucharest (MD 1940). He served as physician in the Romanian army during WWII and moved to New York University as postdoc in 1946 and to the Rockefeller Institute the following year. In 1973 he moved to Yale. He worked on cell fractionisation and electron microscopy, initially in the group of A. CLAUDE. He worked on the fine structure of mitochondria, ER, and synapses, discovered the ribosome and characterised the secretory pathway. He can thus be regarded as one of the fathers of modern cell biology. He was awarded the NOBEL Prize for Physiology and Medicine 1974, together with A. CLAUDE & C. DE DUVE.

PAULING, LINUS United States, 1901–1994. Studied science at the Oregon Agricultural College (now Oregon State University, BSc in chemical engineering 1922) and California Institute of Technology (PhD in chemistry 1925, minors in physics and maths), where he continued as an academic staff member (assistant 1927, associate 1929, and full professor 1931). His field of study was the application of physical principles to chemical problems. In our context relevant is his work on the secondary structure of proteins, on hæmoglobin and on the antigen-antibody reaction. He won the NOBEL Prize for Chemistry in 1954 for his studies on *The Nature of the Chemical Bond* (the title of his most important book, published in 1939) and the NOBEL Prize for Peace 1962 for his stand against nuclear weapons, which makes him the only person ever to win two unshared NOBEL Prizes. Both the personality of PAULING and the problems of his time shine through his following narration: "A couple of days after my talk, there was a man in my office from the FBI saying, 'Who told you how much plutonium there is in an atomic bomb?' And I said, 'Nobody told me, I figured it out'." PAULING will be remembered not only for his scientific achievements, but also for his effort in health education (antismoking, correct nutrition). To the lay public he became known as an advocate for high-dose vitamin C supplementation, taking up to 18 g per day (the recommended dose is 60 mg).

PAYEN, ANSELME French, 1795–1871. He studied at the École Polytechnique, became director of a chemical factory at the age of 20 and professor of technical

and agricultural chemistry in 1835. He invented a synthesis of borax from boric acid and soda, breaking a Dutch monopoly on borax imported from East India. He also invented the use of activated charcoal to decolourise sugar, methods to produce starch and alcohol from potatoes, and an analytical procedure for nitrogen determination. From malt solution he, together with J. F. PERSOZ, isolated the first enzyme, diastase, by alcohol precipitation in 1833, showing it to be heat labile. In 1834 he discovered cellulose.

PERSOZ, JEAN-FRANÇOIS French, 1805–1869. He started his career as an apothecary apprentice, then became an assistant to J. L. THENARD at the Collège de France in Paris. He became professor of chemistry at the Faculty of Science at Strasbourg 1833 and later moved to Paris as professor of technical chemistry. He is best known for his book, "Traité théorique et pratique de l'impression des tissus" (1846). Together with A. PAYEN he isolated the first enzyme, diastase from barley malt (1833), and demonstrated its presence in saliva.

PERUTZ, MAX FERDINAND Austrian, 1914–2002. He started studying chemistry at Vienna University in 1932, then moved to Cambridge in 1936 where he stayed for the rest of his career. He became research assistant to Sir LAWRENCE BRAGG in 1939 and head of the Medical Research Council Unit of Molecular Biology in 1947, with J. C. KENDREW as his only staff. His work on the crystal structure of hæmoglobin started in 1937 and continued until 1959. During that time he solved the phasing problem by isomorphous replacement with heavy atoms (1953). He received the Nobel Prize in Chemistry in 1962.

POMPE, JOANNES CASSIANUS Dutch, 1901–1945. He studied medicine at the University of Amsterdam (MD 1936 with a thesis on what is now called POMPE disease). As a member of the Dutch resistance he was captured by German troops and executed.

PRUSINER, STANLEY BENJAMIN United States, born 1942. He studied chemistry and medicine at the University of Pennsylvania (AB 1964, MD 1968). After internship at the UCSF he went to the National Institutes of Health for three years and returned as resident in neurology to UCSF. He became assistant professor in 1974, lecturer 1976, associate professor 1980 and full professor in 1984. In 1972 he encountered his first patient with CJD, which turned his interest to the spongiforme encephalopathies and their causative agents, then believed to be slow virus. Like T. ALPER before him he found that preparations of the causative agents appeared to consist only of protein, not of nucleic acids. This led to his publication of the prion hypothesis in 1982. He and his collaborators went a long way to characterise this protein. He received the NOBEL Prize for Physiology and Medicine in 1997.

RAMACHANDRAN, GOPALASAMUDRAM NARAYANA IYER Indian, 1922–2001. He studied physics at Madras University (Master 1942, DSc), then Cavendish Laboratory in Cambridge (PhD 1949). He invented an X-ray mirror and an X-ray microscope and made important contributions to the field of crystallography. After returning to Madras University in 1952 his interest shifted to macromolecular structure, discovering the collagen triple helix (1954 with G.

KARTHA) and inventing the RAMACHANDRAN map in 1962. He is considered one of the fathers of the field of molecular biophysics.

RÉAUMUR, RENÉ ANTOINE-FERCHAULT DE French, 1683–1757. He studied law, philosophy, and science, and moved to Paris in 1703, where he became member of the Academy of Science in 1708. One of the great universal scholars in the history of science, he worked on palaeontology, zoology, and chemical technology. He invented the alcohol/water-filled thermometer, a process for steel production, the incubator and recognised corals as animals rather than plants. In his work on mollusca he found that their shells are produced from a juice produced by the animal; he also reintroduced dying with purple, a technique lost in antiquity. He fed food bits packed in small metal cages to seagulls; when the birds regurgitated the cages, the food inside was partly digested. If the cages contained little sponges, a juice pressed from them could hydrolyse meat. Thus RÉAUMUR recognised digestion as a chemical process.

SANGER, FREDERICK British, 1918–2013. He received his BA 1939, PhD 1943, both in Cambridge. He introduced 2,4-dinitrofluorbenzene (SANGER's reagent) into protein chemistry in 1945 and was the first to determine the primary structure of a protein (insulin) after years of work. He received the Nobel Prize in Chemistry 1958. He later switched to nucleic acid sequencing, developing the chain termination method (1975) for which he received a second NOBEL Prize in 1980.

SCHWANN, THEODOR German, 1810–1882. He studied medicine in Berlin (MD 1834), where in 1836 he prepared the first enzyme from animal sources, pepsin, by adding mercury chloride to stomach juice. He became professor at the University of Louvain in 1838 and Liège in 1848. With M. J. SCHLEIDEN he formulated the cell theory, introduced the term "metabolism", and worked on muscle and nerve cell excitation (discovering the myelin sheath around nerve axons named after him) and fermentation. He is considered the founder of embryology.

SKOU, JENS CHRISTIAN Danish, born 1918. He studied medicine at the University of Copenhagen 1937–1944. He then started PhD work in Aarhus on anaesthetics in 1947 (finished in 1954) which led to the discovery of Na^+/K^+-ATPase. Skou stayed in Aarhus, became professor and chairman of the department of physiology in 1963. He received a Nobel Prize in Physiology and Medicine for his discovery of Na^+/K^+-ATPase in 1997, together with P. D. BOYER & J. E. WALKER.

SØRENSEN, SØREN PEDER LAURITZ Danish, 1868–1939. He received his PhD in chemistry in 1899 in Copenhagen. He became head of Carlsberg Laboratory, Copenhagen. He recognised the importance of hydrogen ion concentration for chemical and in particular enzymatic reactions and introduced the logarithmic *p*H-scale (in 1909).

SPALLANZANI, LAZZARO Italian, 1729–1799. He studied law and theology at the University of Bologna, but turned to science, becoming professor of logic, metaphysics, and Greek at the University of Reggio aged 25. In 1768 he proved that cells could not be spontaneously generated but came from air and could

be killed by boiling. However, because he used closed vessels to keep his preparations germ-free his results were not generally accepted until L. PASTEUR repeated the experiment in the open swan-neck bottle. He described the role of sperm and ovum in mammalian reproduction and performed the first artificial insemination (in a dog). He studied the regeneration of body parts in lizards and he proved that the active ingredient in stomach juice (pepsin) lost its activity upon storage. Investigating bat orientation he found that the animals could still fly with their eyesight blocked, but lost their bearing when their ears were covered. The reason of course eluded him and became clear only after the discovery of ultrasound.

STANLEY, WENDELL MEREDITH United States, 1904–1971. He received his BA at Earlham College 1926, MA 1927, and PhD in chemistry 1929, both at the University of Illinois. He did his postdoc in Berlin, and was assistant at the Rockefeller Institute 1931, professor of biochemistry at the University of California 1948. He worked on lepracidal compounds, sterols, tobacco mosaic, and influenza virus. He was awarded the Nobel Prize in Chemistry 1946 together with J. B. SUMNER & J. H. NORTHROP.

STOKES, GEORGE GABRIEL Irish, 1819–1903. He studied at the University of Cambridge where he also became professor of mathematics in 1849. He worked on pure maths and on mathematical and experimental physics, contributing greatly to hydrodynamics, optics and the theory of wave propagation.

SUMNER, JAMES BATCHELLER United States, 1887–1955. He lost his left arm in a hunting accident at age 17, and studied chemistry at Harvard, graduating in 1910. After brief spells at various positions he continued his studies in biochemistry at Harvard, obtaining a PhD in 1914. He became assistant professor at Cornell 1914, and full professor 1929. He started to work on the purification of jackbean urease in 1921, succeeding in 1926. Urease was the first ever enzyme to be purified enough to be crystallised, proving that proteins were defined compounds. He received the Nobel Prize in Chemistry 1946, together with J. H. NORTHROP & W. M. STANLEY.

SVEDBERG, THEODOR (THE) Swedish, 1884–1971. He studied at Uppsala University: BA 1905, MA 1907, PhD 1908. His thesis, "Studien zur Lehre von den kolloiden Lösungen" is still a classic in his field; he tested EINSTEIN's theories on BROWNIAN motion and thereby proved the existence of molecules and founded the field of molecular hydrodynamics. Svedberg became lecturer of physical chemistry in 1909, and professor in 1912. He constructed the first ultracentrifuge (1923) which for the first time allowed the determination of the molecular mass of proteins. Together with TISELIUS he invented free-flow electrophoresis of proteins. He worked on nuclear chemistry, radiation biology, photochemistry, and polymer chemistry. He was awarded the Nobel Prize in Chemistry in 1926 for his work on colloids.

TANAKA, KŌICHI Japanese, born 1959. Started studying electrical engineering at Tōhoku University in 1978 and joined the central research laboratory of Shimadzu Corporation in 1983. In 1984 he started working there on mass spectrometry, by chance inventing matrix-assisted laser desorbtion/ionisation

(MALDI, with glycerol as matrix), one of two methods used to ionise protein molecules for mass spectrometric analysis. For this important contribution to protein science he received the Nobel Prize in Chemistry for the year 2002, together with J. B. FENN and K. WÜTHRICH.

THEORELL, AXEL HUGO THEODOR Swedish, 1903–1982. He studied medicine at the Karolinska Institute from 1921, awarded Bachelor of Medicine 1924, associate assistant 1924, temporary associate professor 1928, and MD 1930 with a dissertation on blood plasma lipids. He moved to Uppsala to work with T. SVEDBERG in 1931 and became associate professor in medical and physiological chemistry at Uppsala University in 1932. From 1933–1935 he worked with O. WARBURG in Berlin-Dalem, discovering FMN. In 1936 he was appointed director of the biochemistry department of the newly founded NOBEL Medical Institute, which opened in 1937. Here he continued his work on the reaction mechanism of redox enzymes, developing new methods of rapid enzyme kinetics, for which he received the NOBEL Prize for Physiology and Medicine in 1955.

TRAUBE, MORITZ German, 1826–1894. A wine merchant and private scholar, he worked on biological oxydation and osmosis and invented an apparatus for measuring the osmotic pressure of solutions in 1867 (incorrectly known as PFEFFER's cell, PFEFFER built a similar device 10 years later). He was the first to recognise that enzymes were proteins.

TSWETT, MIKHAIL Russian botanist, 1872–1920. He invented chromatography to separate leaf pigments.

VIRCHOW, RUDOLF LUDWIG KARL German, 1821–1902. He received his MD 1843 at the Friedrich-Wilhelms-University in Berlin, was prosector at the Charité, founded the *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin* (now *Virchows Archiv*) in 1847. In 1848 he participated in the failed March Revolution and had to leave Berlin. He worked at the University of Würzburg (1849–1856); then he returned as professor of pathology to Berlin, where he remained for the rest of his life. He strongly influenced medicine (cancer originates from activation of dormant cells, meat inspection, wastewater treatment, systematic autopsy, medical care for the entire population), anthropology, and archæology, and virtually founded modern pathology (*Die Cellularpathologie: in ihrer Begründung auf physiologische und pathologische Gewebelehre*, 1858). From 1859 on he was a member of the city parliament in Berlin, from 1862 a member of the Prussian Parliament and from 1880 a member of the German *Reichstag* for the *Fortschrittspartei*, which he cofounded. When OTTO VON BISMARK (then Prussian prime minister) challenged him to a duel in 1865 after a stormy parliamentary session, he replied that this was not a civilised form of discussion (other sources claim that he suggested pork sausages as weapons). Amongst his more than 2000 publications only few were erroneous: he was an antievolutionist, doubted the bacterial cause of infectious disease, and thought that the skeleton found at *Neanderthal* belonged to a deformed modern human.

WALKER, SIR JOHN ERNEST British, born 1941. He received his BA in chemistry from St. Catherine's College, Oxford in 1964, and DPhil. 1969. He did his postdoctoral training in 1969–1974 in various positions, then joined the Medical Research Council's Laboratory of Molecular Biology. He started work on membrane proteins in 1978, in particular ATP-synthase and discovered the ATPase-motive named after him and crystallised the F-domain of ATP-synthase. He received the Nobel Prize in Chemistry 1997 together with P. D. BOYER & J. C. SKOU for his studies on the reaction mechanism of ATP-synthase.

WARBURG, OTTO HEINRICH German, 1883–1970. He studied chemistry in Berlin under E. FISCHER, obtaining a doctorate in 1906. He then went to Heidelberg and obtained an MD in 1911. During WWI he served in the Prussian horse guards, in 1918 he was appointed professor at the Kaiser-Wilhelm Institute of Biology in Berlin-Dahlem; in 1931 he became director of the Kaiser-Wilhelm Institute for Cell Physiology. He worked on polypeptides, biological oxydation, carbon fixation in plants, radiation biology, and tumor-metabolism. He discovered the roles of FMN and $NAD(P)^+$ (together with A. H. T. THEORELL, receiving the NOBEL Prize 1955), and of iron-containing enzymes. Although he held the rank of professor he was never a teacher and spent his time entirely on research. For his research on respiratory enzymes he was awarded the NOBEL Prize in Physiology and Medicine in 1931.

WILLEBRAND, ERIK ADOLF VON Finnish, 1870–1949. An internist, he published his study on "hereditary pseudohæmophilia" in a family of 66 members from the Åland-islands in 1926. Unlike hæmophilia (lack of clotting factor VIII) females too may show this disease. It took 60 years until the defective gene for what is now known as VON WILLEBRAND-factor could be identified.

WILLSTÄTTER, RICHARD German, 1872–1942. He studied science at the University of Munich and the Technical School in Nürnberg, receiving his PhD 1894 as a student of BAEYER. He was lecturer in 1896 and professor in 1902. He moved to ETH Zürich in 1905, Kaiser-Wilhelm Institute in Berlin 1912, and University of Munich 1916. In 1925 he resigned his post as professor in protest against growing anti-Semitism and fled to Switzerland in 1939. He was awarded the Nobel Prize in Chemistry in 1915 for his work on natural dyes, in particular chlorophyll and hæmoglobin, whose structure he solved. His enzymological studies laid the ground for chromatographic purification of proteins, even though he himself never believed enzymes to be proteins.

WÜTHRICH, KURT Swiss, born 1938. He studied science and sports 1957–1962 at the University of Berne, then from 1962–1964 at the University of Basel, where he obtained a PhD in chemistry. His project on the catalytic activity of copper compounds required the use of electron paramagnetic resonance spectroscopy (EPR). For postdoctoral studies he went to Berkeley (1965–1967), where he started using nuclear magnetic resonance (NMR). This work he continued at Bell Laboratories (1967–1969), where he was in charge of the first NMR-spectrometer operating with supercooled coils, working at 220 MHz. Here he started his NMR-work on proteins, initially with hæmoglobin. In 1969 he joined the ETH at Zürich where he remained for the rest of his career.

There he developed new methods of multidimensional NMR. This allowed the deconvolution of the signals of the many different atom-atom interactions in a protein. For this work he received the Nobel Prize in Chemistry in 2002, together with J. B. FENN and K. TANAKA.

ZELLWEGER, HANS ULRICH Swiss, 1909–1990. He studied medicine in Zurich, Hamburg, Berlin, and Rome, and received his MD 1934 in Zurich. He worked with ALBERT SCHWEITZER in Lambaréné (1937–1939), then with FANCONI in Zurich. In 1951 he was professor for pædiatrics at the American University in Beirut, and in 1959 the University of Iowa, where he described the disease named after him in 1964.

Appendix B: List of Symbols

c	Concentration (mol/L)
D	Diffusion coefficient (cm^2/s)
e	Elementary charge (1.6022×10^{-19} C/mol)
$\mathcal{E}$	Absorbance (pure number)
E	Energy (J)
	Potential difference (V)
$\mathfrak{E}$	Electrical field strength (V/m)
$\mathfrak{F}$	Force (N)
F	FARADAY-constant ($96\,484.61$ C/mol $= \mathrm{J\,V^{-1}\,mol^{-1}}$)
f	Friction coefficient (kg/s)
$\mathfrak{g}$	Gravitational acceleration ($9.8067\,\mathrm{m\,s^{-2}}$)
$\Delta G'^0$	GIBBS free energy (under standard biological conditions, J/mol)
h	PLANCK's quantum (6.6262×10^{-34} J/s)
k	BOLTZMANN constant (1.3807×10^{-23} J/K)
	Reaction velocity constant (unit depends on order of reaction)
k_{cat}	Turnover number (s^{-1})
K_a	Association constant (M^{-1})
K_d	Dissociation constant (M)
K_{m}	MICHAELIS-constant (M)
K_p	Partitioning coefficient (pure number)
l	Length (m)
m	Mass (kg)
M	Molecular mass (pure number, but Da is often used)

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

$\bar{m}$	Aggregation number of a detergent (number of molecules per micelle)
N_A	AVOGADRO's number ($6.022 \times 10^{23}\,\mathrm{mol}^{-1}$)
n	HILL-coefficient (pure number)
	number
pH	Hydrogen ion tension (pure number)
pI	Isoelectric point (pure number)
$p\mathrm{K_a}$	Strength of an acid (pure number)
Q	Electrical charge (C)
r	Radius (m)
R	Universal gas constant ($8.3143\,\mathrm{J\,mol^{-1}\,K^{-1}}$)
t	Time (s)
$t_{1/2}$	Half-life period (s)
T	Absolute (thermodynamic) temperature (K)
v	Reaction velocity (mol/s)
$V_{\max}$	Limiting reaction velocity (kat = mol/s)
V	Volume (L)
z	Number of elementary charges transferred in a reaction
θ	Molar fraction of enzyme with bound substrate (pure number)
λ	Wavelength (nm)
τ	Relaxation time (s)
ϕ	Dihedral angle around the N-C^{α}-bond (deg)
ψ	Dihedral angle around the C^{α}-C'-bond (deg)

Appendix C: Greek Alphabet

alpha	α	A
beta	β	B
gamma	γ	Γ
delta	δ	Δ
epsilon	ϵ, ε	E
zeta	ζ	Z
eta	η	H
theta	θ, ϑ	Θ
iota	ι	I
kappa	κ	K
lambda	λ	Λ
mu	μ	M
nu	ν	N
xi	ξ	Ξ
o	o	O
pi	π, ϖ	Π
rho	ρ, ϱ	R
sigma	σ, ς	Σ
tau	τ	T
upsilon	υ	Υ
phi	ϕ, φ	Φ
chi	χ	X
psi	ψ	Ψ
omega	ω	Ω

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

Appendix D: The Genetic Code

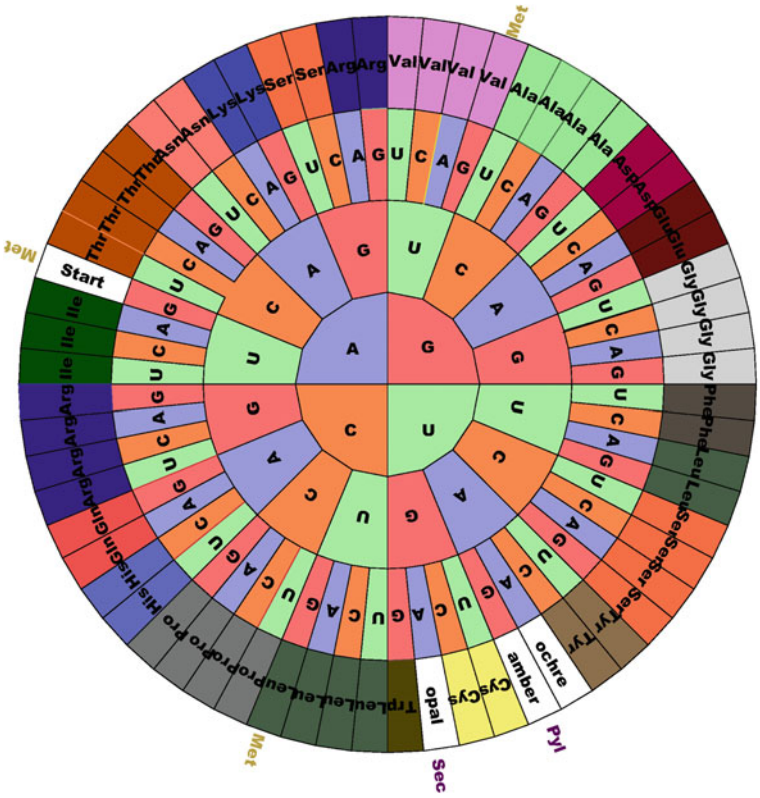


Fig. D.1 mRNA-codons and the corresponding amino acids. Alternative uses of codons are marked on the exterior. The colours used to symbolise different compounds are known as "shapely colour set", a quasi-standard in molecular modelling

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

Appendix E: Acronyms

AAA+	ATPases associated with diverse cellular activities, common motive in proteins that link ATP hydrolysis to some cellular function
ABC	ATP binding cassette, major family of primary active transporters
ACTH	Adrenocorticotropic hormone, (corticotropin) peptide stress hormone
ADP	Adenosine 5'-diphosphate
AE1	Anion exchanger 1, also called band-3 protein, equilibrates chloride and bicarbonate across the erythrocyte membrane
AGE	Advanced glucation end products, substances spontaneously formed from glucated proteins, may be involved in aging
AICD	APP intracellular domain, cytosolic fragment of APP produced by secretases
AIDS	Acquired immune deficiency syndrome, life-threatening sexually transmitted disease caused by HIV
ALS	Amyotrophic lateral sclerosis
AMP	Adenosine 5'-monophosphate
AMPK	AMP-activated protein kinase, switches body from fat synthesis and storage to fat use
AMP-PCP	Adenosine 5'[β, γ-methylene]triphosphate, nonhydrolysable ATP-derivative; the oxygen between the β- and γ-phosphate is replaced by an methylene (-CH_2-) group.
AMP-PNP	Adenosine 5'[β, γ-imino]triphosphate, nonhydrolysable ATP-derivative; the oxygen between the β- and γ-phosphate is replaced by an imino (-NH-) group
AP	Adaptor protein, occur in CCV
apoA-I	Apolipoprotein A-I
APP	β-amyloid precursor protein, involved in plaque formation in ALZHEIMER disease
Aqp	Aquaporin, water-conducting pore in the cell membrane

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

Arf ADP-ribosylating factor, small G-protein involved in COP-vesicle formation

ATP Adenosine 5'-triphosphate, energy carrier for metabolic reactions

ATPase Adenosine triphosphatase, enzymes that hydrolyse ATP to ADP and P_i

CBB-G250 Coomassie Brilliant blue G250, reagent that changes colour after binding to protein

CBB-R250 Coomassie Brilliant blue R250, dye used for staining proteins after electrophoresis

BiP Ig heavy chain binding protein; synonym for Grp78

BMI body mass index, weight over squared height (kg/m^2), clinical indicator of obesity

2,3-BPG 2,3-Bisphosphoglycerate, compound used to modulate the oxygen affinity of hæmoglobin

BSE Bovine spongiforme encephalopathy, prion-disease in cattle, also known as "mad cow disease"

Btu B twelve uptake, ABC-transporter required for the uptake of vitamin B_{12} in bacteria

C1Inh C1-inhibitor, inactivates complement. Lack of this protein causes hereditary angioneurotic œdema

C4BP C4-binding protein, protein which inactivates complement

CAM Cell adhesion molecule, proteins required for cell-cell interactions

cAMP Cyclic adenosine 5'-monophosphate, an important second messenger

CARD Caspase recruitment domain, N-terminal domain of NLR

CCV Clathrin coated vesicles, required for endocytosis, transcytosis and GOLGI-endosome transport

Cdc Cell division cycle, proteins involved in the regulation of mitosis

CDC Centers for Disease Control in Atlanta (Georgia, USA)

Cdk Cyclin-dependent kinase, has two isoforms (Cdk1 and Cdk2)

cDNA Complementary DNA, DNA produced from RNA by reverse transcriptase

CFTR Cystic fibrosis transconductance regulator, a chloride channel

CH Calponin homology, domain in some actin-binding proteins

CIITA Class-II transactivator, regulates MHC-gene expression in an interferon dependent manner. Lack of this protein causes bare lymphocyte syndrome, an inherited immunodeficiency

CJD CREUTZFELDT-JAKOB disease, a naturally occurring prion disease in humans

CK Creatine kinase catalyses the reaction phosphocreatine + ADP $\rightleftharpoons$ creatine + ATP. Phosphocreatine is a storage form of high-energy phosphate especially in muscle, but also in other cells.

CLIP Class II associated invariant chain peptide, stabilises MHC-II until antigenic peptide is bound

cMOAT Canalicular multiple organic anion transporter

CMP Cytosine monophosphate
COMT Catechol-O-methyltransferase, enzyme that inactivates catecholamines
COP β-coat protein, in vesicles for ER/GOLGI transport and for transport between GOLGI-stacks
COPD Chronic obstructive pulmonary disease, for example, chronic bronchitis or emphysema
CPAP Continuous positive airway pressure, method for mechanical respiration
CPEB Cytoplasmic polyadenylation element binding protein, protein required for long-term memory in the sea-snail *Aplysia*
CPK COREY, PAULING & KOLTUN, colour standard for molecular models named after authors
CSF Colony stimulating factors, cytokines that stimulate cells of the immune system
CTAB Cetyltrimethyl ammonium bromide, positively charged detergent which can solubilise membrane proteins without denaturation
CTP Cytidine triphosphate
DAG Diacylglycerol, second messenger
DAF Decay-accelerating factor, protein which inactivates complement. Inability to synthesise its glycolipid tail leads to paroxysmal nocturnal hæmoglobinuria
DAPI 4',6-Diamidino-2-phenylindole, substance that shows blue fluorescence (358/461 nm) after intercalating into DNA
DCCD Dicyclohexylcarbodiimide, used to activate carboxylic acid groups for coupling reactions
DEPC Diethylpyrocarbonate, inactivates enzymes by reacting with essential His-residues
DNA Deoxyribonucleic acid, carrier of genetic information
DnaJ ADP/ATP exchange stimulating factor of DnaK
DnaK Chaperone in prokaryotes and organelles derived from them
DNP 2,4-dinitrophenol, uncoupler
DTT Dithiotreitol, used in the lab as an antioxydant to protect SH-groups in proteins (CLELAND's reagent)
EC Enzyme Commission, body responsible for enzyme nomenclature within IUBMB
ECM Extracellular matrix, "glue" between cells
EDTA Ethylenediamine-N,N,N',N'-tetraacetic acid, reagent that forms strong complexes with some metal ions
EGF Epidermal growth factor
ELISA Enzyme-linked immunosorbent assay; very sensitive method to detect and quantify either antibodies or their antigen
Elk-1 Eph-like kinase 1
EM Electron microscopy

EMIT Enzyme multiplied immunoassay technique, fast homogeneous immunological assay for small molecules such as pharmaceuticals or drugs of abuse

Eph Ephrin receptor tyrosine kinase, receptor for ephrins, a group of small proteins required for morphogenesis of the nervous system

ER Endoplasmic reticulum, membrane system inside a eukaryotic cell

ERAD ER associated protein degradation, mechanism to destroy misfolded proteins in the ER

ERK Extracellular signal-regulated kinase, protein in the MAP-kinase pathway

EU European Union, economic and political association of (at the time of writing) 25 European countries

F_{ab} immunoglobulin molecule whose F_c-end has been removed enzymatically

FACS Fluorescence-activated cell sorter, instrument to count and isolate cells expressing particular antigens

F_c carboxy-terminal end of an immunoglobulin molecule, interacts with effector cells

FAD Flavine adenine dinucleotide, oxydised

$FADH_2$ flavine adenine dinucleotide, reduced

FFI Fatal familial insomnia, naturally occurring prion disease in humans

FITC Fluorescein isothiocyanate, amine-reactive fluorescent probe

FMN Flavine mononucleotide, prosthetic group of some flavoproteins

FMOC- 9-fluorenyl-methoxycarbonyl-, used as protection group during peptide synthesis

Fuc Fucose, a sugar used in glycoproteins and glycolipids

GABA γ-aminobutyric acid, a neurotransmitter

Gal Galactose

GALT Gut-associated lymphoid tissue

GalNAc N-acetylgalactosamine

GAPDH Glyceraldehyde-3-phosphate dehydrogenase, enzyme in glycolysis and gluconeogenesis

GDP Guanosine diphosphate

GFP Green fluorescent protein, favourite marker in molecular biology

Glc Glucose

GlcNAc N-acetyl-glucosamine

GluT Glucose transporter

GPI Glycosyl phosphatidylinositol, C-terminal anchor of type-6 membrane proteins

GRB2 Growth factor receptor-bound protein 2, protein in the MAP-kinase pathway

Grp78 78 kDa glucose-regulated protein, chaperone of the ER

GrpE ATP-hydrolysis stimulating factor of DnaK

GSH Glutathione, reduced

GSS	GERSTMANN-STRÄUSSLER-SCHEINKER-disease, a natural prion disease in humans
GSSG	Glutathione, oxydised
GTP	Guanosine triphosphate
GuHCl	Guanidinium hydrochloride, substance that destroys protein secondary, tertiary and quaternary structure
HbA	normal adult hæmoglobin (α_2/β_2)
HbA_{1c}	glycated HbA
HbC	Hæmoglobin C, Glu-6 → Lys
HbF	Fetal hæmoglobin, ($\alpha_2\gamma_2$)
HbH	Hæmoglobin H, tetrameres of β-subunits, formed in embryos with α^0-thalassæmia
HbS	Sickle cell hæmoglobin, Glu-6 → Val
HDL	High density lipoprotein, VLDL from which cholesterol has been removed
HIV	Human immunodeficiency virus, causative agent of Acquired immune deficiency syndrome (AIDS)
HLA	Haupt-Lymphocyten Antigen, (Ger.), synonym for MHC, today mainly used in tissue typing for transplantation
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A, intermediate of the mevalonate pathway
HMM	HENRI-MICHAELIS-MENTEN, biochemists who described the relationship between substrate concentration and enzyme reaction velocity
HPLC	High-performance liquid chromatography, chromatography on fine-grained, homogenous stationary
Hsc10	10 kDa heat shock cognate, GroEL
Hsc70	70 kDa heat shock cognate, constitutively expressed cytosolic isoform of Hsp70
Hsc60	60 kDa heat shock cognate, GroES
Hsp40	40 kDa heat shock protein, eukaryotic homologue of DnaJ
Hsp70	70 kDa heat shock protein, a molecular chaperone
Hsp90	90 kDa heat shock protein, a molecular chaperone
HSV	Herpes simplex virus
HTLV	Human T-lymphotropic virus, group of retrovirus causing leucæmia/lymphoma
IDP	Intrinsically disordered protein, a protein that has extended structure by itself, but refolds upon contact with a binding partner
IEF	Isoelectric focussing, electrophoretic method to separate proteins by their isoelectric point.
IF	Intermediate filament
IFAP	Intermediate filament associated proteins
IFN	Interferon, a family of cytokines
Ig	Immunoglobulin
IgA	Immunoglobulin A

IgD Immunoglobulin D
IgE Immunoglobulin E
IgG Immunoglobulin G
IgM Immunoglobulin M
Iκ B Inhibitor of NF-κ B
IKK Inhibitor of NF-κ B kinase
IL Interleukin, a family of cytokines
IMAC Immobilised ion affinity chromatography, method to separate proteins by the number of exposed His-groups
IPTG Isopropylthiogalactoside, non-metabolisable inducer of the bacterial *lac* operon
IUBMB International Union for Biochemistry and Molecular Biology, an international body which, amongst other things, is responsible for standardisation of the nomenclature of molecules (in part together with IUPAC)
IUPAC International Union for Pure and Applied Chemistry
JAK Janus kinase aka "just another kinase", protein activated by cytokine receptors
KDO Ketodeoxyoctulosonate, component in the inner core of lipopolysaccharide
Kir6. K^+-channel, inwardly rectifying, together with Sur part of the K_{ATP} channel that regulates insulin secretion.
LAMP Lysosome-associated membrane glycoproteins protects the lysosomal membrane from degradation by hydrolases
LC Liquid chromatography, separation method
LCFA Long chain fatty acid, 14–24 C-atoms
Lck Lymphocyte-specific kinase, tyrosine kinase of the immune system
LDL Low-density lipoprotein
LPS Lipopolysaccharide, component of the membrane of GRAM-negative bacteria
M6P Mannose-6-phosphate, residue required for protein sorting into the lysosome
mAb Monoclonal antibody, obtained from fusing antibody-producing B-cells with immortal hybridoma cells
MAC Membrane attack complex, complex of complement proteins that damages cells
Man Mannose
Mant-ATP Methylantroyl-ATP, ATP labelled with a fluorescent reporter group
MAO Monoamine oxydase
MAP Mitogen activated protein, proteins phosphorylated in response to mitogen stimulation of a cell. Stimulation occurs via a cascade of MAP-kinases (MAPK), MAP-kinase-kinases (MAPKK) and MAP-kinase-kinase-kinases (MAPKKK). The abbreviation is also used for **M**icrotubule **A**ssociated **P**rotein

MASP MBL associated serine protease, protein in the complement pathway
MBL Mannan binding lectin, protein of the complement pathway that recognises bacterial cell walls
MBP Microtubule binding protein
Mdr Multiple drug resistance transporter, group of ABC-type transmembrane ATPases
MEK Mitogen/extracellular signal-regulated kinase, protein in the MAP-kinase pathway
MFS Major facilitator superfamily, secondary active transporters
MHC Major histocompatibility complex, antigen-presenting protein on the cell surface
MMP Matrix specific metalloprotease, breaks integrin/ECM-interactions
MODY Maturity onset diabetes of the young, inherited form of diabetes mellitus
MΦ Macrophage, amœboid cells of the innate immune system
MPF Maturation promoting factor, Cdk1 associated with Cyclin B
Mpp Matrix-processing protease, clips of the signal sequence of nuclear encoded proteins after their import into the mitochondrium
mRNA Messenger ribonucleic acid, working copy of genetic information
Mrp Multidrug resistance related protein
mtDNA Mitochondrial DNA
MS Mass spectrometry, sensitive analytical technique that identifies molecules by their mass
MTOC Microtubule organising centre, central hub where the minus-ends of microtubules meet
MW Molecular mass
NAD^+ Nicotine adenine dinucleotide, oxydised, acceptor of activated hydrogen, usually in catabolic reactions
NADH + H^+ Nicotine adenine dinucleotide, reduced, carrier of activated hydrogen
$NADP^+$ Nicotine adenine dinucleotide phosphate, oxydised, acceptor of activated hydrogen
NADPH + H^+ Nicotine adenine dinucleotide phosphate, reduced, donor of activated hydrogen, usually in anabolic reactions
NAD(P)$^+$ Either NAD^+ or $NADP^+$
NAD(P)H either NADH + H^+ or NADPH + H^+
NANA N-acetylneuraminic acid, sialic acid
NBD Nucleotide binding domain
NMDA N-methyl-D-aspartate, agonist of inotropic glutamate receptors in the nervous system
NEM N-ethylmaleimide, reactive compound used to label SH-groups
neu Neuronal tumor proto-oncogene, synonym for human epidermal growth factor receptor 2 (HER2)
NF-κ B Nuclear factor κ B

NKCC $Na^+/K^+/2Cl$ cotransporter
NLR NOD-like receptor, cytoplasmic receptors for peptidoglycans and their breakdown products
NMR Nuclear magnetic resonance, a method to determine the distance of atoms in a molecule
NOD Nucleotide-binding oligomerisation domain, central domain in NLR
NSAID Nonsteroidal anti-inflammatory drug, pain-killers such as aspirin or paracetamol
NSF NEM-sensitive factor, protein required for vesicle fusion
OMIM Online Mendelian inheritance in man, curated database of human inherited diseases maintained at the NIH
OmpA Outer membrane protein A from *E. coli*
OPA *Ortho*-phtaldialdehyde, reagent that forms fluorescent compounds with amino acids
PAF Platelet activating factor
PAMP Pathogen-associated molecular pattern, molecules characteristic for certain groups of pathogens, that are recognised by our innate immune system (*e.g.*, lipoteichoic acid, lipopolysaccharide, mannose-ending glycoproteins)
PC Phosphatidyl choline
PCR Polymerase chain reaction, highly sensitive method for the detection and quantification of nucleic acids
PDI Protein disulphide isomerase, acts on disulphide bridges in proteins
PE Phosphatidyl ethanolamine
PEG Polyethylene glycol, compound used in protein purification and crystallisation
PEP Phospho*enol*pyruvate, super-energy-rich phosphoester formed during glycolysis
PET Positron emission tomography, medical imaging technique
Perk , regulator of the unfolded protein response
PFIC Progressive familial intrahepatic cholestasis, group of inherited diseases of the liver
Pgp Permeability glycoprotein, synonym for Mdr1
PHHI Persistent hyperinsulinæmic hypoglycæmia of infancy
P_i Inorganic phosphate
PIP Phosphatidylinositol 4-phosphate
PIP_2 Phosphatidylinositol 4,5-bisphosphate
PITC Phenyl isothiocyanate, reagent used in EDMAN-sequencing of proteins
PK Pyruvate kinase, enzyme that catalyses the reaction phospho*enol*pyruvate + ADP → pyruvate + ATP

PKA Protein kinase A, cAMP-dependent Ser/Thr kinase acts on RRxS/T motive to stimulate Glc use (gluconeogenesis, glycogenolysis and glycolysis)

PKC Protein kinase C, Ser/Thr kinase activated by DAG and Ca^{2+}

PPI Peptidyl prolyl *cis/trans*-isomerase, catalyses the conversion between the *cis*- and *trans*-conformation of Pro in protein chains

PP_i Pyrophosphate

prion Proteinaceous infectious agent

PrP Prion protein, comes in two variants, cellular (normal) PrP^C and scrapy (abnormal) PrP^{SC}

PS Phosphatidyl serine

Rab Ras-associated protein group of small G-proteins involved in vesicle fusion, members of the Ras family

Raf Receptor associated factor, protein in the tyrosine kinase receptor signaling cascade

Rag Recombination activating gene, encodes for proteins that are required for recombination of DNA

RAGE Receptor for AGE, activates the NF-κ B-pathway

Ras Rat sarcoma viral antigen, protein in the MAPK-pathway. Constitutively active versions of this protein are encoded by some tumor virus

RFLP restriction fragment length polymorphism, used in molecular diagnosis

RIG Retinoic acid inducible gene, intracellular receptors for viral RNA

RING Really interesting new gene, $-C-X_2-C-X_{9-39}-C-X_{1-3}-H-X_{2-3}-C-X_2-C-X_{4-48}-C-X_2-C-$ Zn-finger domain in some E3 ligases

RNA Ribonucleic acid, can carry genetic information, act as catalyst (ribozyme) or may be involved in regulation

RNase Ribonuclease, enzyme that digests RNA

ROS Reactive oxygen species, molecules resulting from incomplete reduction of oxygen, such as $O_2^{-\cdot}$, $HO_2^{\cdot}$, H_2O_2, and $HO^{\cdot}$

RPC Reversed phase chromatography, separation method for small organic molecules

rRNA Ribosomal RNA

SAH S-adenosyl homocysteine, demethylation product of SAM

SAM S-adenosyl methionine, carrier of C1-bodies in metabolism

Sar Secretion-Associated RAS Super Family 1

SDS Sodium dodecylsulphate, anionic detergent used to destroy the secondary structure of proteins

SEC Size exclusion chromatography, synonym for gel filtration

Sec61 61 kDa secretory protein, transports membrane proteins into the ER

SEM Scanning electron microscopy, gives high-resolution images of 3D-structures

SH2 Src-homology-2, domain in several regulatory proteins

SHR Steroid hormone receptor
Sic1 Substrate/subunit inhibitor of cyclin-dependent protein kinase 1
sHsp Small heat shock proteins
SIR-2 Silent information regulator 2, class III protein deacetylase
SIV Simian immunodeficiency virus, causative agent of an AIDS-like disease in monkeys
SLE Systemic *lupus erythematosus*, debilitating autoimmune disease caused by auto-antibodies against nuclear antigens
SNAP Soluble NSF-attachment protein
SNARE SNAP receptor
SOS Son of sevenless, protein in the MAPK-pathway.
Spgp Sister of Pgp, ABC-type exporter of bile salts
SR Sarcoplasmic reticulum, specialised ER in muscle cells
Src Sarcoma proto-oncogene
SRP Signal recognition particle, recognises membrane proteins during their synthesis and directs them to the ER
STAT Signal transducer and activator of transcription, substrate of JAK
Sur Sulphonyl urea receptor
Tap Transporter associated with antigen processing, ABC-type ATPase that transports antigenic peptides from the cytosol into the ER
TCR T-cell receptor, protein on the surface of T-lymphocytes which binds to the MHC/antigen-complex
TdT Deoxynucleotidyl-transferase
TEM Transmission electron microscope, gives high-resolution images of thin-sections or very small objects
TGN *trans*-GOLGI network, exit compartment of the GOLGI-complex
T_h helper T-cell
TIL-1 Toll-IL-1 receptor, cytosolic domain of TLRs and receptors for the interleukin-1 family of cytokines
TIM Transporter of the inner membrane, required for import of nuclear encoded proteins into mitochondria
TIMP Tissue inhibitor of metalloprotease, regulate the activity of MMPs
T_k killer T-cell
TLR Toll-like receptor, family of PAMP-receptors on immune cells
TMD Trans-membrane domain
TNF Tumour necrosis factor, a family of cytokines
TOM Transporter of the outer membrane, required for import of nuclear encoded proteins into mitochondria
TRAM Translocating chain associated membrane protein
TRAP-T Tripartite ATP-independent periplasmic transporters, secondary active transporters with extracellular binding proteins
TRiC **T**ailless complex polypeptide 1 **Ri**ng **C**omplex, eukaryotic homologue of GroEL/GroES
TRITC Tetramethylrhodamine isothiocyanate, amine-reactive fluorescent probe

γ TU-RC	γ-tubulin ring complex, structure in the MTOC
USA	United States of America
UbA	E1-ubiquitin ligase
UbC	E2-ubiquitin ligase
UbL	Ubiquitin-like modifiers, enzymes that transfer ubiquitin-like peptides to proteins, involved in many regulatory processes
UDP	Uridine 5'-diphosphate
UTP	Uridine 5'-triphosphate
US	United States (of America)
UV	Ultraviolet, light with wavelengths shorter than 400 nm
UVIS	Visible and ultraviolet
vCJD	New variant CREUTZFELDT-JAKOB disease, human prion disease caused by eating meat from animals suffering from Bovine spongiforme encephalopathy
VLCFA	Very long chain fatty acid, $>$ 24 C-atoms
VLDL	Very low-density lipoprotein, transport cholesterol in blood
VopS	Vibrio outer protein S, pathogenicity factor of *Vibrio parahæmolyticus*
VSV	Vesicular stomatitis virus, laboratory model for protein synthesis and -transport studies
WASP	Wiscott-Aldrich syndrome protein, involved in the regulation of actin filament branching
WHO	World Health Organization
XLSA/A	X-linked sideroblastic anaemia/cerebellar ataxia, disturbance of iron metabolism due to mutation in ABCB7

Index

E. Buxbaum, *Fundamentals of Protein Structure and Function*,
DOI 10.1007/978-3-319-19920-7

D

T