Russian Chemical Reviews **64**(12) 1067 – 1104 (1995)
©1995 Russian Academy of Sciences and Turpion Ltd

UDC 547.963; 548.737

Current advances in the X-ray crystallography of proteins

E M Popov

Contents

I.	Introduction	1067
II.	Membrane proteins	1068
III.	Proteins of the immune system	1071
IV.	Aspartate proteases	1075
V.	DNA-binding proteins	1084
VI.	Proteins of skeletal muscles	1087
VII.	Synchrotron radiation	1092
VIII.	Multiwave anomalous diffraction	1098
IX.	Conclusion	1100

Abstract. The present state and prospects for the development of the X-ray crystallography of globular proteins are analysed. The mechanisms of the functioning of membrane receptors, the immune proteins of T-lymphocytes, the histone octameric core of nucleosomes, DNA topoisomerase, the aspartate proteases of retroviruses, and the proteins of the actomyosin complex of skeletal muscles are examined on the basis of the three-dimensional structures of the protein molecules which have become known in recent years. It is shown that the results of the X-ray diffraction analysis of proteins has played a decisive role in the creation of an experimental basis and the development of the approaches necessary for the establishment of theoretical biology. A large proportion of the review is devoted to the advances in X-ray crystallography as regards the sources of radiation, the phase problem, and the experimental conditions. The use of synchrotron radiation and the abandonment of multiplet isomorphous substitution in the determination of phases not only significantly simplify the study of the molecular structures of proteins but permit an approach to securing information about the structures of protein intermediates which had previously been inaccessible. The bibliography includes 324 references.

I. Introduction

One of the most outstanding scientific investigations not only in biology but throughout natural sciences in the course of the passing century was completed at the end of the 1950s and the beginning of the 1960s. Perutz and Kendrew determined the three-dimensional structures of haemoglobin and myoglobin, having first solved the problem, posed more than a quarter of a century earlier by Bernal, of the determination of the three-dimensional structures of protein molecules with the aid of X-ray diffraction. This event implied the establishment of a new direction in the development of biology — the morphology of elementary biosystems, revealing previously invisible possibilities for discovering the structure of living matter. Without knowing the atomic-molecular structures of proteins, it is impossible to understand the nature and specific features of their interactions, and the mechanisms of the functioning of enzymes, the immune system, the genetic apparatus, the membrane proteins, and many others. The advances achieved in the first investigations were of universal importance, which was reflected in the state of virtually all aspects of the research into the organic world. After more than three decades have elapsed since the publications by Perutz and Kendrew, X-ray diffraction still remains the only direct method for the determination of the three-dimensional structures of protein molecules and their complexes at the atomic level. The data obtained with its aid still remain the irreplaceable experimental basis of any search for solutions to fundamental problems such as those of the structural and structural-functional organisation of protein molecules. Taking into account the active role of proteins in vital processes and their universal biochemical and biophysical properties, one may claim that the creation of theories of the structural and structural-functional organisation of proteins will entail the establishment of quantitative theoretical biology.

After the classical studies by Perutz and Kendrew and also Phillips, who established in 1965 the three-dimensional structure of the first enzyme (lysozyme), X-ray diffraction analysis began to develop rapidly in many scientific centres. After only two years, the structures of ribonuclease A,[1,2] α-chymotrypsin,[3] and carboxypeptidase A,[4] were elucidated. By 1975 some 70 proteins had had their crystal structure determined with a resolution of better than 3.5 Å (Blundell and Johnson[5]) and at the present time the number approaches 3000.[6] The rate of advance was determined mainly by the difficulties in obtaining single crystal specimens of sufficient size and with a high degree of order. Proteins do not form regular macroscopic structures in the cells and organisms and the growing of protein crystals under laboratory conditions requires much skill and patience. The difficulties in obtaining high quality specimens are mentioned in almost all investigations. The following is, for example, a typical remark made by investigators[7] who studied the three-dimensional structure of lipoxygenase-1, an enzyme made up of 839 aminoacid residues: 'the protein crystals were obtained after trying more than 1000 different crystallisation conditions'.

E M Popov Laboratory for Enzyme Chemistry,
Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry,
Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10,
117871 Moscow, Russian Federation. Fax (7-095) 310 70 07

Received 23 June 1995
Uspekhi Khimii **64**(12) 1143 – 1182 (1995); translated by A K Grzybowski

This review is devoted to the analysis of the current state of the X-ray crystallography of proteins and the prospects for the development of the method in the immediate future. For this purpose, the studies which in some way constitute the most major advances at the present time were selected from the multiplicity of publications.

Sections II – VI present the results of X-ray diffraction studies on proteins, the scientific importance of which is determined by the importance of the objects analysed themselves. The three-dimensional structures of the molecules of membrane receptors, the immune proteins of T-lymphocytes, the histone octameric core of nucleosomes, DNA topoisomerase, the aspartate proteases of retroviruses, and the proteins of the actomyosin complex of skeletal muscles, all of which have become known only in recent years, are discussed in these sections. Studies of this kind have given rise to an extensive development of the study of the morphology of biosystems at the atomic-molecular level and will continue with increasing intensity.

In Sections VII and VIII, the use of the first methodological improvements in the X-ray diffraction analysis of proteins, concerning the sources of radiation, the exposure time, the phase problem, and the experimental conditions, is emphasised. Novelties of this kind are of special interest because they yield previously inaccessible information and hence lead to a vigorous development of the study of the morphology of proteins.

II. Membrane proteins

One of the greatest advances in the X-ray diffraction analysis of proteins in recent years is the start made on the elucidation of the three-dimensional structures of the first membrane proteins. It is useful to report briefly what had been known about these proteins before their investigation by X-ray diffraction. In the usual plasma membrane, the protein content is about 50% of its mass. However, in certain membranes, for example in the internal membranes of mitochrondria and chloroplasts, the protein content increases to 75%, whilst in others, for example in the myelin membrane, it falls to 25%. The lipid bilayer is threaded by many membrane proteins right through and the proteins are in contact with the aqueous medium on either side of the membrane. The molecules of these proteins, referred to as the transmembrane proteins, as well as the lipid molecules surrounding them possess amphipathic properties because they contain hydrophobic sections, interacting within the bilayer with the hydrophobic tails of lipids, and hydrophilic sections turned towards water on either side of the membrane. Another group of membrane proteins is in contact with water only on one side of the bilayer. Some of them are immersed in the outer or inner layer of the membrane, others are associated via nonvalence interactions with transmembrane proteins, whereas a third set of proteins of this kind are attached to the membrane via the chains of aliphatic acids bound covalently to them and inserted in the lipid layer.

By virtue of the dynamic properties of the membranes, the proteins associated with them can be separated from the lipids. Sometimes this operation is fairly simple and can be performed under mild conditions, for example by extraction with a salt solution. The liberated membrane proteins are referred to as peripheral. In other cases, usually involving transmembrane proteins, the isolation is more complex and is not always successful. Many proteins, called 'integral' on the basis of the method for their purification, can be isolated only after the complete disruption of the bilayer with the aid of detergents and organic solvents. On mixing with the membrane, the nonpolar, hydrophobic ends of the detergent molecules interact with the lipids and, by disrupting their association with the proteins, bind to the hydrophobic sections of the latter. As a result of this, the polar, hydrophilic ends of the proteins become oriented in the direction of the aqueous medium. The solubilised membrane proteins, which acquire the ability to dissolve in water, pass into solution in the form of complexes with the detergent. After the removal of the latter, the proteins lose their solubility and precipitate, forming a mixture of different proteins. As a rule, the membranes are heated to 100 °C in a 1% solution of a strong detergent as a preliminary step, which leads to the disruption of all the noncovalent protein–protein and protein–lipid interactions. Helenius and Simons[8] developed a method for the separation of proteins by electrophoresis in a polyacrylamide gel in the presence of a detergent (sodium dodecyl sulfate).

Further studies on membrane proteins — the investigation of their functions, the mechanisms of their action, and their three-dimensional structures — encountered major difficulties. It was found that the liberation of the protein molecules from the membrane is frequently accompanied by the irreversible denaturation of their native structures and hence by the loss of activity. In the rare favourable cases when the proteins are solubilised under mild conditions and when their concentrations are low, their functional properties are not lost, which indicates at least partial retention of the native conformations. The most detailed information about the proteins in the lipid bilayer has been obtained in the study of the plasma membrane of human erythrocytes, which contain haemoglobin — a nonmembrane protein which is the most abundant constituent of these cells. The convenience of this object is due to the fact that the erythrocyte membranes, called 'shadows', are the only ones in the cell and they can be easily isolated in a pure form with and without rupture and can even be obtained turned inside out. In the study of erythrocyte shadows, it was possible to establish for the first time that the membrane bilayer is threaded right through by certain membrane proteins and that the outer and inner sides of the membrane are asymmetric. Studies on the proteins of the human erythrocyte plasma membrane by one-dimensional electrophoresis in a polyacrylamide gel in the presence of sodium dodecyl sulfate led to the identification of about 15 principal proteins with molecular masses ranging from 15 to 250 kDa. Three of them — spectrin, glycophorin, and the so called band III protein — constitute altogether 60% of the mass of all the membrane proteins in the cell. They are bound in different ways to the membrane and can serve as examples of three typical ways in which proteins are associated with membranes.

Spectrin (240 kDa) can be equally regarded as a membrane protein and as a component of the cytoskeleton of erythrocytes. Two of its polypeptide chains, which are twisted relative to each other, form mobile dimers, which are combined into tetramers, as a result of interactions with the short actin filaments and other proteins, covering the cytoplasmic surface of the membrane in a dense network. The structural organisation enables spectrin to regulate the mobility of proteins and their distribution in the plasma membrane.

Glycophorin is the first membrane protein for which the aminoacid sequence has been determined. This transmembrane glycoprotein contains 131 aminoacid residues and about 100 sugar residues; its molecular mass is 30 kDa. Fig. 1 shows that a large proportion of the polypeptide chain of glycophorin is on the outer surface of the membrane, where the hydrophilic N-terminal fragment is located. The C-terminal section of the chain, also made up predominantly of hydrophilic residues, is immersed in cytoplasm, while the hydrophobic fragment, in the form of a single α-helix comprising 20 aminoacid residues, passes through the nonpolar lipid bilayer.

The band III protein from the human erythrocyte membrane is a transmembrane protein with a molecular mass of about 100 kDa (approximately 800 aminoacid residues). This is a transport protein, two molecules of which form an anionic channel for Cl^- and HCO_3^- ions moving passively through the membrane in accordance with their concentration gradients. The bilayer is threaded several times by the polypeptide chain of the protein in the α-helical conformation; approximately one third of the chain from the N-terminus is immersed in cytoplasm, whilst the short C-terminal section is located in the extracellular space (Fig. 1).

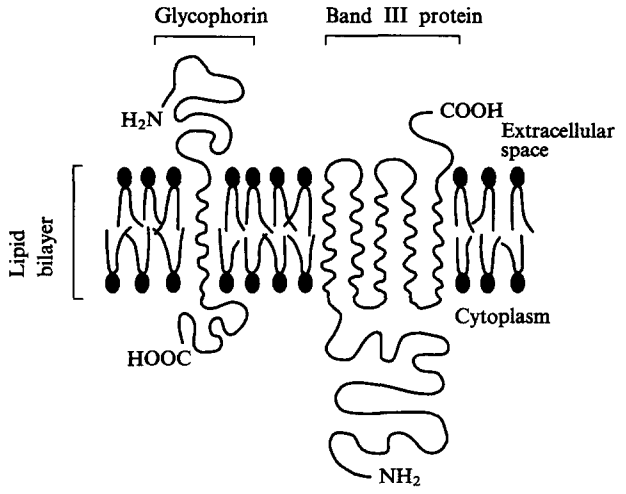


Figure 1. Schematic illustration of the disposition of glycophorin and the band III protein in the human erythrocyte membrane.

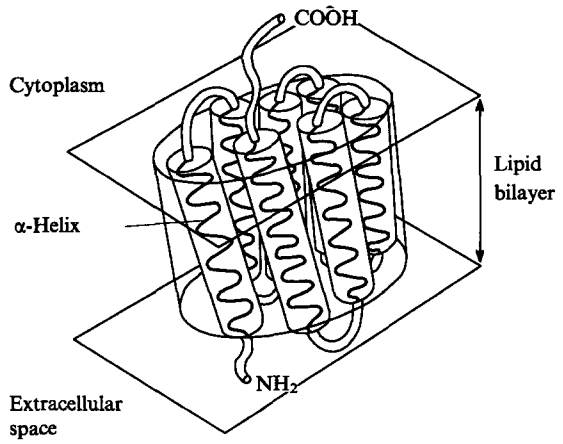


Figure 2. The structure of the bacteriorhodopsin molecule in the lipid bilayer.

In order to understand the mechanism of the functioning of the band III transport protein as well as the mechanisms of the action of other membrane proteins, it is necessary to know at least the three-dimensional structure of the molecule under the conditions of the lipid bilayer. Here it is necessary, first, to separate the transmembrane protein from the lipids and other membrane proteins without simultaneously damaging its three-dimensional molecular structure, which is very difficult; second, a highly ordered single crystal of the required size must be obtained from the isolated protein molecules. An alternative approach to the preparation of protein objects for X-ray diffraction analysis, which also concerned crystallographic studies on membrane proteins, arose in the 1980s. The approach involves genetic engineering based on the site-specific recombination of DNA. As a result, it became possible to devise specific, including artificial, genetic programs and to obtain the required proteins and their fragments in amounts sufficient not only for theoretical but also for applied purposes. However, genetic engineering methods have not radically facilitated the preparation of crystalline specimens made up of the native conformations of protein molecules and giving rise to distinct diffraction patterns. Nevertheless, the prospect of obtaining chemically pure membrane proteins, their mutants, and components of any length and with any variation of the aminoacid sequences greatly expanded the field where solutions to this problem may be sought.

Until recently, the only membrane transport protein for which certain details of its three-dimensional structure, obtained with a resolution of ~7.0 Å, were known was baceteriorhodopsin discovered in the plasma membrane of the purple bacterium *Halobacterium halobium*. Its polypeptide chain consists of 248 residues. Bacteriorhodopsin functions as a photoactive proton pump. After activation by a single light quantum, the excited chromophore induces conformational changes in the protein, as a result of which two protons are transferred from the inner surface of the cell to the outer surface. A proton concentration gradient and an electric potential gradient are created in the membrane and give rise to the synthesis of ATP. The bacteriorhodopsin molecules in the membrane form a planar crystal lattice resembling a two-dimensional crystal. Henderson and Unwin[9] used a combination of low intensity electron microscopy and small angle X-ray scattering to determine the three-dimensional structure of the protein and its orientation in the membrane. As a result of attaining record accuracy for both methods in the above investigation, it was shown that the polypeptide chain of bacteriorhodopsin forms seven α-helices in the membrane, each of which contains about 30 aminoacid residues (Fig. 2). All the helices intersect the lipid bilayer approximately at right angles to the plane of the membrane. They are close-packed and do not form a water-filled channel. The protons therefore apparently pass through the membrane via a series of protein chains of the α-helical residues, interacting with one another, the functioning of which is initiated by the excited chromophore.

A considerable proportion of the membrane proteins are receptors, i.e. are allosteric proteins capable of rearranging in a definite manner the three-dimensional structure of the molecule on interaction with hormones — specific signal mediators. The allosteric change in conformation is transmitted via the transmembrane chain of coupled proteins or domains of a single large protein within the cell and stimulates the corresponding physiological process in the latter. The available information about the three-dimensional structures of receptors and other components of transmembrane proteins is extremely scanty. Until recently, it has been possible to obtain the crystals and to determine the structures of only two specific membrane receptors — the photoreceptor centre[10,11] and porin.[12] In the 1990s, the studies on receptor proteins began to acquire a more systematic character. The information obtained hitherto about their three-dimensional structure refers, as will be shown below, not to the structures of whole membrane receptors but so far only to the outer portions of the molecules — extracellular domains.

1. The structure of the complex between the human growth hormone (hGH) and the outer part of the receptor (hGHbp)

De Vas et al.[13] established by X-ray diffraction with a resolution of 2.8 Å the three-dimensional structure of the protein complex between the hGH ligand and the extracellular domain of the hGHbp receptor, the interaction of which activates the growth and metabolism of the cells in the muscle, cartillage, and bone tissues. The molecule of the hGHbp membrane receptor consists of components separated in space: the extracellular fragment binding the hormone and located on the outer side of the membrane, the transmembrane fragment threaded through the lipid bilayer, and the intracellular component the function of which is as yet not entirely clear. There is also no clear idea about the mechanism of the interaction of all the components with one another and about the transfer into the interior of the cell of the signal which follows the association of the hormone with the outer part of the receptor.

The X-ray diffraction analysis of the hormone-receptor complex showed that the extracellular protein chain of the hGHbp molecule forms two easily distinguishable domains consisting of fragments incorporating the 1-123 and 128-238 aminoacid residues and linked by the 124-127 labile section. Each domain contains seven tiers of β-sheets, which give rise to a sandwich structure. De Vas et al.,[13] attribute a major importance

to the C-terminal octapeptide 239–246 section in the performance of the hormonal function. On sorption of the substrate, it ensures the conformational mobility of two domains of the receptor, which is necessary for the achievement of effective nonvalence contacts. In addition, it assists in the approach of the extracellular component of one receptor to the corresponding component of another and hence involves the two analogous outer domains of the latter in the binding of the hormone. This results in the formation of the complex hGH(hGHbp)$_2$, in which both receptors interact with the hormone virtually to the same extent despite the fact that there is no full structural similarity between the two centres for the binding of the hGH molecule to the pairs of domains of the two receptors. The three-dimensional structure of the complex hGH(hGHbp)$_2$ outwardly resembles the complex of an antigen with two pairs of domains (Fab) of immunoglobulin.

2. The structure of the extracellular domain of the aspartate receptor of the bacterium *Salmonella typhimurium* and its complex with the ligand

The three-dimensional structure of the domain in the free state of the receptor and in the receptor associated with aspartate has been determined by Milburn et al.[14] with resolutions of 2.4 and 2.0 Å. Being a typical transmembrane protein, the whole molecule of the aspartate receptor consists of three components: the outer ligand-binding domain, the cytoplasmic signal domain, and an intermediate domain located in the lipid bilayer. The periplasmic section of the 31–188 protein sequence, containing the Asp binding centre, is linked by two peptide fragments (7–30 and 189–212), threaded through the membrane, to the inner, largest domain, which is made up of the 1–6 and 213–552 residues.[15,16]

The aminoacid sequence of the extracellular 31–188 domain has been synthesised by a genetic engineering method.[17] The purified polypeptide crystallised exclusively in the form of a dimer having the same affinity for the aspartate ligand as the entire protein of the natural type. The diffraction pattern, which was initially diffuse, acquired a sharper outline on cross-linking of the two chains of the dimer by the disulfide bond formed by the Cys residues in positions 36. The preparation thus obtained did not differ from the domain of the native receptor as regards its functional properties.

The three-dimensional structures of the subunits of the dimer not associated with the ligand are identical and have the form illustrated schematically in Fig. 3. Each of them consists of four α-helices forming a cylinder ~20 Å in diameter and > 70 Å long.

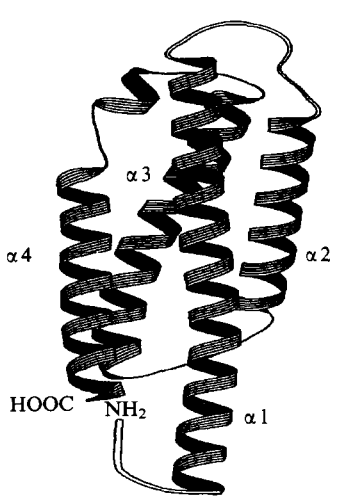


Figure 3. Schematic ribbon representation of the three-dimensional structure of the extracellular domain of the aspartate-binding receptor.[14]

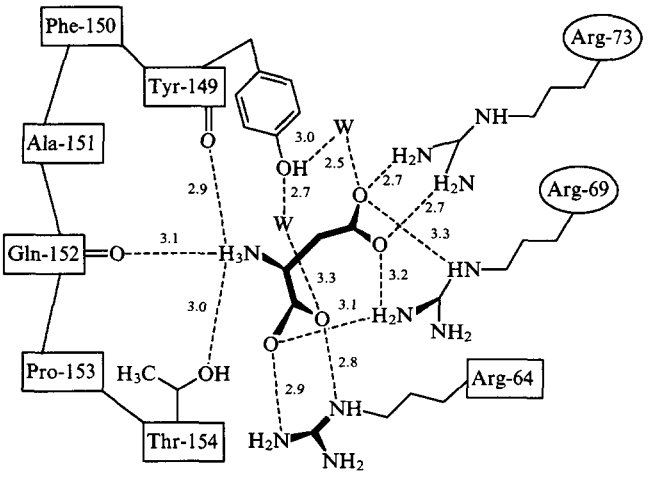


Figure 4. Schematic illustration of the hydrogen bonds between aspartate and the ligand-binding centre formed by the extracellular domains of two aspartate receptors. The aminoacid residues of one of the receptors are indicated in rectangles and those of the other in ovals; [14] W is a molecule of the water of crystallisation.

The dimer cross-linked by the disulfide bridge exhibits C_{2v} symmetry. On interaction with the ligand and formation of a complex, the symmetry breaks down. The centre for the binding of aspartate to the receptor is located near the apex of the dimer between the contact surfaces of the subunits at a height exceeding 60 Å above the postulated surface of the outer phospholipid layer of the membrane. In conformity with its symmetry, the dimer made up of the extracellular domains of two receptors has two hydrophilic ligand-binding pockets. One of them, most clearly manifested on electron density charts, is illustrated in Fig. 4. It incorporates the Arg-64 residue and the 149–154 fragment of one subunit and the Arg-69 and Arg-73 residues of another subunit, together with molecules of water of crystallisation, which form with aspartate a complex network of hydrogen bonds in the ligand–receptor complex. The sorption of two aspartate molecules induces the mutual approach of the residues forming the hydrophilic pockets and an overall compaction of the dimeric complex. The mean square deviation of the C$^\alpha$ atoms of one subunit (without taking into account the 77–85 fragment) is 0.47 Å and that of the C$^\alpha$ atoms of the other subunit (without taking into account the 25–36 fragment) is 0.34 Å. The deviations of the C$^\alpha$ atoms in the 77–85 and 25–36 sections exceed appreciably the above values. The conformational changes in the dimer on binding of aspartate lead to the rotation of the subunit by 4° relative to the short axis perpendicular to the twofold axis and the contact surfaces of the ligand-binding domains. The extent to which the above conformational deviations reflect the specific allosteric effect of the receptor still remains unelucidated.

3. The structure of prostaglandin H_2 synthase (PGHS)

Prostaglandins — aliphatic acid derivatives with 20 carbon atoms in the chain — constitute an important family of local chemical mediators, the source of which is cells of many types. They bind to a large number of receptors on the cell surface and induce the contraction of smooth muscle, diminish the capacity of thrombocytes for aggregation, activate the central nervous system, act on inflammatory processes, and liberate renin.[18] It has been established recently that prostaglandins also play an important role in the pathophysiology of cancer and of cardiovascular diseases.[19–21]

The crystal structure of the first membrane enzyme, prostaglandin H_2 synthase (PGHS), was first established in 1994.[22] The protein catalyses two processes in the first stage of the

biosynthesis of prostaglandins. It initiates and activates, firstly, the cyclooxidase reaction, during which arachidonic acid is converted into prostaglandin PGG_2 and, secondly, the subsequent peroxidase reaction, in which PGG_2 is converted into PGH_2, the latter being rapidly transformed into one of the prostanoids. The integral membrane protein PGHS is located predominantly in the endoplasmic reticulum. Its liberation from the membrane is achieved by extraction with a salt solution and by solubilisation with a nonionic detergent, in the presence of which crystallisation is also carried out. In the resulting specimen, the protein molecules are associated with flurbiprofen, a nonsteroid antiinflammatory preparation, and form dimeric complexes with a twofold symmetry axis. Two isomers of prostaglandin H_2 synthase have been detected in the cells: PGHS-1 and PGHS-2. These are distinguished by their activity and their distribution in the tissues. The three-dimensional structure of the sheep PGHS-1 has been determined[22] with a resolution of 3.5 Å. It is a glycosylated protein with a molecular mass of 70 kDa and 587 aminoacid residues, which participates in the maintenance of tissue homeostasis. As regards the function of PGHS-2, it has been suggested that the enzyme is responsible for the biosynthesis of prostaglandins during cell division and in inflammatory processes.

The three-dimensional structure of the PGHS-1 molecule consists of three largely autonomic and apparently independently coiling subunits: a small N-terminal domain, a membrane-bound section of the chain, and an enzyme domain. The first N-terminal domain comprises the 33–72 section of the polypeptide chain (the 1–24 residues constitute a signal sequence, which can be split off, and do not enter into the mature enzyme; the 25–32 residues could not be interpreted on the basis of the electron density maps obtained). Its conformation, stabilised by three intradomain disulfide bonds, is very similar to the structure of the epidermal growth factor (EGF). Polypeptide modules, similar to EGF, have been found in many extracellular proteins — factors stimulating the growth of neurons and the clotting of blood, and also in the proteins of surface cells necessary for their survival and proliferation. The function of the EGF-like domain of the PGHS-1 molecule is unknown. The second domain, incorporating the 73–116 residues, consists of four α-helices, oriented along one side of the monomer. Its conformation, highly amphipathic properties, and the hydrophobic surface profile quite definitely indicate the possible integration of the enzyme in only one lipid layer, so that the given membrane protein is monotopic according to Blobel's classification.[23] The third, largest catalytic domain of the PGHS-1 molecule (the 117–587 residues) has a globular structure consisting of two lobes twisted relative to one another but nevertheless fairly distinguishable. One of them, the larger, has a V-shape made up of two α-helices held together by an intermediate third helix acting in a manner resembling a pair of forceps. This construction serves as a framework around which a further four α-helices and the loops connecting them are packed. The small 'lobe' of the catalytic domain consists of a packet of six parallel helices.

As a bifunctional enzyme, the PGHS-1 protein has two active centres: the peroxidase and cyclooxidase centres. The former is on the surface of a narrow groove between the large and small 'lobes' of the catalytic domain and contains the prosthetic group — the high-spin Fe(III)-protoporphyrin IX — the proximal surface of which is turned towards the small 'lobe'. The peroxidase activity involves only the reduction of the PGG_2 synthesised in the cyclooxidase centre and the activation of the cyclooxidase reaction. This mechanism is typical for the haem-dependent peroxidases, which reduce hydroperoxidase substrates via a two-electron transfer reaction, leading to the formation of the oxoferrylporphin radical-cation. The oxidised haem returns to the initial stage after two successive one-electron transfers to the substrates being reduced or as a result of the transfer of oxygen to aryl sulfide compounds. The PGHS-1 molecules reduce preferentially PGG_2 and other long-chain aliphatic acids as well as aliphatic hydroperoxidases. The cyclooxidase active centre consists of a long narrow channel (8 × 25 Å) extending from the outer surface of the domain, linked to the membrane via its helix, to the centre of the monomeric PGHS-1 molecule. The diffuse diffraction pattern arising from the crystals, which have so far been grown for onlyone enzyme–inhibitor complex, did not allow the authors[22] to put forward on the basis of the data obtained a mechanism of the catalytic reaction occurring in the cyclooxidase centre.

In conclusion of the brief consideration of the results of the X-ray diffraction analysis of prostaglandin H_2 synthase, one should mention that the importance of this study, as the first investigation of the three-dimensional structure of a membrane enzyme, does not lie solely in the experimental facts themselves, the value of which is quite evident. A no less important feature of the study is associated with the influence of the data obtained on the very formulation of the relevant questions on a qualitatively different, higher level. The search for solutions to important problems such as the incorporation of enzymes in membranes, the function of the EGF-like module, the conformational and chemical aspects of the catalytic steps performed in the cyclooxidase and peroxidase active centres, and a number of others, which are of general interest, can now be based on concrete structural data.

III. Proteins of the immune system

Although the proteins of the immune system are membrane proteins, nevertheless, bearing in mind their functional characteristics, it is useful to consider the latest advances in X-ray diffraction analysis in the study of their three-dimensional structure in a separate section.

It was established in the 1960s that the two main types of immune responses are affected by different types of lymphocytes — T and B. The T-lymphocytes, which are responsible for cellular immunity, multiply in the thymus — the central lymphoid tissue. The antibodies produced and the B-lymphocytes responsible for the humoral response are formed from the stem cells of the haemopoietic tissue in mammals and in the Fabricius sac in birds. They subsequently migrate with the bloodstream to peripheral lymphoid tissues. The immune response of the cellular type implies the proliferation of specialised cells responding to foreign antigens, which are disrupted either directly on the surfaces of the cells produced or with the aid of other cells, such as macrophages. The humoral response consists in the activation of the synthesis of antibodies, circulating in the liquid media of the organism and interacting specifically with the foreign molecules. The binding of the antibody to the antigen facilitates the absorption of the latter by phagocytes and frequently activates the special system of the blood proteins — the complement system, which assists in the destruction of the antigen.[24,25] One of the functional problems in immunology is associated with the study at the atomic-molecular level of the mechanisms of the recognition and of the response reactions of the immune system to the appearance in the organism of infective antigens — foreign proteins, viruses, bacteria, and pathogenic substances.

According to Burnet's theory of clonal selection,[26,27] the specificity of the antigen response of lymphocytes is due to the presence on the surfaces of the T- and B-cells of receptor proteins, which interact selectively with particular antigens.

It is known that, on stimulation by the antigen, the T-lymphocytes begin to divide vigorously and are converted into active effector cells. Depending on the antibodies produced by them on the cell surface, they can be divided into at least three groups.[28] The first group comprises cytotoxic cells, which, by reacting with the foreign viral antigens, kill the infected cells before the onset of the replication of the virus. The molecular mechanism responsible for the destruction of the target cell has

so far not been elucidated. Nevertheless, it is known that this requires direct contact between the cytotoxic T-lymphocyte (the killer cell) and the infected cell (the target cell).[29] The second group of the activated T-lymphocytes includes the so called T-helpers, the helper cells required both by B- and T-cells to generate the humoral and cellular immune responses and also for the activation of macrophages. Finally, the third group is made up of T-suppressors, which suppress the response of the B-cells or other T-cells to antigens. It has been noted that, like the majority of T- and B-lymphocytes, the T-suppressors function only when they are continuously stimulated in this respect by the T-helpers, which, by activating the T-tissues of the suppressor, are themselves inhibited. Such feedback in the complex network of the interactions of lymphocytes ensures the self-regulation of the activities of cells of both types.

A large proportion of the T-lymphocytes recognise foreign agents, provided that they are not associated on the cell surfaces with membrane glycoproteins, which are coded by the genes of the major histocompatibility complex (MHC).[30] There are two classes of MHC molecules, I and II, each of which consists of a set of glycoproteins of the cell surface.[31-33]

The recognition by the T-lymphocytes of an antigen associated with the MHC glycoproteins on the surface of the target cell is effected by the T-cell antigen receptor (TCR) and one of the transmembrane proteins, CD4 or CD8, which are TCR coreceptors.[34-36] It has been suggested that a nonvalence complex is formed as a result of the stabilising contact between variable sections of the TCR molecule, the antigen, and the polymorphic sections of the MHC and also between the CD4 (or CD8) membrane proteins and the nonpolymorphic sections of the class II (or class I) MHC respectively. The coordinated interactions of TCR and CD4 (or CD8) with the MHC proteins of the target cells require a definite set of properties of the T-cells, which appear during the development of lymphocytes in the thymus, their differentiation, and maturation.[37-43]

The interaction of CD8 with MHC-I molecules triggers the intracellular signal system via the $p56^{lck}$ protein, an src-bound tyrosine kinase, associated directly in the lipoid cells with both CD4 and CD8.[44] In this case, the CD8–$p56^{lck}$ complex can function virtually like growth factor receptors, in which the linkage with the intracellular domain activates the intracellular tyrosine kinase.

The X-ray diffraction data on the three-dimensional structures of membrane proteins and their individual domains involved in the functioning of the T-cell immune system began to appear only in the 1990s. Since this concerns mainly the structures of the membrane proteins, the use of the X-ray diffraction method in cellular immunology has encountered major preparative difficulties which have not yet been fully overcome. Incidentally, the situation is also similar in all other aspects of molecular biology concerned with the study of phenomena occurring on the surfaces of cells. Therefore, in assessing the results of the X-ray diffraction analysis discussed below, one must bear in mind that they represent only the beginning of an extensive penetration of the method into the problems of T-cell immunology.

1. The structure of the extracellular fragment of the CD8 coreceptor of T-lymphocytes

The CD8 protein has been detected on the surfaces of T-cells in two forms: the disulfide-linked homodimer α_2 and the heterodimer $\alpha\beta$.[27] Both forms have N-terminal extracellular domains, the aminoacid sequences and structural features of which are similar to those of the variable domains of the immunoglobulin Ig-V.[45,46] The CD8 homodimer α_2 is illustrated schematically in Fig. 5. The extracellular region of CD8 includes 162 residues, 114 of which constitute the N-terminal immunoglobulin domain linked to the transmembrane section by a chain of 48 residues. The intracellular region consists of a highly polar aminoacid sequence interacting directly with tyrosine kinase $p56^{lck}$.

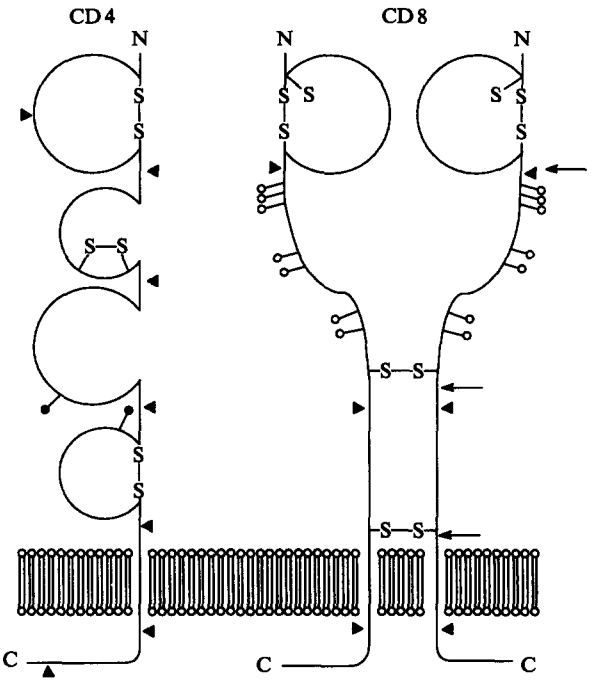


Figure 5. Schematic illustration of the CD4 receptor and the homodimer of the CD8 receptor. Each large circle corresponds to an immunoglobulin domain; the open circles denote the potential *O*-glycosylation sites, whilst the filled circles denote the corresponding *N*-glycosylation sites; the arrows indicate the ends of expression, whilst the dark triangles represent the limits of the intron sequences.[47]

The three-dimensional structure of a large part of the extracellular polypeptide chain of the human CD8 coreceptor, the length of which corresponds to 146 residues (CD8-146), has been established[47] with a resolution to within 2.6 Å. The crystals employed for the determination of the structure had the necessary size and gave rise to a satisfactory diffraction pattern only after the treatment of the protein preparation with neuraminidase, *O*-glycosidase, and V8-protease. In the CD8-146 structure thus determined, the overall character of the coiling of the polypeptide chain of the N-terminal domain proved to be identical to that of the variable domains of the immunoglobulin Ig-V. Its three-dimensional structure also contains, as can be seen from Fig. 6,

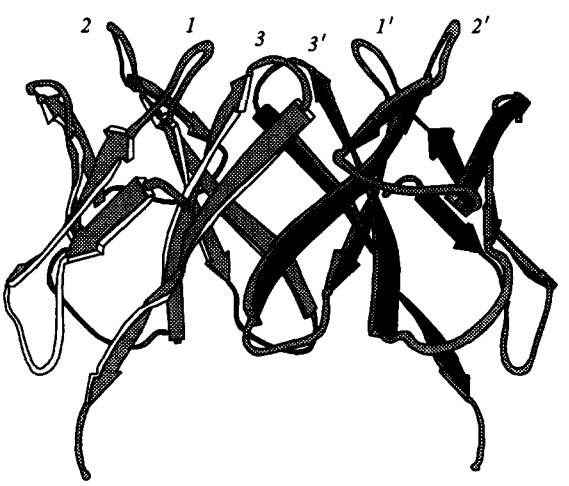


Figure 6. Schematic ribbon illustration of the three-dimensional structure of the CD8-146 homodimer. The numerals *1*, *2*, *3* and *1'*, *2'*, *3'* denote the CDR-1-, CDR-2-, and CDR-3-like loops of the two subunits.[47]

two antiparallel β-pleated sheets made up of five and four β-tiers linked by loops having the same topology. Two domains of the CD8-146 fragment make up a homodimer, similar to the immunoglobulin F_v, with a twofold symmetry axis. However, the electron density maps do not permit the identification of the geometry of the 115 – 141 fragment, which follows the N-terminal domain. The absence of the corresponding clear reflections from the diffraction pattern indicates a disordered state of this fragment in the crystal.

The joint consideration of the steric structure of the CD8 domain discovered and the crystal structure of the MHC-I (HLA-A2) protein established previously by Bjorkman et al.[48] revealed the existence of steric and electrostatic complementarity of the potential energy surface profiles of the two polypeptides and made it possible to reproduce their nonvalence complex and to construct the corresponding atomic model. The binary complex was supplemented by yet another participant — the T-cell receptor (TCR), the surface of which is apparently complementary to the residual unoccupied peptide-binding groove of the HLA-A2 molecule of the target cell. Thus Leahy et al.[47] concluded that the CD8 domain and TCR are linked simultaneously to the same MHC-I molecule in the antibody-antigen complex, while the second domain of the CD8 dimer is linked to a second MHC-I molecule. This scheme of the interactions of the receptors of the killer cell with the MHC-I molecules of the target cell agrees with experimental data.[49]

2. The structure of the extracellular fragment of the CD4 receptor containing domains 3 and 4

It was supposed previously that the three-dimensional structures of the molecules of the CD4 and CD8 receptors are very similar and that their interactions with the molecules of the class II and class I MHC should therefore have the same effect on the activation and development of the T-lymphocytes.[50] However, crystallographic studies on CD4 [51, 52 - 54] and CD8 [47] showed that the three-dimensional structures of these proteins, at any rate in the extracellular region, differ significantly from one another. The principal differences concern two features. Firstly, the N-terminal Ig-V-like domain of CD4 functions in the monomeric state, while the N-terminal domain of CD8 forms the F_v-like homodimer. Secondly, the structure of CD4 in the extracellular region consists of four largely autonomous domains, while CD8 consists of only one domain, linked to the transmembrane component of the receptor by an apparently labile section comprising 48 aminoacid residues, which is virtually undetectable in the diffraction pattern. The only structural feature characteristic of both CD4 and CD8 and absent from the immunoglobulin variable domains is that the length of the CDR2 loop is greater by six aminoacid residues. The functional significance of such local alteration of the sequence remains obscure.

We shall deal in greater detail with the three-dimensional structure of the CD4 receptor. The CD4 molecule is that of a transmembrane glycosylated protein. It is present on the surfaces of T-lymphocytes belonging to the subpopulation which recognises the target cells containing antigens associated with the MHC-II proteins,[27] similarly to the way in which another subpopulation of the T-lymphocytes with CD8 molecules on the surface recognises foreign cells containing antigens linked to the MHC-I proteins.[55 - 57] The cytoplasmic region of CD4 is in close contact with the tyrosine kinase p56, which plays an important role in triggering the production of T-cells. The CD4 molecule is the principal receptor interacting with the human immunodeficiency virus (HIV-1),[58 - 61] which is discussed in detail in Section VI.

On the basis of results of the analysis of aminoacid sequences of CD4 and immunoglobulins, it was predicted that the extracellular part of the receptor would consist of four domains, which have been assigned to the IgSF superfamily.[58, 62] This hypothesis was soon confirmed by the results of direct structural studies. Admittedly, all attempts to carry out the X-ray diffraction analysis of the entire extracellular component of the molecule comprising 371 residues were unsuccessful owing to its poor crystallisation. However, it was possible to obtain crystals giving an entirely satisfactory X-ray diffraction pattern, which were individually made up of the two halves [of one of the two halves? (Ed. of translation)] of this part of the CD4 sequence. Each crystalline specimen incorporated two domains — D1D2 [52, 53] and D3D4.[51, 54] The cryotechnique, which made it possible to determine the three-dimensional structure with a resolution to within 2.6 Å, was used in the crystallographic study of D3D4; at room temperature, the resolution was to within 2.8 Å. Cooling of the crystals to −150 °C was accompanied by a 1% – 2% decrease in the size of the unit cell without disturbance of the symmetry group. The mean square deviation of the C^{α} atoms of the aminoacid residues following the change in temperature from room to −150 °C was 0.86 Å (0.93 Å for the D3 domain and 0.57 Å for the D4 domain). Comparison of the three-dimensional structures of the four domains showed that the three-dimensional structures of the sequences have a series of common features, the D1 and D2 domains resembling more closely the D3 and D4 domains respectively. The differences between the D3D4 and D1D2 geometries consist primarily of the greater distance in D3 between the β-pleated sheets, which is evidently associated with the absence of the disulfide bond in this domain and, secondly, of the change by ~25° in the mutual orientation of the D3 and D4 domains relative to the orientation of D1 and D2.

In the model of the entire extracellular part of the CD4 receptor, proposed by Lange et al.[51] on the basis of the known three-dimensional structures of D1D2 and D3D4 and consistent with a series of indirect data,[63, 64] the four domains have a rod-shaped form with a length of ~120 Å, which is approximately equal to the sum of the lengths of the two D1D2 (65 Å) and D3D4 (60 Å) fragments. The conformation of the intervening section consisting of three residues (179 – 181) remained unestablished. In the D1D2 fragment, these residues are C-terminal and, since they are apparently appreciably mobile in the crystal, they are not manifested on the electron density maps. In the proposed CD4 model (D1D2D3D4), the N-terminal sections of the D1 and D2 domains are accessible to the establishment of contacts with the MHC-II proteins and some sections of the surfaces of the D3 and D4 domains are able to interact with the proteins of the T-cell receptor complex of their own lymphocyte; incidentally, the structures of the two components are so far unknown.

3. The structure of the class II HLA-DR1 antigen

The human MHC antigens HLA (human leukocyte antigens) were first discovered on leukocytes. Like MHC, the antigens of the HLA group can be divided into two classes, I and II, consisting of sets of cell surface glycoproteins. As already mentioned, in transplantation responses, the cytotoxic T-cells mainly respond to foreign class I glycoproteins, whereas the T-helpers respond to foreign class II glycoproteins. The HLA-II molecules are transmembrane αβ heterodimeric proteins with a molecular mass of 60 kDa. They are initially expressed by special antigen-producing cells such as macrophages, stem cells, and B-lymphocytes. The class II of the HLA antigens embraces proteins which are transformed from extracellular antigens into T-cell helpers in the course of the identification of foreign antigens and immune responses.[65, 66] In man, the molecules of this class are polymorphous and can therefore interact specifically with sets of different proteins. The polymorphism of HLA is probably responsible for the different immune responses of different individuals exposed to the action of the same foreign agents; different susceptibilities to diseases, the breakdown of tolerance to own antigens, and the appearance of autoimmune responses are associated with such polymorphism.[67, 68]

The crystals were grown in an aqueous solution of class II HLA-DR1 belonging to several variants of alleles. The protein

was first purified with the aid of immunoaffinity chromatography and detergents and then freed from the transmembrane region and the cytoplasmic domain by treatment with papain.[69,70] The three-dimensional structure of the extracellular part of HLA-DR1 was determined with a resolution to within 3.3 Å by means of a combination of X-ray diffraction data obtained from three different specimens.[70] In all cases, the protein crystallised as the dimer of the DR1 αβ heterodimer. The structure discovered was similar to the molecular structure of HLA-I [71] and was found to correspond to the hypothetical model of the peptide-binding centre of this protein constructed on the basis of crystallographic data for the three-dimensional structure of class I HLA-A2.[72] The three-dimensional structure of the monomeric unit of class II HLA-DR1 in the αβ heterodimer consists of two α-helical domains and two β-structural domains (β_1 and β_2). The molecule is intersected by a long peptide-binding groove, the bottom of which is lined with eight tiers of an antiparallel β-sheet and the walls are lined with two antiparallel helical sections. There are four contact regions between the DR1 molecules in the dimer (Fig. 7). Two of them are located near the two fold axis of rotation and include the interactions of the 49–55 (β_1) and 88–111 (α_2) residues. The other two contact regions, most remote from the membrane, are made up of five pairs of aminoacid residues of two close β_1 domains. The crystal structure of DR1[70] permits the formation of a complex with CD4, postulated previously on the basis of the study of mutagenic derivatives.[73] If in the T-cell recognition of DR1, the latter is in a dimeric form, then, it follows from the X-ray diffraction analysis of the protein that the α_2 domain of the second molecule of the dimer can also be in contact with CD4, giving rise to an additional interaction of DR1 with CD4.

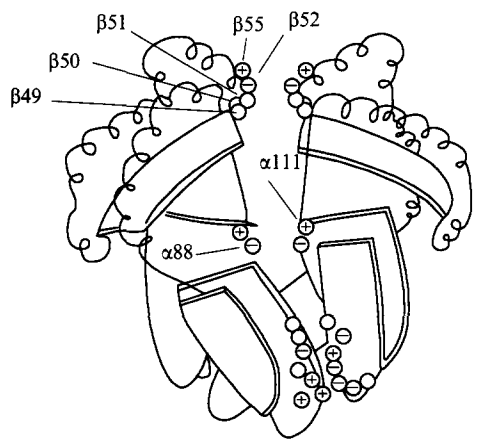


Figure 7. The contact regions of the subunits of the class II HLA-DR1 αβ heterodimer.[70]

Brown et al.[70] proposed a hypothetical mechanism of the assembly of the antigen-specific intercellular complex of HLA-II, TCR, and the CD4 coreceptor. It is suggested that in the first stage the αβ T-cell receptor and its associated CD3 and subunits form weak dimers, similar to the DR1 dimers, which are unstable before the appearance of the DR1–TCR complex. The complex becomes irreversible as soon as the DR1 molecules bind to specific peptides via TCR. The final stabilisation of the complex creates conditions for the appearance of effective interactions of the four-domain CD4 coreceptor simultaneously with both TCR monomers and the MHC-II proteins. The association of CD4 implies the inclusion of the signal system for the triggering of the immune response of the cellular type.

4. The structure of the HLA-DR1 – peptide complex

The antigens of the major histocompatibility complex (MHC or HLA in man) are transmembrane glycoproteins. They bind the peptides within the cell in the cytoplasmic space and supply them to the extracellular space, where they become accessible to the T-cells participating in the mechanism of the functioning of the immune system, in the identification of the foreign antigen, and in the response to its appearance.[74,75] The interaction of the T-cell antigen receptor with the peptide associated with the MHC(HLA) stimulates the expression of the proteins of the major histocompatibility centre in the T-cell. An individual organism has a small set of varieties of such proteins. Furthermore, there are facts indicating that each of them is able to bind an enormous number of pathogenic peptides and of signalling the need for response activities. The apparent contradiction is resolved by virtue of the broad polymorphism of the molecules of the MHC proteins, i.e. the lability of their three-dimensional structures or, in other words, the possibility that individual, predominantly surface sections of the protein globule can assume a large number of different conformational states. The polymorphism of the aminoacid residues of the class I and II HMC proteins forming the binding centres is responsible for their ready adaptability to peptides with different sequences. However, this does not rule out the specificity of the interactions. The MHC-I proteins are as a rule associated with short peptides, usually comprising ten residues, and are distinguished by a high selectivity in relation to aminoacids in fully defined sites of the oligopeptides sequence combined with tolerance in the nature of all other sections of the chain.[76–78] The MHC-I proteins function in the endoplasmic reticulum, where a vigorous peptide-generating proteolytic process takes place simultaneously. The MHC-II proteins act in the endosomal and liposomal compartments, where proteolysis also proceeds simultaneously. Representatives of this class usually bind longer peptides and also exhibit a high sensitivity to the locations of various aminoacids in the sequence.[79–82]

The results of the X-ray diffraction analysis of the extracellular parts of several human HLA-I [83–86] and mouse MHC-I [87,88] proteins indicate the operation of a common mechanism of the binding of peptides by class I proteins. In the structures of all the peptide–protein complexes investigated, the peptides had similar conformations stabilised at the binding centres by what is in principle the same system of hydrogen bonds and by the specific dispersion interactions of the side chains of several residues of the associated peptide with the hydrophobic pockets of the protein globule complementary to them.[86–92]

The binding of peptides to class II MHC (HLA) proteins can be inferred from the results of the X-ray diffraction study by Stern et al.[93] of the three-dimensional structure of the complex between the extracellular part of the HLA-DR1 protein and the peptide of the HA influenza virus — the tridecapeptide fragment (306–318) of haemagglutinin (the membrane glycoprotein of this virus). The diffraction pattern was obtained from a single crystal at −170 °C with the aid of synchronous radiation with $\lambda = 0.910$ Å. The three-dimensional structure of the HLA-DR1 – HA complex was determined with a resolution to within 2.75 Å.

The HA tridecapeptide was sorbed at the HLA-DR1 binding centre in the extended (twisted with a large pitch) form. It represents the best possibilities for interactions, on the one hand, with the aminoacid residues surrounding the peptide and, on the other, with the surface residues of the T-cell receptor. Thus the peptide is open to intercellular interactions. A schematic illustration of the twelve hydrogen bonds formed in the HLA-DR1–HA complex is presented in Fig. 8. The bonds are formed only between the atoms of the main chain of the HA peptide and the atoms of the conservative aminoacid residues of the α-helical and β-structural regions of HLA-DR1. Consequently the system of hydrogen bonds is independent of the particular sequence of the peptide, i.e. is universal for class II HLA-DR1 peptide complexes. This finding also indicates the similar orientations of the peptides in the complexes simultaneously in relation to the binding centres of the HLA-DR1 and TCR proteins belonging to other cells.

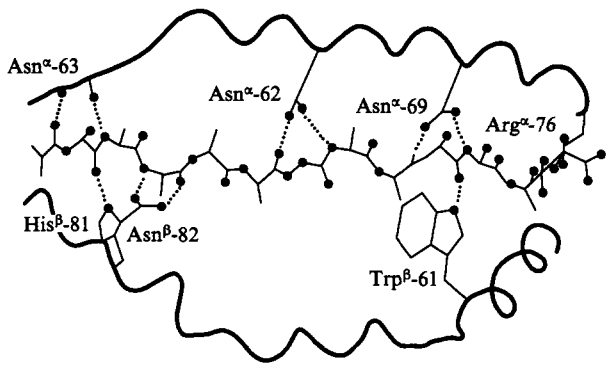


Figure 8. Schematic illustration of the hydrogen bonds between the peptide and the peptide-binding centre of the HLA-DR1 – HA complex.

IV. Aspartate proteases

The first attempt to solve with the aid of X-ray diffraction the problem of the structural organisation of membrane proteins, examined in Sections II and III, ended with relative success. The three-dimensional structures of the periplasmic domains, containing ligand-binding centres, were established for a series of transmembrane receptors and antigen-specific complexes of T-lymphocytes. Further development of studies on this aspect of the field should lead to the determination of the three-dimensional structures of whole receptor molecules, incorporating also transmembrane and cytoplasmic domains. The atomic-molecular foundations for the theoretical and experimental study of the structural-functional organisation of membrane proteins and the mechanisms of their functioning will be ultimately created. The difficulties on this path, which are mainly methodological in character, are still very large.

The latest advances in the X-ray diffraction analysis of proteins, the knowledge of the three-dimensional structures of which is of vital importance for man, are examined below. These are aspartate proteases, primarily the protease of the human immunodeficiency virus (HIV-1) and renin, one of the principal participants in the system for the regulation of the tonus of the blood vessel walls, skeletal muscles, and other tissues. It is becoming increasingly more evident that the search for radical therapeutic means of combating the acquired immunodeficiency syndrome (AIDS), other types of viral diseases, as well as hypertonicity, i.e. the solution of strictly applied medical-biological problems of extraordinary importance, requires a high level of development of theoretical molecular biology. A necessary, although not entirely sufficient, precondition for its attainment is information about the three-dimensional structures of the corresponding protein molecules. In contrast to structural studies on membrane receptors, the retarding factor in this instance is not the difficulties in the isolation, purification, and crystallisation of the proteins but difficulties of a more fundamental character.

Table 1. X-Ray diffraction analysis of native aspartate proteases and their inhibitors.

Enzyme	Inhibitor	Resolution/Å	Ref.
Pepsin	–	2.7, 2.3, 2.0, 1.8	94–98
	AcPhe – TyrOMe	3.0	99
Pepsinogen	–	1.8	100, 101
Chymosin	–	2.3	102–104
Renin	–	2.5, 2.8	105, 106
	His – Pro – Phe – His – Leu – Leu – Tyr – Tyr – Ser	2.5, 2.4, 2.6, 2.0	106–109
Penicillopepsin	–	2.8, 1.8	110–112
	Iva – Val – Val – Lysta – OEt	2.6, 1.8	113, 114
Endothiapepsin	–	3.0, 2.4, 2.1	115–117
	Pro – Thr – Clu – Phe(CH_2NH)Phe – Arg – Glu	2.0	118
	Boc – Thr – His[NHCH(CH_2Ph)CH(OH)N(CH_2CHMe_2)CO]Lys – Phe	1.8	119
	Boc – His – Pro – Phe – His – Sta – Leu – $PheNH_2$	2.2	120
Rhizopuspepsin	–	2.5, 1.8	115, 121
	AcVal – Sta – Ala – StaOH	2.5	122, 123
	D – His – Pro – Pro – Phe – His – Phe(CH_2NH)Phe – Val – Tyr	1.8	124
Proteinase of the bird myeloblastosis virus (BMV)	–	2.4	125
Proteinase of the Rous sarcoma virus (RSV)	–	2.0	126
Proteinase of the human immunodeficiency virus (HIV-1)	–	2.0, 2.8	127, 128
	Val – Ser – Gln – Asn – Leu ψ[CH(OH)CH_2]Val – Ile – Val (U-85548e)	2.5	129
	AcSer – Leu – Asn – Phe ψ[CH(OH)CH_2]Pro – Ile – Val – OMe (JG-365)	2.4	130
	AcThr – Ile – Nle – CH_2 – Nle – Gln – Arg – NHMe (MVT-101)	2.3	131
	PR [a] – Gly – Gly – Ser – Ser – Gly – PR	1.8	132

[a] PR = HIV-1 proteinase.

Before considering the results of the X-ray diffraction analysis of the molecules of the HIV-1 protease, which belongs to the family of the RNA-containing viruses (*Retroviridae*) and of renin (a mammalian proteolytic enzyme), which have attracted exceptional attention in recent years as promising therapeutic targets, we shall deal briefly with the overall state of structural research into aspartate proteases.

The first crystalline structures of aspartate proteases became known in 1977, when the results of X-ray diffraction analysis of the molecular structures of four enzymes were published simultaneously: pepsin, penicillopepsin, endothiapepsin, and rhizopuspepsin (Table 1). The structures were determined with a resolution of 2.5–3.0 Å. About ten years elapsed before the resolution could be improved to 1.8–2.1 Å, which made it possible to infer with adequate reliability the details of the structures both of the enzymes themselves and of the inhibitor and water molecules interacting with them. The structures of a further two aspartate proteases were elucidated during this period — chymosin and renin, as well as the structures of a series of inhibitor complexes of the four enzymes mentioned above.

X-Ray diffraction analysis showed that, despite the significant differences between the aminoacid sequences, the three-dimensional molecular structures of the aspartate proteases of animals and microorganisms are similar. They are all made up of two symmetrically arranged domains which consist almost wholly of β-structures. Both domains reproduce the same structural motif and contain the Asp–Thr–Gly sections, which form part of the active centre of the enzyme. In the contact region, the domains form a groove intersecting a major part of the protein globule. The two catalytically important residues Asp-32 and Asp-215 (the numbering is that for pepsin), the carboxy-groups of which are close to one another and to several other residues as well as the molecules of water of crystallisation, are located at the centre of the groove. The system is stabilised by a network of hydrogen bonds and electrostatic and dispersion interactions (Fig. 9).

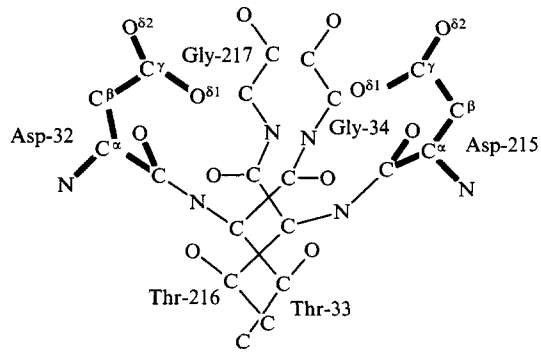


Figure 9. Catalytically important residues of the active centre in aspartate proteinases.

The studies on the three-dimensional structures of the complexes between aspartate proteases and substrate-like inhibitors (Table 1) made it possible to identify the sites of the active centre for each enzyme and to relate them to the sections of the substrate responsible for the primary and secondary specificity of the enzymic reaction. Comparison of the structures of native proteases with their inhibitor complexes showed that the binding of the oligopeptide ligands, differing only slightly from the substrates, is accompanied by the rotation of one domain relative to another by ~4° and its displacement by ~0.3 Å along the axis passing through the C^{α} atoms of the Asp-32 and Asp-215 residues. The change in the mutual orientation of the domains leads to the squeezing of the groove and an appreciable deformation of the active centre, an increase in the distance between certain residues by ~2.0 Å, and a displacement of the side chains by up to 3.5 Å. It was postulated that such changes are necessary for the occurrence of enzymic reactions and hence take place also on formation of enzyme–substrate complexes.[133]

The concept of the purely chemical aspects of the enzymic reactions of aspartate proteases was formulated even before the establishment of protein crystallography. Detailed studies on the nonenzymic hydrolysis of a wide variety of esters, amides, and model peptides, carried out mainly in the 1960s, showed that the catalysis is effective only in a medium incorporating the following three components: a nucleophile attacking the carbonyl carbon atom of the peptide bond to be cleaved, a basic catalyst promoting the elimination of a proton from the nucleophile, and an acid catalyst ensuring the protonation of the leaving group.[134] These investigations as well as the study of the kinetics and thermodynamics of the enzymic reactions of proteases permitted yet another important conclusion of a general character: the chemical mechanisms proper of the activity of all the proteolytic enzymes can be reduced to two types — covalent catalysis and general base catalysis. In the former case, the nucleophile of the protein molecule, interacting directly with the substrate and forming intermediate covalent compounds with the latter, is the component initiating the enzymic reaction. In the proteolytic hydrolytic reactions of the second type, the enzyme does not form covalent compounds with the substrate throughout the entire catalytic step. Its nucleophilic atom group behaves in this case as a basic catalyst, which, by interacting with a water molecule, deprotonates it and thereby converts it into an effective nucleophile, which attacks directly the carbonyl carbon atom of the peptide bond to be cleaved.

The employment by Antonov, Ginodman, and co-workers[134–137] of the method based on the incorporation of the heavy oxygen atom of the $H_2{}^{18}O$ molecule into the products of enzymic reactions has apparently played a decisive role in the determination of the type of catalytic reaction of aspartate proteases. The first comparative studies on the hydrolysis catalysed by various amide hydrolases[134] showed that the serine and cystein proteases function in accordance with the covalent mechanism, while the aspartate and metal-containing proteases function in accordance with the mechanism of the general base catalysis. This conclusion agrees with the most unambiguously interpreted results of chemical, kinetic, and thermodynamic studies on the catalytic reactions of endoproteases. It was subsequently confirmed by the results of the X-ray diffraction analysis of enzymes and is at present generally accepted.

1. The structure of the HIV-1 aspartate protease

Interest in the mechanism of the action of aspartate proteases has grown particularly in recent times in connection with the detection of enzymes of this group in the retroviruses inducing in human and animal organisms diseases such as AIDS, certain types of leukemia, sarcomas, and malignant mammary gland tumours. It is very significant that the aspartate proteases of these viruses play a key role in their life cycles.[138–141] In the host cell, they hydrolyse the proteins of the nuclear capsules surrounding the viral RNA and split the polyprotein chains into mature structural proteins and enzymes participating in the replication of the retroviruses. The viral proteases are also able to hydrolyse the proteins of the infiltrated cells, disturbing thereby the integrity of their structural-functional organisation. On the other hand, the cellular proteolytic enzymes are unable to cleave the viral proteases and to perform their functions. Thus, in the absence of proteases or when the latter are inhibited, the immunodeficiency viruses in man (HIV-1, HIV-2) and in monkeys (SIV), as well as the tumour-generating viruses would be unable to assume infective forms. This is why the proteases of retroviruses have become the object of such constant attention by enzymologists and medical scientists. The aspartate proteases of retroviruses have become some of the most thoroughly

Table 2. Distances (Å) between the atoms of the Asp-32 and Asp-215 side chains (numbering for pepsin) in the crystal structures of aspartate proteases.

Enzyme	$O^{\delta1}_{32}-O^{\delta1}_{215}$	$O^{\delta2}_{32}-O^{\delta2}_{215}$	$C^{\gamma}_{32}-C^{\gamma}_{215}$	$O^{\delta1}_{32}-O^{\delta2}_{215}$	$O^{\delta2}_{32}-O^{\delta1}_{215}$	$C^{\beta}_{32}-C^{\beta}_{215}$	$C^{\alpha}_{32}-C^{\alpha}_{215}$	Ref.
Pepsin	2.8	5.8	5.1	4.4	4.8	7.4	6.7	98
Renin	2.9	6.2	5.3	4.8	4.8	7.4	6.6	105
Penicillo-pepsin	2.9	5.8	5.2	4.4	4.8	7.5	6.6	112
Endothia-pepsin	2.9	5.7	5.2	4.4	4.8	7.6	6.6	117
Rhizopus-pepsin	3.1	6.1	5.4	4.4	5.0	7.7	6.8	121
HIV-1 protease	3.0	5.8	5.2	4.7	4.7	7.4	6.6	128
RSV protease	2.8	5.6	5.0	4.6	4.2	7.4	6.4	126

experimentally investigated enzymes after only 2–3 years since their discovery. Attention has been concentrated on the HIV-1 aspartate protease, about the structure and function of which more is perhaps known nowadays than about a classical object such as pepsin, which was discovered as early as 1836.

The aminoacid sequence of the HIV-1 protease consists of 99 aminoacid residues, whereas the aspartate proteases of vertebrates and microorganisms contain more than 300 aminoacids. The Asp–Thr–Gly fragment, present in each domain in all pepsin-like enzymes, is encountered in this sequence. According to Pearl and Taylor,[142] the physiologically active form of the HIV-1 protease includes two molecules associated via nonvalence interactions. This hypothesis has been confirmed experimentally.[143] A polypeptide chain of 203 aminoacid residues, constituting two molecules of the HIV-1 protease, the C- and N-ends of which are linked by the Gly–Gly–Ser–Ser–Gly pentapeptide section, has been synthesised by genetic engineering methods. This dimer, linked by valence bonds, assumed spontaneously the conformation of the native dimer under physiological conditions and exhibited an activity identical with that of the latter. This conclusion has been recently confirmed in a study[132] where the three-dimensional structure of a similarly linked dimer of the HIV-1 protease was determined with a resolution of 1.8 Å. When one of the two catalytically important Asp-25 residues is replaced by Gly, the activity is lost. Another investigation[144] of the valence-linked HIV-1 protease dimer with a Gly–Gly dipeptide bridge and with the replacement of Asp-25 by Asn and Glu led to the same results.

Tomasselli et al.[139] determined the dissociation constant of the native dimer of the HIV-1 protease, which proved to be 2 pM. According to the data of Cheng et al.,[143] its value is 50 pM. Despite the considerable discrepancy between the dissociation constants obtained, one can nevertheless postulate that the HIV-1 protease dimerises in the nanomolar region, i.e. at a concentration much lower than that in the viral particle. The study of Seelmeier et al.[145] showed that the catalytic activity of the HIV-1 protease is suppressed by pepstatin inhibitors, which is characteristic of all aspartate proteases. It was noted in the same study that the replacement of the Asp-25 residue in the aminoacid sequence of the viral protease by Thr leads to the complete loss of activity. The same effect has been observed following the replacement of the Asp-25 residue in the aminoacid sequence of the HIV-1 protease by Asn,[140,145] Ala,[146,147] Tyr, and His.[146] The activity also falls to zero in pepsin following a similar mutagenesis of one of its Asp catalytic residues.[148]

Thus one may claim that the dimeric form of the HIV-1 protease, in which both aspartic acid residues (Asp-25 and Asp-125) play a catalytically important role, is physiologically active. Only this form possesses all the distinctive features of aspartate proteases. Consequently, like pepsin, renin, rhizopus-pepsin, and other enzymes of the given group, the dimeric form of the HIV-1 protease should function in accordance with a general base catalysis mechanism. To some extent, the conclusion that there is a unique chemical mechanism of the catalytic reactions involving aspartate proteases has been confirmed by the results of the X-ray diffraction analysis of the retroviral proteolytic enzymes.

The three-dimensional structures of three retroviral aspartate proteases (HIV-1, RSV, and BMV) and of three complexes of the HIV-1 protease with substrate-like inhibitors (U-85548e, JG-365, and MVT-101) are known at the present time (Table 1). They are all dimers and are structurally related to one another and to the aspartate proteases from other sources. Furthermore, one can assume that all the known proteases have geometrically identical active centres (these are understood as the direct sites of the sorption of the peptide group to be cleaved). This is indicated by the similarity of the distances between the corresponding atoms of the two functional residues of aspartic acid in the enzyme molecules (Table 2). Thus one may conclude that the enzymic reactions involving aspartate proteases of the vertebrates, microorganisms, and retroviruses have the same initial conditions and hence should proceed by the same mechanism, namely the general base catalysis mechanism.

The crystal structures of the native HIV-1 protease and of its inhibitor complexes are compared in Fig. 10, and Fig. 11 presents a map of the distances between the C^{α} atoms (up to 10 Å) in two associated protease molecules. The numerous and varied, as regards their nature, nonvalence contacts (dispersion and electrostatic contacts and contacts via hydrogen bonds) between two pairs of N-terminal (1–8, 101–108) and C-terminal (87–99, 187–199) fragments make the greatest contribution to the stabilisation of the dimer. The arrangement achieved during

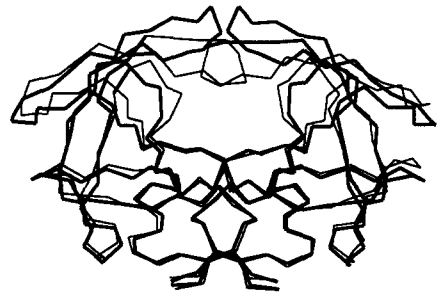


Figure 10. The principal chains of the native HIV-1 protease (thick line) and of its complex with the JG-365 inhibitor (thin line).

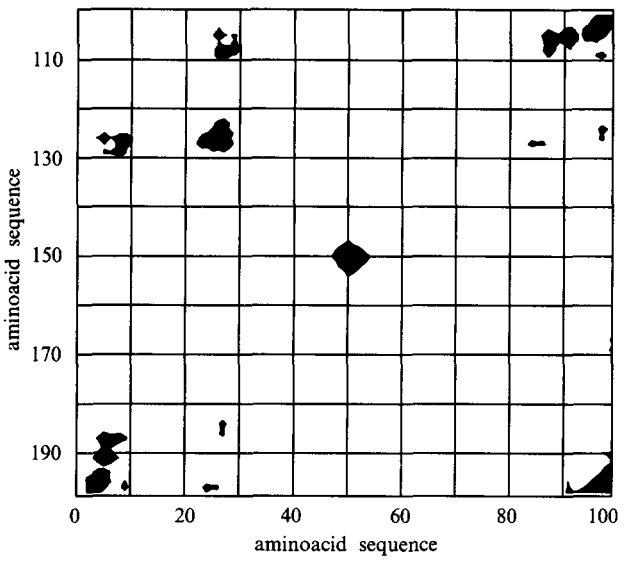


Figure 11. Map of the distances between the C^{α} atoms of the HLA-1 molecules of the protease in the homodimer (the shaded regions correspond to distances < 10 Å).

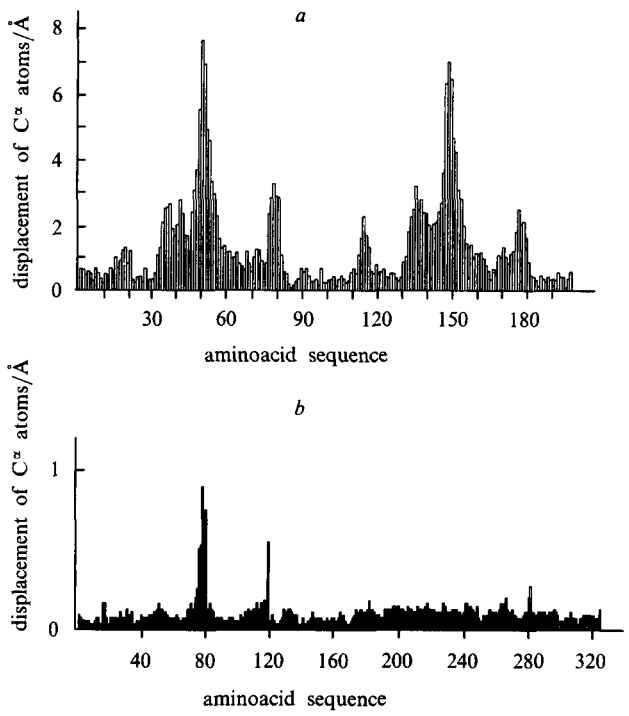


Figure 12. The displacements of the C^{α} atoms in the molecules of the dimeric HIV-1 protease on formation of a nonvalence complex with the JG-365 inhibitors (*a*) and in the rhizopuspepsin molecule on formation of a nonvalence complex with an inhibitor (*b*).

the dimerisation process in this close-packed and apparently most stable part of the HIV-1 protease exerts an appreciable and direct influence on the formation of the active centre in close proximity. The interactions of more than 40 aminoacid residues of the four terminal fragments are extremely selective. This is indicated, for example, by the absence of catalytic activity from the dimeric complex made up of mixed chains of the HIV-1 and HIV-2 viruses, which are closely homologous.[149] Three changes, in the 6, 92, and 95 positions, were found to be sufficient for activity to be lost, the sequences of the enzymes in the two viruses being identical in all other sections which are in contact within the dimer.

The 23–28 and 123–128 fragments also come close to one another in the dimeric complex of the HIV-1 protease. They include the effectively interacting tripeptide fragments Asp-25–Thr-26–Gly-27 and Asp-125–Thr-126–Gly-127—the permanent components of the active centres of all the aspartate proteases. This contact region of the two molecules is stabilised additionally by nonvalence interactions with the 7–9 and 107–109 residues of the N-terminal fragments and the 84–85 and 184–185 residues of the C-terminal fragments.

Finally, there is yet another region of intermolecular interactions in the dimeric HIV-1 protease. It is made up of two β-pleated sheets (45–57 and 145–157) overhanging, like open wings, the active centre (the term 'flaps' has been applied to them). In the native state of the enzymes, they do not participate in the stabilisation of the dimeric form and their position is maintained by the interactions of the 48–52 and 148–152 pentapeptide sections, which are in close proximity (Figs 10 and 11), and also apparently by the contacts with the neighbouring dimer in the crystal lattice.

The flaps play an important role in the stage involving the formation of the nonvalence complex. The enzyme–substrate (inhibitor) interactions entail an enormous, without precedent in the catalytic reactions of other enzymes (at any rate proteolytic enzymes), conformational rearrangement of a considerable proportion of the native structure. It involves not only the side chains of the residues in the active centre, which usually occurs, but also the main chains of both molecules of the HIV-1 protease. The sections 45–57 and 145–157 undergo the greatest changes on complex formation. Judging from the shifts of the C^{α} atoms, they reach ~8.0 Å in the central 50 and 51 and 150 and 151 residues, gradually diminishing to 1.0 Å in the terminal 45 and 57 and 145 and 157 β-structure residues (Fig. 12*a*). The ends of the flaps projecting towards one another, as in cantilevers, are displaced towards the substrate (inhibitor) molecule sorbed in the active centre. The 77–81 and 177–181 fragments undergo synchronously appreciable displacements (2.0–3.0 Å), which also diminish the size of the cavity in the active centre. The displacements of the C^{α} atoms in the two HIV-1 protease molecules in the dimeric complex are similar but not identical.

The changes in the positions of the flaps on formation of complexes were observed also in other aspartate proteases. However, as can be seen from Fig. 12*b*, which illustrates the displacements of the C^{α} atoms in the enzyme–inhibitor complex of rhizopuspepsin, here the differences are much less appreciable. As regards other sections of the main chain of the protein, the latter hardly respond to the association with the inhibitor.

The knowledge of the HIV-1 protease dimerisation process is of considerable interest, primarily in the search for antiviral medicinal preparations. The question whether the association of the molecules occurs without change in their three-dimensional structure or whether the process is accompanied by the conformational adaptation of the partners merits special attention in this instance. If the changes occur, then it becomes possible to prevent the formation of the active complex by a specific action on the enzyme molecules, which would preclude the conformational rearrangements necessary for their combination. It may be that this aspect of the development of an antiviral therapy will prove to be the most promising by virtue of the exceptional nature of the mechanism of the formation of the active forms of the retroviral proteases compared with the aspartate proteases from other sources.

However, it is not possible to observe directly the three-dimensional structure of an individual HIV-1 protease molecule in order to compare it with the structure of a subunit in the dimeric complex, because the enzyme crystallises only in the form of the dimer. Only a theoretical procedure remains for the solution of the problem: a priori calculation of the conformation of a free HIV-1 protease molecule from the known aminoacid sequence.

Fig. 13 presents projections of the three-dimensional structure of one of the subunits in the native dimeric associated species, from which it is seen that the first nine N-terminal and subsequent six C-terminal residues of the sequence do not interact either with one another or with the remainder of the molecule because of their remoteness. The uncoiled conformational states of these fragments are more favourable for the intermolecular interactions which they undergo in the structure of the catalytically active HIV-1 protease dimer (Figs 10 and 11). It was noted above that the intermolecular contacts of the four terminal groups of the two sequences actually make the main contribution to the stabilisation of the dimer. On the other hand, the uncoiled forms of the 1–9 and 94–99 sections are energetically unfavourable for the free enzyme molecules and their occurrence in a dilute solution is unlikely. Consequently, the dimerisation process should involve a conformational transition from the coiled forms, favourable for the free molecules, to the uncoiled forms preferred by the associated molecules. It is seen from Fig. 13 that the conformational state of the 45–57 flap should change simultaneously. In the complex, this state is determined by the interaction of a second molecule with the flap and possibly by the interactions with the dimer of an adjoining unit cell.

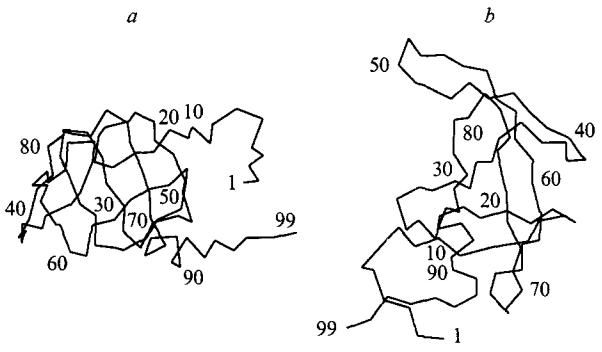


Figure 13. Projections (*a*, *b*) of the HIV-1 protease molecule in the dimer.

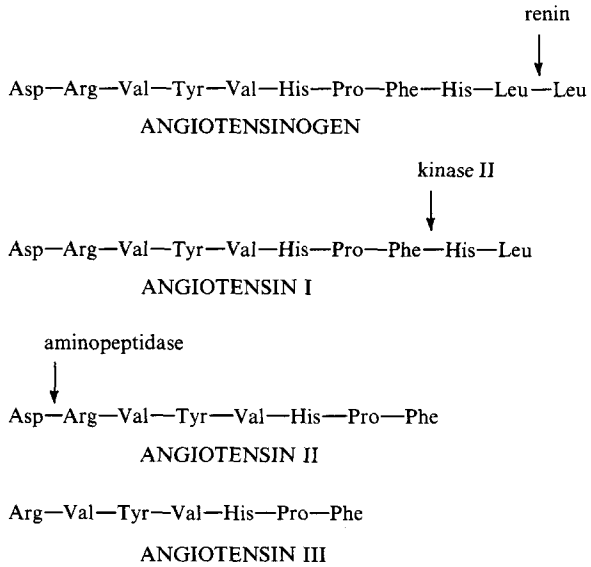
From the foregoing, one may conclude that there is the possibility of a search for therapeutic agents for protection against retroviruses, in particular the human immunodeficiency virus — of devising preparations which would prevent the dimerisation of the molecules of aspartate proteases. Without this, the functioning of the enzyme and hence the operation of the life cycle of the virus are impossible.

Another approach to the preparation of the necessary medicines involves the creation of special inhibitors of the retroviral aspartate proteases. This is an extremely complex task and at the present time has not yet been achieved. Among the requirements which must be met by the properties of the inhibitors, the main and most difficult one concerns the selectivity of their action. Inhibitors exhibiting a therapeutic effect should be primarily highly specific to such an extent that, while deactivating the retroviral protease, they do not disturb the normal functioning of both aspartate and other proteolytic enzymes of the host cell. The specific search for inhibitors satisfying this requirement requires the availability of quantitative data on all the stages of the catalytic step performed by the viral protease, similar to the data concerning the mechanisms of the functioning of the proteases of the infiltrated cell and the methods for the solution of the reverse structural problem, i.e. the development of the chemical structure of the inhibitor on the basis of the specified three-dimensional form. In connection with the problem of the specificity of the enzyme–substrate interactions, X-ray diffraction data on renin — the aspartate protease of blood plasma — are especially important.

In contrast to not only the other aspartase proteases but in general amidohydrolases, this enzyme exhibits a virtually absolute specificity. Renin is known to hydrolyse only its own natural substrate angiotensinogen, abstracting from its N-terminal part of a decapeptide, as well as a small number of synthetic analogues which are highly homologous with respect to the N-terminal aminoacid sequence of angiotensin.[134,150] Almost any aminoacid replacement in the pentapeptide fragment of the substrate before the bond to be cleaved ($P_5 - P_1$) and in the tetrapeptide section beyond it ($P'_1 - P'_4$) leads to a nonproductive Michaelis complex and the absence of hydrolysis. The reason for such phenomenal specificity of renin still remains unelucidated.

2. The structure of renin and of its inhibitor complexes

Renin, a glycosylated protein, forms part of the enzyme system which effects the biogenesis and decomposition of the octapeptide tissue hormone angiotensin II, the most powerful of the known pressor agents in the blood circulation system. Angiotensin II stimulates the narrowing of the peripheral arterioles throughout the organism and hence increases the arteriole pressure. Apart from this, it activates the secretion of a series of hormones (mainly aldosterone) and influences the work of the heart, the liver, the central and peripheral components of the nervous system, etc. Its direct precursor, angiotensin I, is formed from the globular blood protein angiotensinogen, already mentioned, as a result of the abstraction from it of the N-terminal decapeptide by renin. According to the scheme presented below, the hydrolysis of the peptide bond between the Leu-10 and Leu-11 residues is the first and rate-determining stage in the cascade of enzymic reactions leading to the conversion of angiotensinogen into angiotensin II with its subsequent decomposition.



The three-dimensional structure of renin has been investigated in a number of studies.[105–109] The three-dimensional structure of the partially deglycosylated native human recombinant renin has been identified with a resolution of 2.5 Å. Some of the peptide sections on the periphery of the globule were disregarded on the electron density maps and for this reason they were assigned to the group of conformationally labile loops lacking defined structures. The X-ray diffraction analysis of the glycosylated native human renin and of its inhibitor complexes was subsequently carried out[106] with resolutions of 2.8 and 2.4 Å respectively. The structure of the enzyme–inhibitor complex of this protein was also determined in another study[109] with a resolution of 2.6 Å. The stabilising interactions of specific pockets of the active centre of the enzyme with the side chains of the $P_4 - P'_4$ residues on the octapeptide sections of the inhibitors were described in the above communications.

Table 3. Dihedral angles of rotation φ, ψ, and ω (deg) of the main chain of the CH-66 inhibitor in the active centre of renin and of the U-85548t, JG-365, and MVT-101 inhibitors in the active centre of the HIV-1 protease.[108,129-131]

Designation of residue	Renin				HIV-1 protease											
	residue	CH-66			residue	U-85548e			residue	JG-365			residue	MVT-101		
		φ	ψ	ω		φ	ψ	ω		φ	ψ	ω		φ	ψ	ω
P_5	His	−76	114	−170	Val	–	11	175	–	–	–	–	–	–	–	–
P_4	Pro	−81	147	−175	Ser	−106	126	178	Ser	−64	169	180	–	–	–	–
P_3	Phe	−103	123	180	Gln	−108	148	−177	Leu	−123	138	178	Thr	–	125	−176
P_2	His	−111	92	−178	Asn	−138	101	−178	Asn	−143	104	180	Ile	−121	92	−178
P_1	Leu(ψ)	−101	44[a]	167[a]	Leu(ψ)	−97	63[a]	133[a]	Phe(ψ)	−94	47[a]	151[a]	Nle(ψ)	−92	72[a]	142[a]
P_1'	Leu	−89[a]	130	177	Val	−56[a]	150	173	Pro	−57[a]	174	175	Nle	−92[a]	154	−179
P_2'	Tyr	−75	133	174	Ile	−139	124	177	Ile	−130	122	−177	Gln	−151	123	178
P_3'	Tyr	−130	145	−174	Val	−126	157	–	Val	−143	138	–	Arg	−81	129	–
P_4'	Ser	−85	–	–	–	–	–	–	–							

[a] The angles correspond to the nonstandard P_1-P_1' sections of the inhibitor chain.

The crystal structure of the submaxillary mouse renin, inhibited by the substrate-like CH-66 decapeptide, has been studied in greatest detail and with the best resolution (2.0 Å).[151]

Piv-His-Pro-Phe-His-[Leu(ψ)-CH(OH)CH_2]-Leu-Tyr-Tyr-Ser-NH_2

($P_6-P_5-P_4-P_3-P_2-P_1$——————$P_1'-P_2'-P_3'-P_4'$),

Here Piv is a pivaloyl blocking group and CH(OH)CH_2 is a hydroxyethylene group with the (*S*)-configuration of the second asymmetric carbon atom of Leu(ψ). The positions of 10038 and 325 aminoacid residues of the glycosylated enzyme, 364 atoms of the inhibitor, and 613 oxygen atoms of the water of crystallisation were determined. The mean square error in the coordination of the atoms was 0.21 Å and the standard deviations from the standard bond lengths and angular distances were 0.026 Å and 0.54 Å respectively.

The three-dimensional structures of the renins investigated have the same characteristic features as other aspartate proteases. They are also structural proteins consisting of two domains. Between them, there is a substrate-binding groove, in the centre of which the catalytically important Asp-32 and Asp-215 residues are accommodated and, as in other enzymes of this group, these residues are followed by Thr and Gly. The three-dimensional structures of the renin molecules from different sources are very similar. For example, the difference in the orientations of the domains in the structures of mouse and human renins does not exceed 1.7°. The positions of the domains in pepsin and endothiapepsin differ from that in mouse renin by 6.9° and 18.8° respectively.

The decapeptide ligand in the enzyme–inhibitor complex has the β-pleated sheet conformation extended along the substrate-binding groove (Table 3).[128-131,151] Its carboxyethylene group between two Leu residues occupies in the active centre of the enzyme a sensitive site of the substrate and forms hydrogen bonds with the carboxy-groups of the Asp-32 and Asp-215 catalytic sections. The conformational states of the main and side chains of all the inhibitor residues in the complex with renin correspond to the lowest-energy forms of the free aminoacids and, being uncoiled structures, provide at the same time the best conditions for effective interactions with the residues of the active centre of the enzyme. Table 3 lists the dihedral angles φ, ψ, and ω, which determine the geometry of the main chain of the renin inhibitor (CH-66)[151] and three inhibitors of the HIV-1 aspartate protease (U-85548e, JG-365, MVT-101).[129-131] It follows from the comparison of the data in Table 3 that, on association with the active centres of the mammalian and retroviral aspartase proteases, the completely different inhibitors assume similar conformational states. At any rate, the differences between the dihedral angles of the inhibitors of renin and the HIV-1 protease are not greater than the differences between the angles of the inhibitors of a single enzyme. This finding as well as the data in Table 2 indicate that the structural states of the substrate-binding centres of the aspartate proteases are similar regardless of their source.

X-Ray diffraction analysis of the inhibitor complex between mouse renin and the decapeptide[151] made it possible to identify the aminoacid residues of the enzyme which form specific pockets and to make certain quantitative assessments of their interactions with the side chains of the inhibitor residues. Bearing in mind the similarity of the latter to the N-terminal aminoacid sequence of angiotensinogen, the data presented in this instance are of undoubted interest for the subsequent theoretical study of the nonvalence interactions of renin with a true substrate and for the solution of the problem of the nature of its exceptional specificity.

Table 4 presents the 34 renin residues which form ten substrate-binding pockets (S_6-S_4'), the interatomic contact distances between which and the inhibitor in the complex do not exceed 4.2 Å. The residues presented in Table 4 are located on the surface of the globule of the water-soluble protein and yet they are all hydrophobic (with the natural exception of the catalytically important Asp-32 and Asp-215), like the substrate

Table 4. The aminoacid residues of renin, the atoms of which form close contacts with the P_6-P_4' residues of the CH-66 inhibitor (< 4.2 Å).[151]

Designation of pocket	Renin residues forming substrate-binding pockets
S_6	Leu-10, Ser-12, Phe-220, Leu-276, W7, W8, W138
S_5	Ser-12, Pro-111, Leu-114, Ser-219, W7, W8
S_4	Ser-219, Phe-220, His-287, W77, W138, W147
S_3	Gln-13, Ser-77, Pro-111, Leu-114, Ala-115, Phe-117, Gly-217, Ser-218, Ser-219, W7, W11, W77
S_2	Tyr-75, Gly-76, Ser-77, Ser-218, Ser-222, His-287, Met-289, Val-300, W77, W147
S_1	Asp-32, Gly-34, Tyr-75, Phe-117, Val-120, Asp-215, Gly-217
S_1'	Gly-34, Ile-73, His-74, Tyr-75, Gly-76, Glu-128, Val-130, Val-213, Asp-215, Ile-291, Val-298, Val-300, W76, W87
S_2'	Gly-34, Ser-35, Ile-73, His-74, Tyr-75, Glu-128, Val-130, Ile-291, Thr-295, W28, W73, W74, W76
S_3'	His-74, Tyr-75, Gly-76, Ile-291, Pro-292, Thr-295, W74, W75
S_4'	Pro-294, Thr-295, W75

Table 5. Changes in the contact area of the renin active centre residues accessible to the solvent during inhibition by CH-66.[151]

Residue	Area/Å²			Residue	Area/Å²		
	free renin	CH-66 complex	surface area accessible to solvent		free renin	CH-66 complex	surface area accessible to solvent
Leu-10	37.3	30.1	7.2	Gln-128	14.0	5.5	8.5
Ser-12	7.1	0.2	6.9	Val-130	14.2	9.9	4.3
Gln-13	7.3	1.0	6.3	Val-213	5.1	0.8	4.3
Ile-30	4.8	1.8	3.0	Asp-215	4.8	0.0	4.8
Asp-32	3.2	0.0	3.2	Ser-218	5.7	0.0	5.7
Ser-35	2.7	0.0	2.7	Ser-219	10.2	0.4	9.8
Ile-73	3.3	0.1	3.2	Phe-220	12.4	2.5	9.9
His-74	33.1	19.2	13.9	Leu-276	12.7	7.1	5.6
Tyr-75	11.7	0.0	11.7	Thr-283	9.3	5.7	3.6
Gly-76	4.2	0.3	3.9	His-287	16.0	4.8	10.2
Ser-77	15.2	4.2	11.0	Met-289	18.5	14.3	4.2
Ile-110	36.1	32.2	3.9	Ile-291	10.3	0.5	9.8
Pro-111	13.1	1.2	11.9	Pro-292	30.0	19.0	11.0
Phe-112	3.6	0.6	3.0	Pro-294	37.6	29.5	8.1
Leu-114	34.9	21.5	13.4	Thr-295	25.4	10.0	15.4
Phe-117	7.2	0.1	7.1	Val-300	5.6	0.2	5.4
Val-120	3.6	0.0	3.6				

residues interacting with them, so that the complexes are stabilised mainly via dispersion forces.

The highly complementary nature of the potential energy surfaces of the active centre of renin and of the decapeptide inhibitor (substrate) and hence their predisposition to undergo effective nonvalence interactions can be inferred from the numerous interatomic contacts between the enzyme and the ligand. This is confirmed quantitatively by the data presented in Table 5. They demonstrate a sharp contraction, on sorption of the inhibitor, of the surface (accessible to the solvent) of the aminoacid residues forming the S_6-S_4' loci. Approximately half of the residues, including Asp-32, and Asp-215, and certain residues of the flap, behave as ligands fully shielded from contact with the external medium. The overall area of the active centre accessible to the solvent, which is ~460 Å², diminishes to ~220 Å² in the complex.

The accessible surface areas of the side chains of the P_5-P_4' residues of the CH-66 inhibitor in two complexes — with renin and with endothiapepsin — are compared in Table 6. According to a calculation,[151] this area is ~64 Å² in the former case and ~130 Å² in the latter. The data in this table indicate an appreciably greater packing density of the same inhibitor in the active centre of renin compared with that of endothiapepsin. Exceptionally pronounced complementary conditions in the renin complex occur in the middle part of the decapeptide ligand, i.e. in the region of the peptide bond being cleaved.

Table 6. Comparison of the accessible surface areas of the side chains of the P_5-P_4' residues of the CH-66 inhibitor in complexes with renin and endothiapepsin.

Designation of residue	Residue	Surface area of complex accessible to solvent/Å²	
		renin	endothiapepsin
P_5	His	20.6	35.7
P_4	Pro	7.7	11.3
P_3	Phe	0.4	6.9
P_2	His	0.5	8.7
P_1	Leu(ψ)	0.0	0.5
P_1'	Leu	0.6	0.9
P_2'	Tyr	7.8	7.2
P_3'	Tyr	12.4	38.0
P_4'	Ser	13.9	20.2

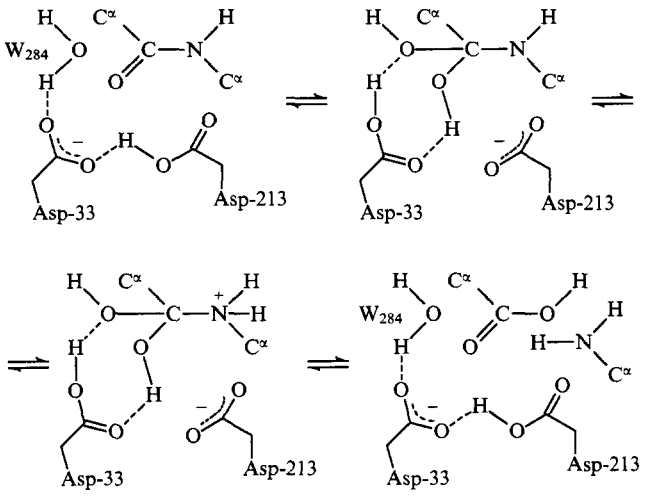
We shall attempt to answer the question of the advances contributed by X-ray diffraction analysis in the study of aspartate proteases and to what extent the knowledge of the three-dimensional structures of these enzymes and of their inhibitor complexes can deepen our understanding of the mechanisms of the catalytic reaction of aspartate proteases. The answer to this question is of general importance for enzymology, because, as already mentioned, proteases are objects in biocatalysis which have been most thoroughly investigated in all respects. We shall consider the hypothetical models of the mechanisms of the action of aspartate proteases, which have been developed on the basis of data on the three-dimensional structures of the enzymes and their complexes with substrate-like inhibitors.

3. The mechanism of the catalytic reaction of aspartate proteases

The first stereochemical model was proposed in 1985 by James and Sielecki,[112] who used for this purpose the crystal structure of penicillopepsin, solved with a resolution of 1.8 Å.[111,112] A series of considerations, based on the knowledge of the three-dimensional structures of the enzymes, had been put forward earlier concerning the mechanism of the action of aspartate proteases.[111,123,152-158] However, they were all to some extent incomplete and did not claim to provide a consistent description of the entire process involving the hydrolysis of the peptide bond catalysed by aspartate proteases. The W_{284} molecule of the water of crystallisation plays an important role in the stereochemical model of the productive covalent Michaelis complex devised by James and Sielecki. In the active centre of penicillopepsin, it occupies a site which is extremely convenient also for the formation of an effective hydrogen bond with the side chain of Asp-33, which increases its nucleophilicity, and for attack on the carbonyl carbon atom of the bond being hydrolysed, which weakens the latter. The protonated carboxy-group of the Asp-213 residue, polarising the carboxy-group of the substrate via a hydrogen bond and hence acting as an electrophilic component of the system, acts in the same direction (Scheme 1).

The three interactions noted above initiate the disruption of the delocalised $\pi-n$ electron system of the peptide bond being cleaved, the pyramidalisation of the nitrogen and carbonyl carbon atoms, and the appearance at the latter of *gem*-hydroxy-groups, which are involved in hydrogen bonding with the protonated carboxy-group of Asp-33. The return of the lone electron pair to the nitrogen atom enhances its tendency to be protonated and, according to James and Sielecki,[112] can be

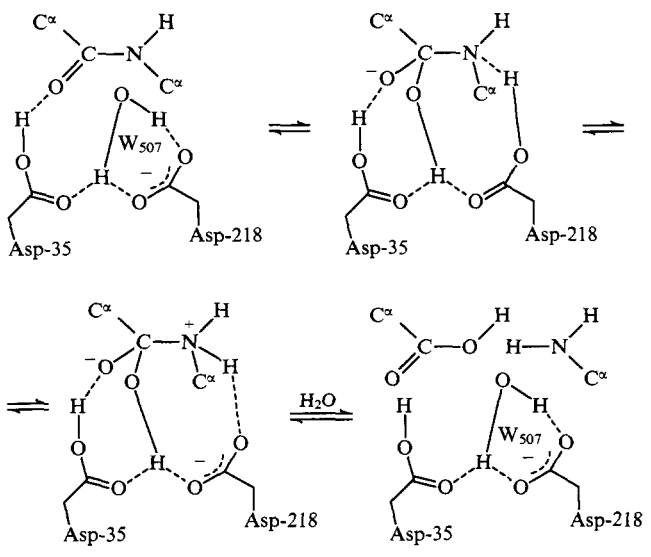
Scheme 1



achieved in two equally probable ways. In one of them, the inversion of the configuration of the bonds involving the nitrogen atom, occurring via the rotation of the amino-group of the substrate by ~60° around the C–N bond, is postulated. As a result, the lone pair is oriented towards the solvent and the nitrogen is readily protonated, particularly in aspartate proteases with a highly acid pH. The alternative procedure involves the transfer of a proton to the nitrogen atom from one of the hydroxy-groups of the tetrahedral adduct, which is in the *syn*-configuration relative to the lone pair of the nitrogen atom. The compound formed is unstable and decomposes into two parts, one of which includes the amino-group and the other the carboxy-group.

The initial hypothesis of James and Sielecki[112] that the nucleophile in the catalytic reaction of aspartate proteases is the W_{284} group has been rejected by Suguna et al.,[124] who hold a different view on the mechanism of the action of the enzymes. The stereochemical model which they constructed is based on crystallographic data for the three-dimensional structures of native rhizopuspepsin[122] and its complex with an octapeptide inhibitor having a reduced peptide bond, which is cleaved on interaction with substrates. Suguna et al.[124] chose as the nucleophile the W_{507} water molecule (W_{39} in terms of the numbering for penicillopepsin), which is located between the two carboxy-groups of the Asp-35 and Asp-218 residues and forms three hydrogen bonds with the latter (Scheme 2). It is suggested

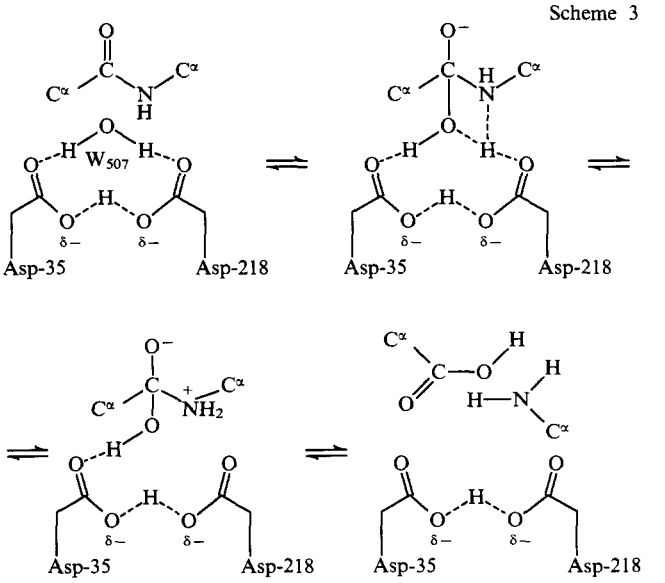
Scheme 2



that the W_{507} molecule is not displaced by the substrate because, following its sorption, the peptide bond to be hydrolysed is subjected to steric hindrance and assumes a nonplanar energetically unfavourable configuration. In this position of the substrate, the carbonyl carbon atom of the bond being cleaved is located opposite the attacking oxygen atom of the W_{507} molecule and the oxygen of the carbonyl group is placed at a distance from the hydroxyl of the Asp-35 side chain which is optimal for the formation of a hydrogen bond. As a result of the formation of the tetrahedral adduct, the OH group of the water molecule passes to the substrate, while the hydrogen atoms initially pass to the carboxy-group of the Asp-218 residue and then to the nitrogen atom. The reorganisation of the intermediate leads to the rupture of the peptide bond of the substrate, the desorption of the reaction products, and the regeneration of the active centre.

In the mechanism of the catalytic reaction of aspartate proteases proposed by Pearl,[159] the nucleophile is the W_{507} water molecule, as in Scheme 2 examined above (Scheme 3). The new feature in this model consists in the hypothesis of the equality of the acidity constants and hence of the electronic equivalence of the carboxy-groups of the Asp-35 and Asp-218 residues having a delocalised proton and, furthermore, in the introduction as an electrophilic component, giving rise to an additional polarisation of the carbonyl oxygen atom of the bond being cleaved, of the two NH groups of the Gly-78 and Asp-79 residues (according to the numbering for rhizopuspepsin), which enter into the aminoacid sequence of the flap overhanging the substrate. According to the previous models of the mechanism of the functioning of aspartate proteases, the effects of the nucleophile and the electrophile lead to the formation of an unstable intermediate compound, which decomposes from the N- and C-ends. In both this case and in those discussed above, the author follows the concept of stress and enforced deformation, i.e. he postulates the possibility of the transformation of the energy of enzyme–substrate nonvalence interactions into an energy of distortion of the peptide bond, which brings it closer to the transition state and hence facilitates the hydrolysis.

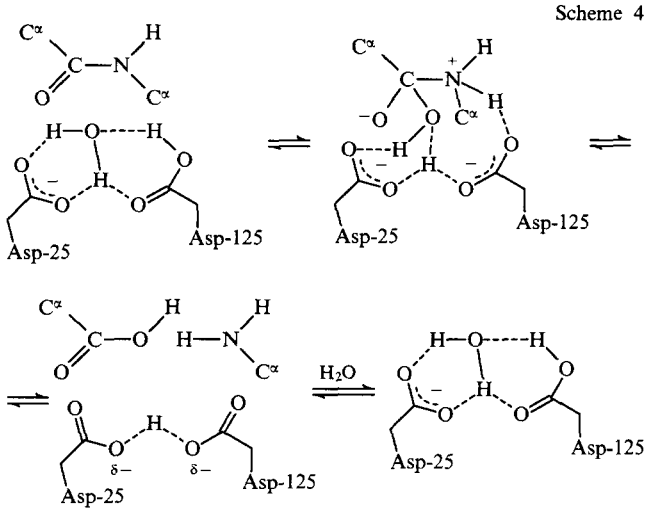
Scheme 3



The model of the mechanism of action of aspartate proteases proposed by Jaskolski et al.[129] is based on the three-dimensional structure of the HIV-1 retroviral protease. The W_{507} molecule (the numbering is that for penicillopepsin), located in an equidistant position between the side chains of the Asp-25 and Asp-125 residues (Scheme 4), was chosen as a nucleophile in this case also. Bearing in mind the symmetry of the enzyme, the authors placed one of the hydroxy-groups of H_2O along the twofold axis in such a way that it is able to form a

forked hydrogen bond with the inner carboxyl oxygen atoms of the aspartic acid residues. The second hydroxy-group of H_2O is directed towards the outer oxygen atom of the Asp-25 carboxylate. The acid proton of the unionised Asp-125 side chain forms a hydrogen bond with the oxygen atom of the W_{507} molecule of the water of crystallisation. Thus, according to the dynamic model,[129] the native HIV-1 protease in the equilibrium state contains a symmetrical system of five oxygen atoms, in the field of which the hydrogen atoms migrate in an appropriate manner. The symmetry of the active centre breaks down on interaction of the enzyme with the substrate. As a result of nucleophilic attack, there is a simultaneous disruption of the carbonyl group of the bond being cleaved, the hybridisation of the carbon and nitrogen atoms changes ($sp^2 \rightarrow sp^3$), a C–OH valence bond is formed, and the nitrogen is protonated. In other words, the enzyme–substrate complex passes in one stage through an unstable intermediate state, which is completed by the formation of the reaction products and the return of the enzyme to its initial state.

Scheme 4



Despite the significant differences between the models proposed, they are unified as regards their initial postulates.

Firstly, it is suggested that the aspartate proteases function through general catalysis and therefore do not form covalent intermediate complexes with the substrate during the reaction.

Secondly, it is claimed that the manifestation of the catalytic activity of proteases required the presence in the system of a nucleophile — basic and acid catalysts (the differences concern their selection).

Thirdly, it is postulated in all the hypotheses that the enzymic reaction involving the hydrolysis of the peptide bond proceeds via a stage of an unstable intermediate (a tetrahedral adduct).

Fourthly, a common feature of all the proposed stereochemical models is the treatment of the stimuli to the functioning of aspartate proteases based on the concept of deformation and stress.[160]

These postulates were formulated and confirmed experimentally in the study of enzymic and model nonenzymic reactions even before the X-ray diffraction analysis had been carried out on even one aspartate protease. The determination of the three-dimensional structures of the enzymes and their complexes with substrate-like inhibitors provided additional confirmation of some of the earlier conclusions and made it possible to visualise the ideas already established.

Thus the mechanism developed for enzyme catalysis are interesting in the sense that they arose, in essence, from the same experimental data, including X-ray diffraction data obtained in recent years, and are based on the same theoretical ideas concerning the nature of biocatalysis which arose before the establishment of protein crystallography and have long become traditional. The imperfection of the generalisations made is indicated by the fact that, whilst the initial experimental data and theoretical foundations were the same and also a single approach was followed, not one but four different models of the functioning of aspartate proteases have been proposed and their number may, incidentally, be even greater. They do not form a series indicating a quantitative development of knowledge about the mechanism of the action of enzymes, but are more likely to represent different points of view. The hypotheses examined correspond to a single level of understanding of the phenomenon investigated and are equivalent as regards the arguments adduced and their predictive power. The question arises what new advances have been contributed by the X-ray crystallography of proteins to the study of the catalytic reactions of aspartate proteases and other enzymes.

In the first place, one should note that X-ray diffraction data have made it possible to determine not only the region of the active centre of the enzyme but also its geometry at the level of atomic coordinates. It would at present be impossible to achieve this quite so fully and reliably by any other procedure. Furthermore, the use of this method to solve the three-dimensional structures of the native enzyme and its complexes with various inhibitors made it possible to discover the conformational rearrangements of the active centres which may occur at different stages of the catalytic step. These data are also unique and informative.

X-Ray diffraction played a major role in the study of the specificity of enzyme–substrate interactions and of the locations of the regions of the active centres which bind to the sensitive sites of the substrate and the residues adjoining it on either side. As regards the mutual orientations of the bond being cleaved and of the catalytically active residues of the enzyme corresponding to the productive binding of the true substrate, the problem remains unsolved. As can be seen in relation to the four different stereochemical models of the functioning of aspartate proteases which have been proposed, the views expressed can be divided into more or less probable, but, despite the knowledge of the conformations of the enzyme in the native state and in the complex with the inhibitor, both are hypothetical in character.

The data concerning aspartate proteases examined here as well as the similar data on chymotrypsin, trypsin, carboxypeptidase, and lysozyme described previously[161] provide grounds for the conclusion that the appearance of unique quantitative information about the three-dimensional structures of enzymes and their complexes did not entail a conceptual development of enzymology and revision of the existing ideas about the nature of biocatalysis.

With the advent of X-ray diffraction analysis of enzymes, there was no transition from speculative ideas about enzyme catalysis to a rigorous quantitative description of this phenomenon. Nor has there been any change in the direction of biocatalytic investigations, which, as before, follow from function to structure, which is not fortuitous because X-ray diffraction studies can only lead to the knowledge of the morphology of biosystems at the atomic-molecular level, which by itself is not the final aim of the study of enzymes. The morphology of the object is always a preliminary feature which is completely necessary for the subsequent investigation of the structural and structural-functional organisation of biosystems. For this reason, the three-dimensional structures of many hundreds of enzyme molecules discovered with the aid of X-ray diffraction did not solve the problem of biocatalysis, but led to a real development of an approach to the study of enzyme catalysis in the direction from structure to function and to an a priori quantitative description of the mechanism of the catalytic reaction. By virtue of the development of the crystallography of proteins, the problem of creating a general theory of biocatalysis and the corresponding computational methods acquired for the first time the form of a truly scientific problem soluble at the present level of scientific knowledge.

V. DNA-binding proteins

This section describes certain crystallographic studies on proteins carried out in recent years. The studies examined below demonstrate the latest advances in crystallography which have made it possible to solve at the atomic level the three-dimensional structures of proteins and complexes consisting of thousands and more of aminoacid residues. The technical development of the method was manifested in the transition noted at the beginning of the 1990s, from X-ray diffraction studies on individual protein molecules to the no less detailed investigation of the three-dimensional structures of macromolecular protein oligomers and their aggregates with other compounds and subcellular systems. In other words, structures of more complex biological systems began to be understood at the atomic level.

Objects which were of interest as regards their structural-functional organisation and at the same time illustrated clearly the current advances in protein crystallography were chosen for consideration. These undoubtedly include the DNA-binding proteins, which participate directly both in the establishment of the genetic apparatus and in all the stages of its functioning. In eukaryotic cells, such proteins are traditionally subdivided into two groups: the histone proteins participating in the establishment of a definite structure of DNA within the cell nucleus and the nonhistone proteins ensuring the biosynthesis of RNA. The results of X-ray diffraction analysis of histones and β-galactosidase, which are representatives of both groups of the DNA-binding proteins, are considered below.

1. The structure of the histone core of the nuclesome

Histones constitute five varieties (H1, H2A, H2B, H3, and H4) of small structural proteins, the aminoacid sequences of which contain respectively 220, 128, 124, 134, and 102 residues. They ensure the close packing of the DNA double helix, which is very long in the extended state. For example, in each human chromosome it is on average 5 cm long. Histones are necessary primarily for the ordered disposition of the long double-stranded polynucleic acid within the small volume of the cell nucleus. However, this is not their only purpose: the nature of the packing of DNA influences the activity of the corresponding sections of the genome. Consequently its histone structural organisation is one of the ways of regulating and controlling the transcription of RNA from DNA. The aminoacid sequences of histones contain about 40 positively charged Lys and Arg residues, which enables them to bind effectively to the DNA double helix regardless of its nucleotide composition.

Histones, especially H2A, H2B, H3, and H4, are some of the most conservative of the known proteins. The rates of change in the aminoacid sequences for a series of proteins and the histone H4 throughout the period of evolution are compared in Table 7. The figures presented suggest that approximately four out of each five random aminoacid replacements are harmful in haemoglobin and 17 out of each 18 replacements are harmful in cytochrome *c*, while in histone H4 there are virtually no harmless replacements.

Table 7. *Changes in the aminoacid sequences of proteins throughout evolutionary time.*[162]

Protein	Unit of evolutionary time/million years[a]
Fibrinopeptide	1.2
Haemoglobin	6.1
Cytochrome-*c*	21.0
Histone H4	600.0

[a] The unit of evolutionary time is the average time necessary for the appearance of one acceptable replacement in the given protein for each 100 aminoacid residues.

Since all the DNA base pairs are equally subject to random mutations, the evolutionary stability of histones suggests that they play an extremely responsible role in the organisation and functioning of the genetic apparatus.

Histones constitute about half of the mass of the chromosome, where they participate in the organisation of several levels of packing of the DNA double helix. Together with other proteins, histones form complexes with DNA, which are called chromatin. In 1974, Kornberg isolated the repeating structure of the chromatin unit for the first time[163] and later established[164] that it consists of an octameric protein core comprising two molecules of each of the histones H2A, H2B, H3, and H4 and a fragment of the DNA double helix. It was postulated[165] that the double-stranded DNA is twisted around the histone core and forms two turns of a superhelix made up of 165 base pairs. This figure was later corrected to 146.[166] The observed chromatin structural unit was called a nucleosome.[167] Studies on the histone core with the aid of different physicochemical methods showed that the octamer consists of a heterogeneous protein ensemble $(H2A-H2B-H3-H4)_2$ comprising three structural subunits: a tetramer $(H3-H4)_2$ and two dimers (H2A–H2B). The DNA double helix in chromatin extends as a continuous filament from one nucleosome to another. The linear DNA linker sections located between the nucleosomes have different lengths, which are usually small, amounting on average to 60 nucleotides. The nucleosomal filament is responsible for the higher levels of compaction of chromatin. Ultimately it appears under the electron microscope in the form of the so called 30-nanometer chromatin fibril.

The histone H1 is responsible for the packing of the nucleosomal filament within the chromatin fibril. The central part of its molecule consists of a globule, which interacts with a specific section of the nucleosome. The N- and C-terminal fragments of the protein have extended forms. The N-terminal fragment is associated with the preceding linker DNA in the region of its contact with the nucleosomal particle, while the C-terminal fragment is in contact with the histone core of the next nucleosome. Furthermore, it has been noted[168] that a free H1 molecule interacts more readily with DNA in the vicinity of another histone molecule, which has already been added, than in the region separate from it. This is the reason for the tendency of the H1 proteins to bind to DNA in groups of eight and more molecules. It has been suggested that the interaction of this type, called cooperative binding, constitutes the basis of the activation of genes. The condensation of the H1 molecules induced by such interaction converts certain chromatin sections into a kind of microcrystals, the 'fusion' of which under the influence of an external regulator signal leads to the disruption of the histone cluster and a local rearrangement of chromatin.[162,169]

The elucidation of the internal structural organisation of the histone core and of its interactions with the DNA double helix, began, as already mentioned, in 1974.[163] Soon the structure of the octamer was investigated in solution by a small angle neutron and X-ray scattering methods with a resolution of ~20 Å[170-172] and in the crystalline state by X-ray diffraction[166,173,174] (with resolutions of 22 Å,[175] 7.0 Å,[176,177] and 3.5 Å). According to crystallographic data,[177] the pure histone core of the chick erythrocyte is an octamer having the form of an ellipsoid with axes 110 and 65–70 Å long. The model proposed by the authors of the above study consisted of three parts: the central tetramer $(H3-H4)_2$ and two dimers (H2A–H2B) disposed on its opposite sides. However, both the form and the size of the octamer differed significantly from the parameters obtained in a series of other investigations.[166,173-175] Repeated diffraction analysis[178,179] established the external dimensions of the histone core, which agreed with the data obtained from other studies (height ~ 55 Å, diameter ~ 70 Å).

According to some workers, the three-dimensional structure of the histone core resembles a flat disc, others see it as a pin or a spool, while according to a third group it is a wedge. Therefore,

only one thing is clear: the structure of the octamer is fairly complex and has no direct analogies. A more detailed discussion of the geometry of the complex $(H2A-H2B-H3-H4)_2$ may be based only on the results of the later studies,[178,179] since they were obtained with a resolution of 3.1 Å, which already permits the identification of the conformational state of the aminoacid residues and even of individual side chains.

In an earlier X-ray diffraction analysis of a whole nucleosome,[166,170,173] it was shown that the right-handed DNA double helix is twisted around the histone core in the form of a left-handed superhelix ~55 Å long and with an internal diameter of ~70 Å. This was confirmed in a study of the structure of an isolated octamer,[178,179] on the surface of which the groove of a protein superhelix was observed. It had an external diameter of ~65 Å, a pitch of ~28 Å, and a left-handed twist, which is fully consistent with the parameters of the DNA superhelix in the nucleosome. Thus, the surfaces of the histone core and of the DNA double helix twisted round it bear the same relation to one another as the surfaces of a screw and a nut. Nevertheless, the formation of a nucleosome can hardly be regarded as the screwing of the protein octamer into a strictly defined nucleotide structure. This is prevented firstly, by the presence on the helical groove of the nucleosome core of alternating hillocks and hollows, and secondly, by the fact that the superhelix of the DNA molecule in a free state is not a stable form, although it is apparently a low-energy conformation. It is more likely that the nucleosome is formed by the screwing of the labile DNA double helix onto the histone core, in a manner resembling the winding of a thread on a spool. The mechanism of the opposite process involving the liberation of the double-stranded DNA from the protein complex is the same. This model is supported by the finding that the hillocks and hollows on the inner central part of the nucleosome are complementary in relation to the profile of the contact surface of the DNA superhelix formed by the B-form of its double helix.

In connection with the determination of the three-dimensional structure of the histone octamer $(H2A-H2B-H3-H4)_2$ and of its steric interrelations with DNA, a whole series of fundamental problems have arisen. For example, the question arises: what is the mechanism and the causes of the spontaneous appearance of the protein complex and the self-assembly of the nucleosome as a whole. No less interesting is the question how the nucleotide chain is liberated from the histone core. The point is that the accessibility of the DNA forming part of the composition of nucleosomes is significantly restricted on those sections where the double helix is in contact with the surface of the octamer. The attachment of specific regulator proteins to the functionally active nucleotide sequence becomes possible only when the corresponding DNA binding sections are liberated from the nucleosomes. The elucidation of the cause of the decomposition of the nucleoprotein complexes is therefore just as important as the study of the cause of their appearance. One may assume that, after the mechanisms of the creation and decomposition of nucleosomes has been subjected to a quantitative treatment, one of the most intriguing questions concerning histone proteins will also be solved, namely why the histones H2A, H2B, H3, and H4 are some of the most conservative proteins in nature as regards their aminoacid sequences (Table 7). One cannot rule out the possibility that the nucleosome is a unique cell subunit in terms of its structural organisation. On the grounds of general considerations, one may postulate that it should combine an ideal consistency between the intramolecular and intermolecular interactions of the proteins constituting the histone octamer, complementary surfaces of the nucleosomal core, and the contact surface of the DNA superhelix, together with a fine balance between forces acting from opposite directions, the disturbance of which, combined with the corresponding changes in external forces, leads to the rapid shift of the equilibrium towards the appearance or decomposition of the nucleoprotein complex. The conservative nature of the histones H2A, H2B, H3, and H4 shows that the normal functioning of such a system virtually rules out changes in aminoacids.

2. The structure of the N-terminal fragment of DNA topoisomerase I

In the replication of DNA, the rate of polymerisation in bacteria and mammals varies from 500 to 50 nucleotides per second. For each ten base pairs of the newly synthesised chain, the parent DNA double helix must execute one complete turn around its axis in order to avoid being tangled up (Fig. 14). Hence, in order to achieve an obstacle-free movement of the replication fork at the experimentally observed rate, the entire chromosome, ahead of the fork, must execute 50 ± 5 turns per second. Exactly the same 'twisting problem' evidently arises in the transcription and recombination of DNA. In all cases, it is solved with the aid of special enzymes, DNA topoisomerases, which form a kind of 'hinge' within the helices.

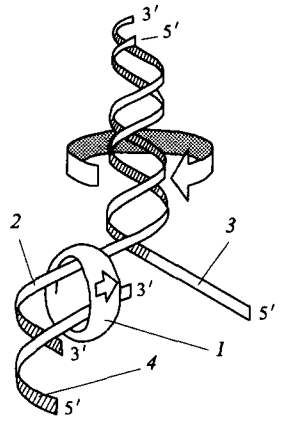


Figure 14. Rotation of the DNA double helix during its replication; (*1*) DNA polymerase on the leading chain; (*2*) matrix for the synthesis of the leading chain; (*3*) matrix for the synthesis of the lagging chain; (*4*) newly synthesised DNA chain.

DNA topoisomerase I cleaves the phosphodiester bond in one of the DNA chains and becomes covalently attached, as a rule via a tyrosine side chain, to the free 5'-end formed. Chain termination enables the two halves of the double helix to rotate freely around the phosphodiester bond located opposite the break and hence to release the accumulated stress. The protein–DNA covalent bond is of the high-energy type and the enzymic hydrolysis of the phosphorus diester bond is therefore a reversible reaction and does not require the involvement of ATP. After the removal of stress, the broken ends of the DNA chain are rejoined. Thus, in terms of its function, the DNA topoisomerase constitutes something in the nature of a 'reversible nuclease' or a 'reversible DNA polymerase'. DNA topoisomerase was first isolated from the cells of *E. coli* in 1971.[180] It still remains the most widely investigated enzyme of the subfamily of type I DNA topoisomerases, which includes the topoisomerases III of the bacteria *E. coli*[181] and *Saccharomyces cerevisiae*[182,183] as well as the reversible gyrase of the archebacterium *Sulfolobus acidocaldarius*,[184,185] which were discovered many years later. Apart from the reaction described, enzymes of this type catalyse the relaxation of an incorrectly twisted linear DNA superhelix, the disentanglement and ring closure of a single-stranded DNA chain, the pairing of complementary rings, and a series of other related operations. A common feature of the type I DNA topoisomerases is the mechanism of the catalytic step, which proceeds without the employment of the energy of ATP and invariably includes the stage in which only one chain of the DNA double helix is temporarily broken. Enzymes with similar

functions, which, however, cleave not one but two phosphodiester bonds and are bound covalently to both chains of the double helix, constitute the subfamily of the type II DNA topoisomerases.[186]

The results of the X-ray diffraction analysis of the three-dimensional structure of the N-terminal fragment of the DNA topoisomerase I of *E. coli* were published in 1994.[187] The three-dimensional structure of the fragment of the polypeptide chain comprising 553 aminoacid residues and having a molecular mass of 67 kDa (that of the whole protein is 97 kDa) was determined with a resolution of 2.2 Å following the interpretation of the diffraction patterns obtained at a temperature of −168 °C using synchroton radiation. This was the first crystallographic study of a topoisomerase.

DNA topoisomerase I is a metalloenzyme, the C-terminal fragment of which contains three zinc(II) atoms and three sections with two disulfide bonds.[188] The catalytic reaction involving the cleavage of the DNA nucleotide chain includes transesterification between the phosphodiester bond of the substrate and the Tyr-319 side chain of the enzyme.[189] The C-terminal fragment of the protein with disulfide bonds does not participate in this reaction. The zinc atoms also do not affect the phosphodiester bond cleavage stage, but are necessary for the reconnection of the broken chain after the relaxation of DNA.[190] The fragment of DNA topoisomerase I with a molecular mass of 67 kDa, investigated by Lima et al.,[187] can therefore catalyse the cleavage of a single nucleotide chain, but is unable to restore the integrity of the DNA superhelix. Its substrate specificity and kinetic characteristics are the same as those of the native protein or the zinc-free protein.[190,191]

The three-dimensional structure of the N-terminal fragment consists of an irregular extended torus with external dimensions of ~67 × 43 × 86 Å. It is made up of domains I–IV, which are separated from one another and are easily observable. Domain III includes the catalytically important Tyr-319 residue (Fig. 15, Scheme 5).

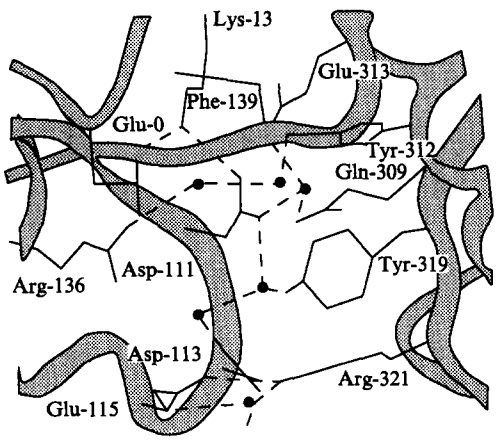
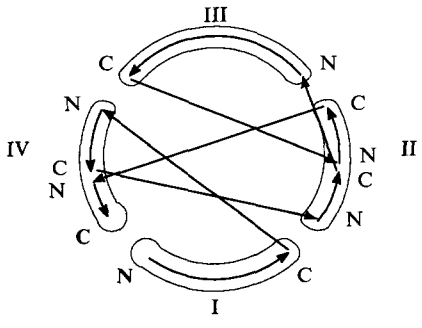


Figure 15. The active centre of DNA topoisomerase formed by domains I and II. The hydrogen bonds are designated by dashed lines.[187]

Scheme 5



The aperture in the torus, free from the side chains of the residues, has a diameter of ~28 Å, so that the insertion of the DNA double helix in the B-form into it should not be accompanied by steric hindrance. The arrangement of the effectively interacting domains I, III, and IV led to the formation of a large substrate-binding groove and of yet another aperture in the immediate vicinity of the principal aperture in the torus and the active centre with the Tyr-319 aminoacid residue. The second aperture has a small diameter but it is quite sufficient for the passage of a single DNA chain.

The view that the appearance of new complex proteins in the course of evolution takes place as a result of the merging of preexisting proteins has come to be generally accepted.[192,193] The three-dimensional structure of the multidomain protein is thus represented as an ensemble of largely independent globular domains which had previously existed as individual proteins, and the aminoacid sequence is represented by a set of polypeptide chains of the domains in which the C-terminus of the preceding domain is linked to the N-terminus of the next domain. The structural organisation of DNA topoisomerase is in no way consistent with this method of the formation of complex proteins, and its polypeptide chain cannot be represented by the sequential and independent formation of domains.

Assembly of the aminoacid sequence of the 67 kDa fragment apparently begins from the N-terminal domain I, which consists of a continuous polypeptide section (Scheme 5). Then, instead of proceeding immediately to domain II, the chain extends to domain IV and only then proceeds to domain II, having, however, completed the formation of domain IV only to the extent of one third. Next, having participated in this instance in the construction of only half of domain II, the chain passes to domain III. After completing its organisation, it returns again initially to domain II and then to IV, completing their construction. In the aminoacid sequence described, all the interactions of domain III with domains I and IV are restricted to nonvalence contacts. The covalently linked domains II and III, the polypeptide chain of which is not intertwined with the chains of other domains, can migrate as a solid body relative to the remainder of the protein. Such structural organisation of the enzyme molecule with two forms of the torus (open and closed) is, according to the authors,[187] of decisive importance for the functioning of the DNA topoisomerase I, since the migration of domains II and III makes the active centre accessible to both single-stranded and double-stranded DNA.

The hypothesis that appreciable conformational changes in the molecule of the DNA topoisomerase in the course of the formation of the nonvalence Michaelis complex are necessary is confirmed also by a number of other no less convincing structural features of the enzyme. One of them concerns the catalytically most important Tyr-319 residue. It belongs to domain III and is located next to the substrate-binding chain. In the native enzyme, Tyr-319 is accommodated in the narrow gap between domains I and III and is shielded from contact with the external medium by the Arg-321 protein chain, forming salt bridges simultaneously with the Asp-113 and Glu-115 side chains. The transition from the closed form of the molecule to the open form not only permits the penetration by DNA into the inner part of the torus but also moves the Tyr-319 residue onto the surface, providing conditions for the approach of the latter to the nucleotide chain of the substrate and attack on the phosphodiester bond. The migration of domains II and III as a single whole is stimulated by the nonvalence protein–DNA interactions, because a cofactor such as ATP is absent in the given instance. This feature distinguishes DNA topoisomerase I from DNA topoisomerase II, the catalytic action of which also involves conformational changes, but these require the participation of ATP.

In a study of the three-dimensional structure of the N-terminal fragment of DNA topoisomerase I,[187] the authors devoted much attention to the description of many details of the

mechanism of the catalytic step. However, these data, like the results of other similar investigations, are phenomenological in character.[194] The importance of X-ray diffraction data for enzymes clearly does not lie merely in making possible the adjustment of the existing theoretical mechanisms of catalytic action for consistency with them or in inventing new ones. These data are valuable by virtue of the potential possibility of their being used as a unique experimental basis for an a priori approach to a quantitative description of the catalytic step as a spontaneous process.[195]

3. The structure of β-galactosidase

The following two factors led me to include, in the section devoted to DNA-binding proteins, an enzyme which strictly speaking does not belong to this group: first, the structural organisation of the β-galactosidase molecule, which it had proved interesting to compare with the organisation of DNA topoisomerase just examined above; and second, the involvement of the enzyme in the regulation of the transcription activity of DNA, manifested by its influence on the effectiveness of the reading process in the course of the transcription of a whole cluster of genes (including its own gene) responsible for the metabolism of lactose. β-Galactosidase has played an historical role in the development of molecular biology, specifically in the concept of the general molecular principles of the regulation of the action of genes. It is in fact the study of the biosynthesis of this enzyme at the transcription level in bacteria which enabled L'vov, Jacobi, and Monod at the beginning of the 1960s to develop a theory of the transfer of genetic information and of the regulation of the protein synthesis (the operon concept). According to this concept, each operon incorporates a regulator gene, a promoter, an operator, and the structural genes of specific proteins (cystrones), in which the corresponding iRNA (polycystrone RNA) is transcribed and for the operation of which the three preceding regulator sections of the DNA double helix are responsible.

The first studies on the mechanism of genetic control were devoted to the synthesis of β-galactosidase, which hydrolyses lactose (a disaccharide) to glucose and galactose (monosaccharides) in the *E. coli* cell. The experiments led to the discovery of the repressor protein for the lactose operon initiating the transcription of structural genes (in the given instance the gene of β-galactosidase, as well as permease and galactoside transacetylase). This is achieved by binding the repressor to the operator section of DNA with a length equivalent to 21 nucleotides, which overlaps the aminoacid sequence of the promoter. This blocks the access by RNA polymerase to its binding section and the transcription of cystrones becomes impossible. The induction and repression of the synthesis of the protein, i.e. the changes in the rate of the process in opposite directions, require the presence in the model of the regulator mechanism of yet another inductor element, which should, on the one hand, control the action of the lactose operon repressor protein and, on the other, should be directly or indirectly related to the function of the enzyme being synthesised. Such as inductor has indeed been discovered and it proved to be the substrate of β-galactosidase, namely lactose or, more precisely, allolactose.

β-Galactosidase is an inducible enzyme, the rate of synthesis of which depends on the composition of the medium and the conditions in the latter, in particular on the lactose concentration. The amount of enzyme increases rapidly with increase in the substrate concentration, so that the relation between the inductor and β-galactosidase is evident. The effect of lactose, via allolactose, on the repressor is allosteric in character. When the intracellular lactose content begins to exceed the norm, the concentration of allolactose increases, which induces conformational changes in the repressor molecule. Its linkage with DNA is weakened and it liberates the promoter and hence enables RNA polymerase to transcribe the β-galactosidase genes. Depression of the gene takes place, iRNA is formed, and the enzyme, which is necessary only when the cell contains lactose, begins to be synthesised.

The three-dimensional structure of the *E. coli* β-galactosidase has been recently investigated with a resolution of 2.5 Å.[196] X-Ray diffraction analysis demonstrated that the aminoacid sequence of the protein comprising 1023 residues has an extended N-terminal 1–51 α-peptide section and five structurally autonomous domains: 51–217 (I), 220–334 (II), 334–627 (III), 627–736 (IV), and 736–1023 (V). In contrast to the type I DNA topoisomerase, where the polypeptide chain passes from one incomplete domain to another and, without completing its construction, is directed to the third, etc. (see Scheme 5), in galactosidase the chain passes from the N-terminus of the molecule to its C-terminus in strict succession through all the domains. The assembly of the long polypeptide chain of the protein is thereby facilitated, because in the modular structural organisation of the native enzyme each domain may act as an independent subunit capable of coiling up.

The quaternary structure of β-galactosidase in the form of a homotetrameric nonvalence complex with a molecular mass of 465 kDa exhibits an enzyme activity. Structural analysis showed that the complex belongs to the 222 symmetry point group. The four tetramers constitute an asymmetric unit cell belonging to the space group $P2_1$ with the parameters a = 107.9 Å, b = 207.5 Å, c = 509.9 Å, and β = 94.7°. At the present time, among all the three-dimensional structures of proteins identified at the atomic level, the molecular structure of β-galactosidase has the longest polypeptide chain. The study by Jacobson et al.[196] demonstrated for the first time the possibility of investigating protein crystals with a unit cell size exceeding 500 Å and a relative molecular mass of ~2000 kDa per asymmetric unit.

It had been shown earlier that the Glu-461, Met-502, Tyr-503, and Glu-537 residues should play an important role either directly in the catalytic reaction itself or in the construction of the active centre.[197,198] Indeed, in the observed three-dimensional structure of the enzyme they proved to be close to one another, being located around a deep well, which apparently represents a substrate binding pocket. The residues constituting it, like the four residues noted above, are highly conservative.

VI. Proteins of skeletal muscles

One of the most surprising properties of proteins, which are not manifested by other compounds, is undoubtedly their capacity for a macroscopic displacement i.e. greatly exceeding the dimensions of their molecules. This is due to the presence in a series of proteins, which have been called allosteric of several energetically preferable and comparatively readily interconvertible conformations which are sensitive to changes in the external medium and interact selectively with other molecules. When energy is consumed, the conformational transitions lose their reversibility and become ordered; the displacement becomes directional and the protein performs useful work.[199] For example, it can 'pass' through a thin filament, a microtube, or a DNA double helix. Such allosteric 'striding' proteins, employing the energy of the hydrolysis of ATP, are DNA helicase participating in the replication of DNA, kinesin, and dynein, which ensure intracellular mobility, and the muscle protein myosin, to the structure and functions of which this section is mainly devoted.

Myosin is a multifunctional protein. In the contraction of skeletal, hard, and smooth muscles, and in intracellular movement, it performs simultaneously at least three key functions: structural, allosteric, and enzymic. The most useful information about the functions of myosin may be obtained by investigating the tranversely striped skeletal muscle, which contracts spontaneously, as well as the analogous tissues of the invertebrates, primarily the flying muscles of insects. Electron microscope study of the longitudinal and transverse thin sections through skeletal muscles, carried out for the first time in 1953 by Huxley,[200]

revealed a high level of their structural organisation. By the following year, Huxley[201] had already proposed the so called sliding filament model, which was of fundamental importance for an understanding of the nature and the molecular mechanism of muscular contractions.

Skeletal muscles are bundles of muscle fibres, the largest repeating structural element of which is the myofibril, a cylindrical filament 1–2 μm (1000–2000 Å) in diameter, which extends from one end of the cell to the other. The myofibril contains in its turn protein filaments of two types — thick and thin. The principal proteins of the thick and thin filaments are myosin and actin filaments in the myofibril are strictly ordered. The functional contractile unit of the myofibril is the sarcomer, which is about 2.5 μm long and is subdivided into discs I and A (Fig. 16). The thick filaments (~1.6 μm long and 0.015 μm thick) extend from one edge of disc A to the other, while the thin filaments (~1.0 μm long and 0.008 μm thick) extend from the *Z* line and enter the gaps between the thick filaments of disc A. In muscle contraction, each sarcomer is shortened in proportion to the entire muscle, only disc I being compressed, while the dense disc A does not change its dimensions. The essential feature of the filament sliding model[201] is that the contraction of the myofibril and hence of the muscle fibre takes place as a result of the sliding of the thick filaments relative to the thin ones, whilst the lengths of both remain the same.

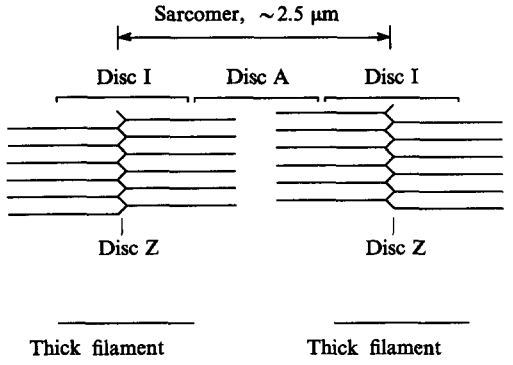


Figure 16. Schematic illustration of the structure of the sarcomer.

The filament sliding model has undergone prolonged experimental tests and has been confirmed most convincingly by data obtained by the direct electron microscope and X-ray diffraction methods. They demonstrated that the shortening of the muscle is indeed unaccompanied by changes in the intrinsic length of the filaments and in the nature of their packing in the sarcomer. The force developed by the muscle proved to be proportional to the degree of mutual overlap of the myosin and actin filaments and is therefore due to their interactions over the entire overlapping section. Following the introduction of high-resolution electron microscopy (second half of the 1960s; resolution 20–40 Å), it became possible to observe a multiplicity of lateral projections, forming ~0.013 μm (~130 Å) transverse bridges between the thick and thin filaments. It became evident that the relative displacement of the filaments is effected with the aid of these bridges. They belong to myosin and operate by employing the energy of the hydrolysis of ATP, like miniature oars. The presence of ATP in muscle fibres has been known since 1929, because it was in fact isolated for the first time from muscles. The fact that myosin catalyses the hydrolysis of ATP, i.e. is an ATPase, was established by Engelhardt and Ljubimova in 1939.[202] This discovery was the first indication of the source of energy for the contraction of muscles and of the role of myosin in the utilisation of the energy.

Towards the beginning of the 1970s, the phenomenological stage in the study of the functioning of muscles was largely complete; the overall mechanism of the process, its principal participants, and the source of energy were discovered.[203–206] Further study consisted in the search for answers to the following questions: what is the molecular mechanism of the contraction of muscles, how is the chemical energy of the hydrolysis of ATP transformed into mechanical work at the molecular level, and how does a nerve pulse trigger the movement of the muscle fibres?

The study of this problem began with experimental structural investigations by electron microscope methods, especially by cryoelectron microscopy, which can lead to a resolution at the atomic level under favourable conditions,[207,208] X-ray diffraction of oriented fibres, and X-ray crystallography. The combined approach, employing the results of both electron microscopy and X-ray diffraction analysis for the construction of models of the three-dimensional structures of the protein molecules consistent with the data obtained by these methods, has been recently (the 1990s) gradually introduced into practical use.[209] This approach has already found an application in studies on muscle proteins and spherical viruses.[210–212]

1. The structure of actin filaments

The thin filaments include as part of their composition actin, which, as already stated, constitutes the basic component of the filaments, tropomyosin, and troponin. The standard procedure for the isolation of actin involves the extraction of the dried and ground muscle tissue with a dilute salt solution. Such treatment cleaves the actin filaments into globular subunits, each of which is made up of a single polypeptide chain with a molecular mass of 41.8 kDa; this is globular actin or G-actin. One Ca^{2+} ion is bound to each molecule of G-actin, stabilising its globular conformation. In addition, one ATP molecule is associated with G-actin via a nonvalence interaction. In the nonvalence polymerisation of globular actin, the terminal phosphate of ATP is split off and fibrillar actin or F-actin is formed. The polymerisation may be induced by increasing the salt concentration to nearly the physiological level. The process does not require the consumption of energy, although it is in fact accompanied by the hydrolysis of ATP, which appreciably increases the rate of polymerisation and influences its dynamics. According to data obtained by electron microscopy, the actin filaments consist of two chains of globular molecules, the length of which is ~65 Å and the thickness in the widest part is ~40 Å. The F-actin chains form a double helix consisting of 13 molecules in six turns repeated at 360 Å intervals.[213]

After attempts in the course of many years, the G-actin molecules were crystallised but only together with the molecules of deoxyribonuclease I in proportions of 1:1. The three-dimensional structure of monomeric actin at the atomic level was therefore discovered from X-ray crystallographic data for the G-actin–deoxyribonuclease I complex.[214] The X-ray scattering pattern for the F-actin fibre was obtained simultaneously with a resolution of 6 Å and the orientation of the G-actin monomer and of the double helix of the polymer was found by comparing the calculated and observed X-ray diffraction patterns.[213] The remaining uneliminated differences reflect the fact that the polymerisation of actin is accompanied by slight conformational changes, but only to some extent, because the use of a refined structural model of G-actin in the construction of a model of the F-actin filament led recently to a better agreement between the calculated and experimental results.[215] Further refinement of the model is difficult owing to the lack of the X-ray diffraction pattern for F-actin with a resolution better than to within 6 Å. A solution may be found by resorting to a theoretical approach and to the employment of conformational analysis[216] and molecular dynamics[217] methods. The atomic model of F-actin, constructed by adjusting for consistency the results of X-ray

diffraction of G-actin crystals and thin F-actin filaments, agreed with the model reconstructed from cryoelectron photomicrographs of the actin filament.[218,219]

Two auxiliary proteins, tropomyosin and troponin, are closely associated with each actin filament. Their principal function is the regulation of muscle contractions as a function of the concentration of Ca^{2+} ions and hence nerve impulses. The rigid rod-shaped α-helical tropomyosin molecules ~400 Å long are located in longitudinal grooves extending along the actin filament, imparting thereby the necessary rigidity to the F-actin double helix (Fig. 17). The globular form of troponin is made up of three different subunits (troponins T, I, and C). Troponin T has a section for the binding of tropomyosin and is responsible for the attachment of the entire complex to the actin filament. When troponin I is added to troponin T and tropomyosin, a complex preventing the interaction of actin with myosin even in the presence of Ca^{2+} ions is formed. In the presence of troponin C, this complex continues to block the interaction of myosin with actin, but only in the absence of Ca^{2+}. When free calcium appears in the medium, troponin C binds four Ca^{2+} ions and removes the blockade in this state. Precisely this effect actually constitutes the basis of the excitation of the muscle contraction by calcium ions. We shall return to the problem of the way in which tropomyosin and troponin control the interaction of myosin with actin after considering the structural organisation of the thick filaments.

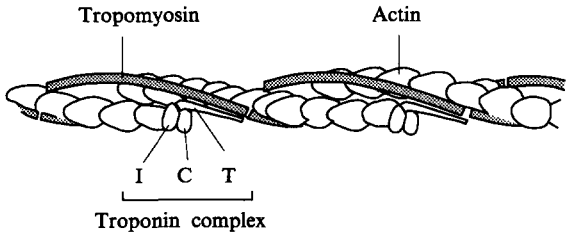


Figure 17. Schematic illustration of the thin filament.

2. The structure of myosin filaments

The contents of myosin, actin, tropomyosin, and troponin in the myofibrils are approximately 55%, 25%, 15%, and 5% respectively. A distinctive feature of the myosin in the skeletal muscles is its ability to form spontaneously, under in vitro conditions, giant polymeric complexes greatly exceeding in size the myosin aggregates from nonmuscle tissues. Myosin can be extracted from skeletal muscles with concentrated salt solutions, in which it is readily soluble. Muscle treated in this way loses only the thick filaments, which decompose into the component myosin molecules having a molecular mass of ~520 kDa. On treatment with a concentrated solution of urea or another detergent, the myosin molecule decomposes into six polypeptide chains: two identical heavy chains with a molecular mass of ~220 kDa and two pairs of light chains with molecular masses of ~22 and

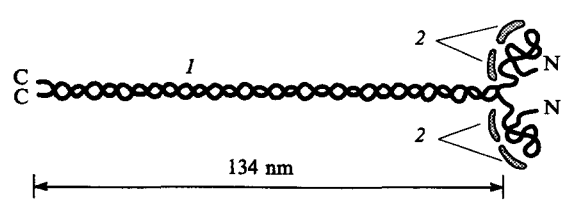


Figure 18. Schematic illustration of the structure of the myosin molecule: (*1*) twisted α-helices; (*2*) light chains.

~15 kDa.[220-224] It was established for the first time in 1963 with the aid of electron microscopy[223] that myosin consists of two globular 'heads', each of which is attached to a heavy chain, containing a long α-helical section. In the native myosin molecule, the α-helices of two heavy chains are twisted round one another into a super helix, forming a rod-shaped tail, from which two heads project. Each head is made up of the globular part of the heavy chain (~95 kDa) and includes one molecule of each of the two types of the light chain (Fig. 18).

The thick filaments are formed as a result of the association of the myosin molecules via their rod-shaped tails: several hundreds of molecules, arranged in parallel with lengthwise displacements, form a bundle from which laterally disposed heads project (Fig. 19). The entire structure is dipolar. The projecting heads are absent only along the small middle section of the filament, where two oppositely directed bundles of myosin tails are joined together. The globular heads of the myosin molecules interact with actin and form transverse bridges between the thick and thin filaments.

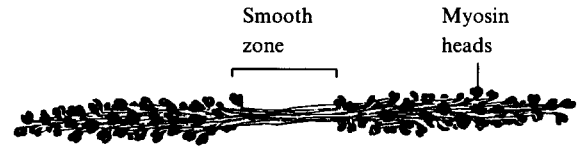


Figure 19. Schematic illustration of the thick filament of the skeletal muscle.

The proteolysis of the heavy filaments showed that the myosin head can be readily separated from the remainder of the protein, forming the so called subfragment-1 (S1) and a superhelical rod, responsible for the self-assembly of myosin and for the creation of the structural base of the thick filaments.[220,225] The subfragment-1 contains the nucleotide-binding active centre for ATP, an actin-binding surface, and two regions corresponding to contacts with the light chains. Thus the globular head retains all its function even outside the filament. The light chains incorporated in it differ from one another in structure and properties. In some forms of myosin, they do not exert an appreciable influence on the ATPase activity of the enzyme.[226] However, in certain types of myosin the light chains of its head may participate in the regulation of the hydrolysis of ATP, but only in the presence of actin.[227,228]

As early as the end of the past century, it was noted that myosin exhibits double refraction. However, until recently, the entire knowledge about the three-dimensional structure was limited to information about the external outline of the molecule and its size obtained with the aid of electron microscopy. In particular, it was known that the myosin head has a pear-shaped form approximately ~190 Å long and 50 Å wide, while the double helix of the tail section is ~1500 Å long and ~20 Å wide.[228-231] Rayment et al.[232] were able to obtain the crystals needed for analysis by employing the procedure, altogether unusual in protein crystallography, involving *N*-methylation under mild conditions, of the side chains of all the Lys residues of the myosin fragment-1. In order to make sure that the methylation did not lead to a drastic alteration of the conformational and enzyme properties of the protein, the authors subjected lysozyme to a similar chemical modification and did not observe after this procedure significant changes in the three-dimensional structure of the enzyme. In addition, the kinetic properties of the methylated myosin S1 were tested.[233] It retained its catalytic activity, although deviations of the kinetic parameters of the hydrolysis of ATP from the values for the native specimen were in fact observed.

The three-dimensional structure of the myosin subfragment-1 was determined[232,234,235] by X-ray diffraction with a resolution of 2.8 Å. This resulted in the refinement of not only the external dimensions of the myosin head, which confirmed that it lacks any kind of symmetry, is 165 Å long, and has a cross-section of 65 × 40 Å in the thickest part, but also led to the construction of an atomic model comprising 1072 aminoacid residues (out of the total number of 1157). The structure contains many α-helices (~48%), the longest of which extends over a distance of 85 Å from the thickest part in the direction of the C-end of the S1 heavy chain. Most of this helix is surrounded by light chains, ELC (essential light chain) and RLC (regulatory light chain), one of the principal purposes of which is conformational stabilisation of the heavy chain of the myosin head (for example, when RLC is removed, the S1 structure breaks down).[236] In addition, the light chains promote the conformational rearrangements which occur when the nucleotide is bound to and liberated from the thickened part of the head. In order to facilitate the description of the three-dimensional structure of subfragment-1, its heavy chain can be usefully divided into three parts: the N-terminal part incorporating the Asp-4–Glu-204 residues, the central part with Gly-216–Tyr-626, and the C-terminal part with Gln-647–Lys-843. Loops lacking secondary structures are located between them. These are subjected to tryptic cleavage, which gives rise to heavy chain fragments with molecular masses of 25, 50, and 20 kDa.

The thickened part of the myosin head contains nucleotide and active binding centres and thus has all the components necessary for the generation of motion. The essential moving, motor part of myosin is a complex of secondary structures disposed around the β-layer consisting of seven largely parallel β-tiers. It includes the components of all three trypsin-cleaved fragments of the heavy chain. The nucleotide-binding packet is located at the end of the central β-pleated sheet and adjoins the 25 kDa fragment on one side and the 50 kDa fragment on the other. The location of the aminoacid residues forming the nucleotide active centre of myosin has been determined by the tracer atom method[237] and by analogy with the sequences of the phosphate loops in the adenylate kinase and ribonuclease structures.[238,239] The binding pocket is ~13 Å deep and approximately just as wide. Since the crystals were grown in the absence of ATP, the active centre corresponds to an open conformation in the observed structure of the myosin subfragment-1.

The interrelation between the hydrolysis of ATP and muscle contraction, discovered, as already mentioned, as early as 1939, was considered in a large series of kinetic studies. The results, obtained for the first time in 1971[240] and in subsequent years,[234,241–243] have not undergone fundamental changes and lead to the same idea about the actin-myosin mechanism of muscle contraction. It is illustrated in Fig. 20 and corresponds to a four-stage process involving the conversion of the energy of the hydrolysis of ATP into mechanical work. In the first stage (*a*) of the proposed mechanism, the myosin head, carrying the ATP-hydrolysis products (ADP and phosphate), is attached to the actin filament. The second stage (*b*) (the 'working step') begins with a change in the conformation of the head induced by its interaction with actin, which liberates the hydrolysis products and results in the appearance of stress in the actomyosin complex, which is released during the longitudinal displacement of the thin filament. After the completion of the motion, conditions arise for the addition of a new ATP molecule to the myosin head, which is accompanied by the breakdown of the actin–myosin contacts (stage *c*). The cycle terminates with the hydrolysis of ATP and a change in the conformation of the head, which assumes the initial position (stage *d*). It has been suggested that each head of the myosin molecule executes its own cycle independently.

Thus the most significant features of the mechanism of the action of the actomyosin complex are that, firstly, the hydrolysis of ATP is completed at the time when myosin does not interact with actin, and, secondly, the driving force of the process arises on liberation of the ADP and γ-phosphate produced.

Not only the results of kinetic studies but also a multiplicity of experimental data obtained by other methods agree with the idea described above that the myosin heads 'stride' along the actin filaments. In particular, the generally accepted idea is confirmed by the three-dimensional structures of monomeric actin[213,214] and the myosin head,[232,234] which were discovered recently, as well as the structural model of the actomyosin complex constructed, with a resolution of 5–8 Å on the basis of these structures using cryoelectron microscope data.[235] One of the key features of this complex, which became evident after the construction of the model, is the existence of nonvalence contacts between actin and the C-terminal domain of the 50 kDa tryptic fragment. Interactions of this type may stimulate conformational changes in myosin at the instant of its mutual approach with actin in the second stage of the muscle contraction process (Fig. 20). In a study of skeletal muscles by the small angle X-ray scattering method, data were obtained recently, indicating convincingly a significant change in the form of the myosin head during the binding of nucleotides.[244]

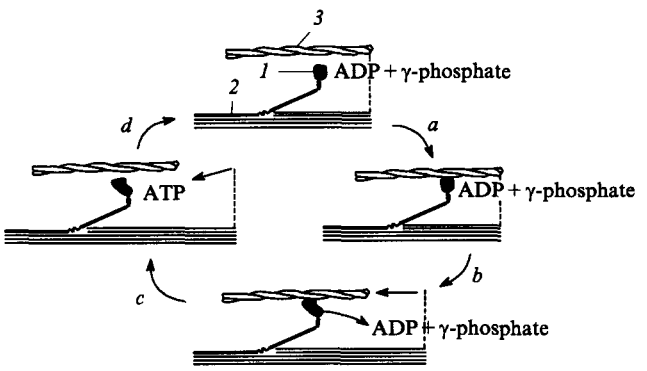


Figure 20. Schematic illustration of the mechanism of the utilisation of the energy of the hydrolysis of ATP for the movement of the actin filament relative to the myosin filament: (*1*) myosin head; (*2*) thick filament; (*3*) thin filament.

The conformational rearrangement of myosin is accompanied by an allosteric effect — the closure of the long narrow groove at the boundary between the thick and thin halves of the pear-shaped head. This results in the mutual approach of actin and the nucleotide centre of myosin and in the liberation of ADP and inorganic phosphate. This is the reason why myosin is an actin-dependent ADPase. It was noted long ago that the rate-limiting stage of the catalytic reaction is not the binding of the ATP molecule to the active centre and not the hydrolysis proper, but the liberation of the products of the enzymic reaction, remaining firmly bound to myosin, which in fact prevents the onset of the next catalytic step.[203,204] The ATPase activity of the purified myosin is low: the enzyme requires approximately 30 s for the hydrolysis of one ATP molecule. In the presence of actin, each myosin molecule is capable of hydrolysing from 5 to 10 ATP molecules per second.

The thick filament contains ~500 myosin heads and they all operate simultaneously but not synchronously, drawing in the actin filament. In the new cycle of the interaction with actin, the head is attached to another monomer of the thin filament located closer to the *Z*-plate. The limiting contraction of the muscle develops in hundredths of a second (~0.02 s). The muscle at rest is elastic and can be easily stretched. The contracted muscle, on the other hand, is inelastic and rigid; its extension is prevented by the bonds between the actin and myosin filaments, which persist when the ATP concentration is inadequate.

Skeletal muscle contracts when the action potential of the nerve fibre is switched on and when a signal is obtained from the corresponding motoneuron. The electrical excitation spreads rapidly over the so called T-tubes, which surround each myofibril in the region of the *Z*-plates. Here the electric signal is transmitted to the sarcoplasmic reticulum—a system of branched intercommunicating vesicles. At the instant of activation by the electric pulse which has passed through the T-tubes, the sarcoplasmic reticulum injects into the cytosol large amounts of Ca^{2+} ions stored in the vesicles, which initiates the contraction of the myofibrils. The signal from the muscle cell membrane reaches the sarcomers in several milliseconds via the T-tubes and the sarcoplasmic reticulum, so that all the myofibrils of the muscle cell contact almost simultaneously. The relation between the muscle contraction and the changes in the Ca^{2+} concentration is due to the functions of the auxiliary proteins tropomyosin and troponin, which are associated with the actin filaments (Fig. 17). They participate in the regulation of the muscle contraction of the Ca^{2+} ions and thereby render the ATPase activity of myosin sensitive to the concentration of these ions.

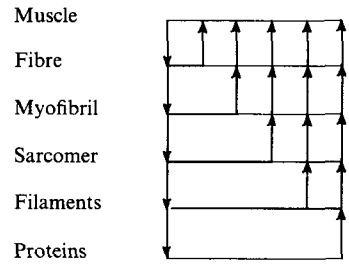
It has been suggested that tropomyosin and troponin effect the Ca^{2+}-dependent regulation of the muscle contraction via steric hindrance to the mutual contacts between actin and myosin. With increase in the concentration of Ca^{2+} ion, the positions of the tropomyosin molecules change, possibly as a result of the change in the shape of the troponin molecules bound to them, in the first place troponin C molecules. In the absence of Ca^{2+}, tropomyosin, troponin T, and troponin I form a complex, which prevents the interaction of actin with myosin. After the appearance of calcium ions, troponin C binds four Ca^{2+} ions and in this state removes the blockade and enables the myosin heads to interact with actin.[245]

The determination of the three-dimensional atomic structures of the proteins G-actin and the myosin subfragment-1, and the reproduction on this basis, resorting to electron microscopy and small angle X-ray scattering, of the three-dimensional structures of the thin and thick filaments, implies the virtually complete visualisation of the elementary constituent sarcomer, capable of demonstrating the principles of the contraction of skeletal muscles and a concrete conformational-electronic mechanism of the transformation of chemical energy into mechanical work. The development of the atomic-molecular model of the actomyosin complex has been one of the outstanding achievements in molecular biology during recent years both in methodological and in purely scientific respects. Without dwelling in detail on the possibilities for the experimental study of the muscle contraction process which have greatly expanded in recent years,[246] we shall only observe that the most notable event in this instance has been the application of the molecular genetics[247] and 'laser trap'[248] methods, which effectively supplement X-ray diffraction and electron microscopy. The 'laser trap' method makes it possible to follow the 100–200 Å displacements of single proteins and to measure the strength of their interactions up to 7 pN (7×10^{-7} dyn). We shall consider below the further improvement of the model itself—the problem of the transition from the phenomenological description of the mechanism of action of the molecular muscle motor to the creation of a rigorous theory of the contraction of muscles and of a quantitative method for the analysis of this process.

The model of Rayment and Holden does not merely constitute a general description of the authors' and coworkers' own work,[232,234,235] but constitutes a logical completion of a series of numerous investigations during the last four decades and in the first place the studies during the 1990s on G- and F-actins by X-ray diffraction and cryoelectron microscopy.[213,214,218,246]

The relation between the knowledge of the structure of the muscle system and the understanding of the mechanism of its action, i.e. between the morphology of the various levels of the structural organisation and the physiology of the muscle, is illustrated in Scheme 6. The thick arrows indicate the direction of the successive steps in the discovery of the structure and the oppositely oriented thin arrows denote the steps in learning about the function of the biosystem. It is seen from the scheme presented that the more complete and detailed the information about the form and composition of the simplest subsystems, the deeper the understanding of the mechanism of the functioning of the entire system. This nature of the dependence of the depth of understanding of the essential features of the physiological phenomenon and the knowledge of the morphology of its participants is a common feature of the processes involved in learning about biological systems. The treatment of the properties of a biological system is always relative and is based on information about the structures of the underlying subsystems.

Scheme 6



The existence of common characteristics and of a single fundamental plan in the development of research into biosystems leads, despite their perhaps self-evident nature, to nontrivial conclusions which are of special interest at the present time. The development of biology throughout history has been determined by two countercurrent streams of investigations: morphological, directed from objects and phenomena with a complex organisation to simpler ones, and physiological, with the opposite direction. The first and second streams were the driving and driven components respectively. However, at the beginning of the second half of the 20th century, in its progress from complex to simple constructions based on the advances in morphology, biology reached the most universal and simplest level in biosystems—a kind of molecular 'bottom'. Further development on the same lines (from morphology to physiology) and in the same direction (from complex to simple constructions) has become impossible. The usual source of qualitatively new information has fizzled out, because the morphological data about the possible underlying subsystems (individual aminoacids, atomic groups, and atoms) contain no specific biological features and hence are not directly linked to the function of any of the biosystems considered.

Thus the approach employed in biology is completed by the determination with the aid of X-ray diffraction, electron microscopy, or another experimental method of the three-dimensional structures of proteins. The preexisting or newly formulated more detailed and better founded ideas about the mechanism of the functioning of the biosystem are confirmed, an atomic-molecular model of the process is developed, and the corresponding hypothesis is formulated on the basis of experimental data on the structure of the biosystem at the molecular level. All that was done previously as regards biosystems with a complex organisation is in essence done in securing new information about the morphology of the underlying subsystems.

The importance of the role of X-ray diffraction analysis of proteins consists in the fact that the study of biosystems at the molecular level was initiated with its aid, and without which it is difficult to imagine the establishment of molecular biology.

The question arises: what is required for further progress in biology and for its transition to the next improved qualitative state. Since the paths leading to the development of studies on all biosystems are in principle the same, the answer to the question posed above does not lose its significance and generality if it is obtained by considering some kind of concrete object, for example, the actomyosin complex. Another question is what additional information (apart from morphological) about the structures of myosin, actin, and certain other proteins is needed in order to not only compare the hypothetical idea about the operation of the actomyosin complex, about the interactions with the ATP molecules supplying it with energy, and about the Ca^{2+} ions regulating its functioning but also to proceed further and attempt to understand why this occurs. This requires data (quantitative data) on the conformational and chemical changes which have taken place, on their sequence, and on the stimuli and their interdependence. With the three-dimensional structure of the actomyosin complex already known, the quantitative description of the mechanism of its action as a spontaneous periodic process may serve as a criterion of the reliability of our knowledge. Here the main problem is not so much the amount and even the quality of the experimental data as its physicochemical interpretation, i.e. the availability of the corresponding theories which provide a logical and quantitative link between the aminoacid sequence, the three-dimensional structure, and the function of the protein molecule.

VII. Synchrotron radiation

The application of synchrotron radiation, which is much more powerful than the radiation from the traditional X-ray tubes, in protein crystallography may be regarded as one of the most significant methodological advances in X-ray diffraction structural analysis during the last decade. In the 1960s–1970s, the majority of the three-dimensional structures of proteins were solved with the aid of sealed sharp-focus tubes with a copper anode giving rise to a characteristic wavelength $\lambda(K_\alpha)$ of approximately 1.54 Å. Their intensity, estimated from the stream of photons in a collimated X-ray beam (10^9–10^{10} photons s^{-1}), is several times higher than for the usual tubes and the focussing of the beam onto a much smaller area makes it possible to obtain diffraction patterns with a higher resolution and a shorter exposure.

In 1959, Parratt proposed that a synchrotron be used as a source of X-rays in crystallography.[249] Moving in a circle within the synchrotron at relativistic velocities in a homogeneous magnetic field, charged particles begin to emit, following retardation and curvature of their trajectories, a continuous spectrum of electromagnetic waves in the form of a very sharp conical beam with a divergence angle of $\sim 1'$ and an axis which coincides with the instantaneous (tangential) direction of motion. The short-wavelength part of the synchrotron (magnetic retardation) radiation spectrum has a broad maximum in the region of interest for X-ray diffraction analysis. In contrast to traditional sources, the polychromatic radiation of the synchrotron may be rendered monochromatic over a much wider range of wavelengths and hence can give rise to a powerful X-ray flux. Thus, in the synchrotron in Daresbury (Great Britain), the useful flux $N(\lambda)$ is at least 10^2–10^3 times more intense than in the best current X-ray tubes with a rotating anode. Parratt's idea became the subject of serious discussions, but remained unrealised for a long time. Both the possibility of the actual increase in the intensity of the diffraction beams and the ability of protein crystals to withstand an intense radiation gave rise to doubts.[250] It was not until 1971 that synchrotron radiation was applied in diffraction studies on biological objects. Studies on muscle fibres by the small angle scattering method showed that, after being rendered monochromatic and after focussing, the intensity of the synchrotron radiation employed is two orders of magnitude higher than the maximum intensity of the collimated filtered CuK_α radiation from an X-ray tube and does not cause the disruption of the irradiated object.[251–253]

The question of the possible application of synchrotron radiation in the X-ray diffraction analysis of protein single crystals was first examined theoretically by Wykoff[254] and Harrison[255] and experimentally by Phillips et al.[256] The last group of workers obtained X-ray diffraction patterns for protein single crystals comprising those of rubredoxin, azurin, the nerve cell growth factor, and L-glutamine-aspariginase, using an electron–positron synchrotron (Stanford, USA).

The authors showed that, when the radiation from an X-ray tube is replaced by synchrotron radiation (keeping all other experimental conditions the same), the intensity of the scattering by the same protein crystals increases by a factor greater than 60 and hence the exposure time diminishes by the same factor, but the risk of the rapid decomposition of the crystals was not eliminated. On the other hand, owing to the decrease in exposure time, the rate of decomposition of the test object proved to be significantly lower than in the case where the radiation from an X-ray tube is used. Together with the above advantages, the authors also noted the small angular divergence of the synchrotron beam and the ease of tuning the synchrotron to a chosen wavelength, which facilitates the selective optimisation of the anomalous contributions to the scattering factor. More than one third of the studies on X-ray crystallography of proteins are nowadays carried out with the aid of synchrotron radiation and multichannel recording systems.

A no less important factor in the X-ray diffraction of proteins (as is incidentally the case in the entire modern X-ray diffraction analysis) was the appearance of the so called coordinate (multichannel) X-ray detectors (or 'area' detectors), which make it possible to record simultaneously all the diffraction maxima arising from the crystal in the given orientation. Such detectors, which are as a rule complex electronic systems with simultaneous recording of a multiplicity of reflections, are the modern analogues of photographic films used previously at the dawn of X-ray diffraction analysis. The recording of a single diffraction pattern (with a large number of diffraction spots) for a specified orientation of the crystal may last several seconds or less, which, together with a powerful radiation source (synchrotron), diminishes the duration of the X-ray diffraction experiment by several orders of magnitude.

Table 8 presents certain results of the X-ray diffraction analysis of proteins, the diffraction patterns of which were obtained both with the aid of synchrotron radiation (Nos. 1–8 and 17–20) and the radiation from X-ray tubes with rotating copper anodes (Nos. 9–16). The table reflects only a small proportion of the publications from the end of 1993 until the first half of 1994 devoted to the solution of the three-dimensional structures of protein molecules and their complexes. If the advantages of the two methods are judged on the basis of the resolutions and the divergence factors (R), then one must conclude that the nature of the radiation source hardly affects the final result. With the aid of synchrotron radiation and an X-ray tube, it has been possible to elucidate the three-dimensional structures of protein molecules with a record resolution, having overcome in both cases the one angstrom barrier [Table 8, bacterial trypsin (No. 2) and crambin (No. 13)] and having attained at the same time a divergence factor R of about 10%—a value close to that of the factor R for the refined structures of small organic molecules.

The methods also do not differ appreciably as regards the size of the irradiated crystal and the diameter of the focussed spot. Thus, in the determination of the three-dimensional structure of the antibody Fab 17/9 (Table 8, No. 16) by means of an X-ray tube with a rotating anode, the size of the specimen was 0.30 × 0.04 × 0.02 mm and the size of the focussed spot was only 0.1 mm. These parameters are not inferior to the experimental conditions when a synchrotron is employed.

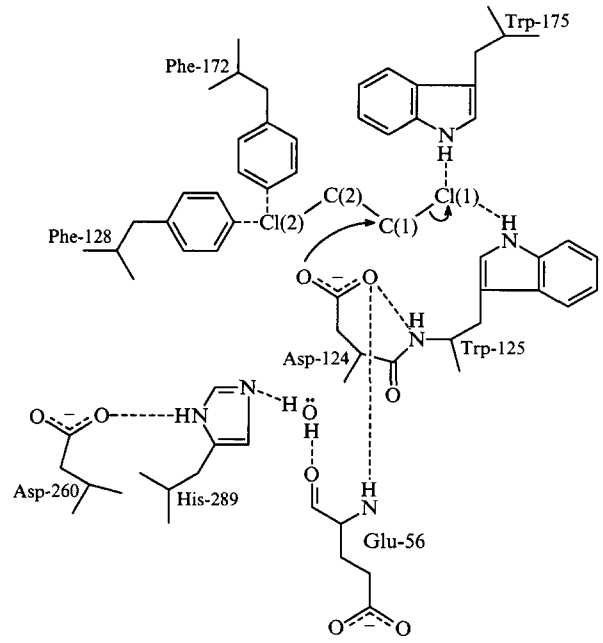
However, one can assume that the powerful synchrotron radiation and its sharp focus are somewhat preferable, since it is then possible to use smaller crystals. Nowadays the problem of the crystallisation of proteins is in both cases just as acute as decades ago. As already mentioned, the situation is particularly complex in the crystallisation of membrane proteins (for a description of the efforts made in this connection and of the relative successes, see Refs 233 and 236–238).

Comparison of the data presented in Table 8 leads to the conclusion that the X-ray diffraction methods employing synchrotron radiation and the radiation from an X-ray tube have no appreciable advantage relative to one another in the study of the native conformations of protein molecules and their stable complexes, i.e. in all those cases where the static problem of the determination of atomic three-dimensional structures is being solved. Until recently, it has been assumed that, in principle, X-ray diffraction analysis cannot be applied to the consideration of problems involving the dynamics of proteins. The use of synchrotron radiation and of modern systems for the simultaneous rapid measurement of the intensities of multiple reflections made it necessary to question the incontrovertible nature of this view if not demand its complete reversal. As can be seen from Table 8, it is possible to shorten the exposure time from many tens of hours (No. 9), days (No. 16), and even weeks (No. 11) to fractions of a second (Nos 3 and 5). According to Thomas,[277] the diffraction reflections from protein crystals exposed to synchrotron radiation may be measured over a period of 100 ps (i.e. 10^{-10} s). The appreciable shortening of the exposure not only makes it possible to shorten significantly the duration of the analysis of the three-dimensional protein structure but also leads to a fundamentally new prospect for the study of the mechanisms of the processes occurring in protein crystals by investigating the structures of the intermediates, when their lifetimes are fairly long or can be prolonged by altering the external conditions. The possibilities which have arisen render the X-ray diffraction method still more unique. One may postulate that the employment of synchrotron radiation promotes the direct monitoring of the kinetics of the conformational and chemical changes in the structures of proteins at the atomic level in the course of, for example, enzymic reactions. A study has been carried out (discussed below) in which an attempt has been made to solve this kind of problem.

1. The mechanism of the functioning of haloalkane dehalogenase

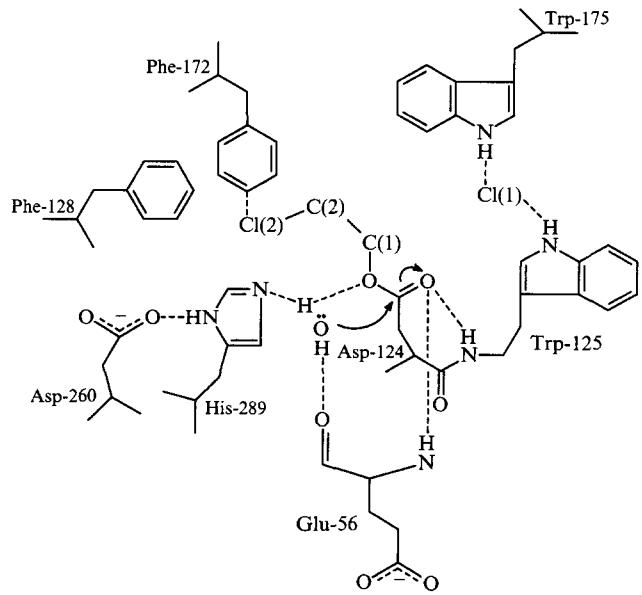
Verschueren and coworkers[175,265] applied X-ray crystallography to the study of the mechanism of the catalytic step involving haloalkane dehalogenase (Table 8, No. 10). The structures of the enzyme, a single crystal of which was kept permanently in the mother liquor, were solved both in the absence and in the presence of the substrate (1,2-dichloroethane) in the solution. At pH 5 and $T = 4$ °C, i.e. far from the optimum conditions (pH 8.2 and 22 °C, the substrate was bound to the active centre, but the catalytic reaction did not develop. The nonvalence enzyme–substrate Michaelis complex remained stable indefinitely and its structure could therefore be determined with the aid of the radiation from an X-ray tube, for an exposure time of 48 h, and with retention of all other conditions in the analysis of the native enzyme. The observed disposition of the substrate in the active centre is illustrated schematically in Scheme 7. One of the chlorine atoms [Cl(1)] in the complex is located at distances of 3.6 and 3.2 Å from the nitrogen atoms of the Trp-125 and Trp-175 side chains and interacts with the hydrogen atoms of two N–H bonds. The other chlorine atom [Cl(2)] is stabilised by dispersion interactions with the Phe-128 and Phe-172 benzene rings. In the observed conformation of the substrate in the active centre, the C(1) carbon atom has been brought closer to the $O^{\delta 1}$ oxygen atom of the Asp-124 side chain (3.8 Å), which is in fact the main reason for the productivity of the nonvalence complex.

Scheme 7



When a crystalline specimen is heated to room temperature, the C(1)–C(1) bond dissociates, which is accompanied by the formation of the $Cl(1)^-$ ion and the alkylation of the $O^{\delta 1}$ oxygen atom of the Asp-124 residue. Such a change in the complex is in fact indicated by the solved three-dimensional structure, which arises and remains stable at pH 5 and 22 °C (Scheme 8). Finally, at pH 6.2 and 22 °C, the alkylated fragment is hydrolysed by a water molecule activated by the side chains of the His-289 and Asp-260 residues interacting with one another. The reaction products are removed and the enzyme returns to the initial state, which is in fact revealed by the X-ray diffraction analysis.

Scheme 8



Thus the two-step mechanism of the reaction involving haloalkane dehalogenase, proposed previously,[278] has been confirmed experimentally.[265] In the catalytic reaction involving this enzyme and occurring under in vivo conditions, the lifetime of the intermediates is too short for crystallographic study. By varying the pH and the temperature, the authors were able to slow down the process, separate it into two stages, and investigate

Table 8. Certain X-ray diffraction data for proteins obtained with the aid of synchrotron radiation and radiation from X-ray tubes with a rotating anode.

No.	Protein	Source	Molecular mass/kDa	Structural data	Resolution/Å	*R* factor (%)	Remarks	Ref.
1	Haemaglutinin	*Influenza virus*	60	Homotrimeric complexes of glycoprotein with receptor analogues	2.15	21.6	Crystal dimensions $0.70 \times 0.70 \times 0.70$ mm; synchrotron radiation with $\lambda = 0.905$ Å; temperature 108 K	257
2	Bacterial trypsin	*Noncardiopsis*	20	Monomer	0.96	10.7	Homotrimeric complexes of glycoprotein with receptor analogues. Tube, CuK_α	258
3	Cytochrome *c* peroxidase (CCP)	Yeast	34	Haem(Fe)-containing enzyme	2.2	15.8	Crystal dimensions $0.40 \times 0.40 \times 1.0$ mm; polychromatic synchrotron radiation (Laue diffraction); single exposure time 300–400 ms; temperature 279 K	259
4	Transketolase	*Saccharomyces cerevisiae*	148	Homodimer	2.0	15.7	Synchrotron radiation	260
5	*p*-Guanidinobenzoyl-substituted trypsin	Bull	24	Monomer	2.0	22.1	Crystal dimensions $1.0 \times 0.40 \times 0.40$ mm; polychromatic synchrotron radiation (Laue diffraction); single exposure time 800 ms; pH 5.5–8.5	261
6	β-Galactosidase	*Escherichia coli*	465	Homotetramer	2.5	19.3	Synchrotron radiation	262
7	Glucose dehydrogenase	*Thermoplasm acidophilum*	160	Homotetramer	2.9	12.4	Crystal dimensions $1.50 \times 0.30 \times 0.20$ mm; tube, CuK_α and synchrotron radiation	263
8	Renin	Mouse	40	Enzyme–inhibitor complex	2.0	18.0	Tube CuK_α (resolution 3.0 Å) and synchrotron radiation resolution 2.0 Å)	264
9	Haloalkane dehydrogenase	*Xanthobacter autotrophicus*	36	Monomer	1.9	16.6	Crystal dimensions $0.40 \times 0.40 \times 0.15$ mm; tube, CuK_α; exposure time 48 h	265
10	Growth hormone	Man	22	Monomer	2.0	18.5	Crystal dimensions $0.25 \times 0.25 \times 0.15$ mm; tube, CuK_α	266
11	Aconitase	Bull, pig	83	Enzyme–*trans*-aconitate and enzyme–nitrocitrate complexes	2.05	17.0	Crystal dimensions $0.70 \times 0.70 \times 0.20$ mm; tube, CuK_α; exposure time 12–18 days	267
12	Pyruvate oxidase	*Lactobacillus plantarum*	270	Homotetramer	2.1	16.2	Crystal dimensions $0.70 \times 0.70 \times 0.70$ mm; tube CuK_α and synchrotron radiation; temperature 277 K	268
13	Crambin	*Crambe abyssinica*	5.5	Monomer	0.83	10.5	Tube, CuK_α; temperature 130 K	269
14	Toxin II	*Androctonus australis*	7	Monomer	1.3	14.8	Tube, CuK_α	270
15	Oncomodulin	Tumour cells	12	Monomer	1.3	16.9	Crystal dimensions $0.20 \times 0.60 \times 0.80$ mm; tube, CuK_α	271
16	Fab 17/9 antibody	IgG2 *a*, *k*	26	Complex between Fab 17/9 and the 101–107 peptide of the influenza virus haemagglutinin	2.8	19.0	Minimum crystal dimensions $0.30 \times 0.04 \times 0.02$ mm; tube, CuK_α; temperature 123 K; exposure time in excess of 10 days	272

Table 8 (continued)

No.	Protein	Source	Molecular mass/kDa	Structural data	Resolution/Å	*R* factor (%)	Remarks	Ref.
17	GH-5 domain of the linker histone H5 (24–96 fragment)	Bird erythrocytes	8.5	Monomer	2.5	25.0	Synchrotron radiation at the three wavelengths 0.9802, 0.9795, and 0.9150 Å corresponding to the Se absorption band. Phase determination by the MAD method. Recombinant [Se-Met]-GH5 sequence produced by *E. coli* plasmids	273
18	BamH1 endonuclease (native and Se-Met)	*Bacillus amiloliqueta-ciens*	25	Homodimer	1.95	19.0	Tube CuK_α and synchrotron radiation at the wavelengths corresponding to the Se absorption edge (0.979 Å), the absorption peak (0.978 Å), and at a point remote from the absorption region (0.900 Å); temperature 278 K	274
19	Gonadotropin (native and Se-Met)	Human embryo	28	Heterodimer (αβ)	2.6	19.9	Synchrotron radiation; anomalous diffraction at four Se wavelengths (MAD); temperatures 277, 293, and 110 K	275
20	Glutamine-5-phosphoribosyl-1-pyrophosphate (PRPP)-amidotransferase	*Bacillus subtilis*	210	Homotetramer	3.0	22.4	Synchrotron radiation; anomalous diffraction at the three wavelengths corresponding to the Fe atoms in the four [4Fe–4S]-clusters of the tetramer (MAD)	276

the atomic three-dimensional structures of the corresponding enzyme–substrate complexes with a resolution of 2.0–2.4 Å. It was shown that the central role is assumed by the carboxy-group of the Asp-124 residue as the nucleophile in the first (alkylation) stage and by a water molecule, attacking the central carbon atom of the ester group of the covalent intermediate, in the second (dealkylation) stage.

2. Deacylation of trypsin

A similar study has been carried out with a classical object in enzymology — trypsin.[261] The Laue method for the recording of synchrotron radiation with time resolution was used in the above study.

The chemical mechanism of the catalytic step in the reactions involving serine proteases, including trypsin, presented below has been known for a fairly long time.[279–281]

$$\underset{\text{I}}{E+S} \rightleftharpoons \underset{\text{II}}{ES} \rightleftharpoons \underset{\text{III}}{ES^*} \rightleftharpoons \underset{\text{IV}}{EA + P_1} \rightleftharpoons \underset{\text{V}}{EA^*} \rightleftharpoons \underset{\text{VI}}{EA} \rightleftharpoons \underset{\text{VII}}{E+P_2}$$

E = enzyme, S = substrate, A = acylenzyme, P_1 and P_2 = hydrolysis products.

According to the proposed scheme, the mechanism of the tryptic hydrolysis includes a consecutive chain of chemical interrelations between the enzyme and the substrate via covalent intermediates. The catalytic process begins with the formation of a nonvalence Michaelis complex (II) in which the hydroxyl of Ser-195 and the imidazole ring of His-57 in the enzyme are respectively close to the carbonyl and amide groups of the peptide or ester bond of the substrate to be cleaved. As a result of their concerted interactions, the nonvalence enzyme–substrate bonding is converted into covalent bonding with the initial formation of a relatively unstable intermediate — the so called tetrahedral adduct (III). The latter decomposes into the acylenzyme and the amine (IV), while on hydrolysis of the ester it decomposes into the acylenzyme and the alcohol. Next follows deacylation, which proceeds similarly in principle, but in the opposite sequence and with participation of a water molecule as the nucleophile and not the O^γ atom of the Ser-195 side chain. A metastable tetrahedral adduct (V) is again formed and then decomposes, via the nonvalence complex (VI) stage, into the initial enzyme and the second reaction product (VII). The chemical mechanism for the proteolytic serine enzymes was developed before the results of the X-ray diffraction analysis of the three-dimensional structure of the molecules became available.

The crystal structure of native trypsin has been established[282–287] with a resolution of 1.8 Å and refined to within 1.5 Å. The geometry of the active centre was identified on the basis of the results of X-ray diffraction analysis of complexes of trypsin with protein inhibitors of animal and plant origin[288–290] and with irreversible synthetic inhibitors.[291] It was shown that the hydroxy-group of the Ser-195 residue and the imidazole ring of His-57, located on the surface of the protein globule and forming the $N^{\varepsilon 2}\ldots H-O^\gamma$ hydrogen bond, participate directly in the interaction with the substrate and in the formation of intermediate products of the catalytic reaction. The His-57 residue is also close to Asp-102, the $N^{\delta 1}-H\ldots O^{\delta 2}$ interaction occurring between them.

We shall now return to the study[261] in which the mechanism of action of trypsin in the second stage of the catalytic step (IV–VII) was investigated by X-ray diffraction with synchrotron radiation. The study began with the preparation of a stable intermediate of the enzymic reaction — the acylenzyme

(VI) — by the addition of *p*-guanidinobenzoate (GB) to the Ser-195 residue of trypsin present in the crystalline state in the mother liquor at pH 5.5. On rapid increase in the pH to 8.5, the process, monitored by three-dimensional Laue diffraction, was initiated. The diffraction of X-rays was detected in three series of irradiations of various sections of the crystalline specimen with a five-second exposure for one photograph in each series. The first series of exposures (t_0) was carried out before injecting a buffer solution at pH 8.5 into the flow cell, while the second and third series were carried out after 3 and 90 min respectively (t_3, t_{90}). The three-dimensional structures of three enzyme–substrate complexes were solved with a resolution of 1.8 Å and *R* factors of ~20%. Comparison of their structures before and after increasing the pH showed that one of the water molecules in the active centre of the enzyme is predisposed to attack the acyl group in the stage involving the initiation of hydrolysis of the acylenzyme and exhibits a tendency to approach the carbonyl carbon atom during the reaction. Thus the claim that water is involved in the deacylation of the trypsin intermediate is transformed from an extremely probable hypothesis into an experimental fact.

3. The structure of the intermediate compound in the first stage of the reaction of cytochrome *c* peroxidase with hydrogen peroxide

According to existing idea, cytochrome *c* peroxidase (CCP) catalyses the reduction of peroxides via a mechanism consisting of the following three stages:[292–295]

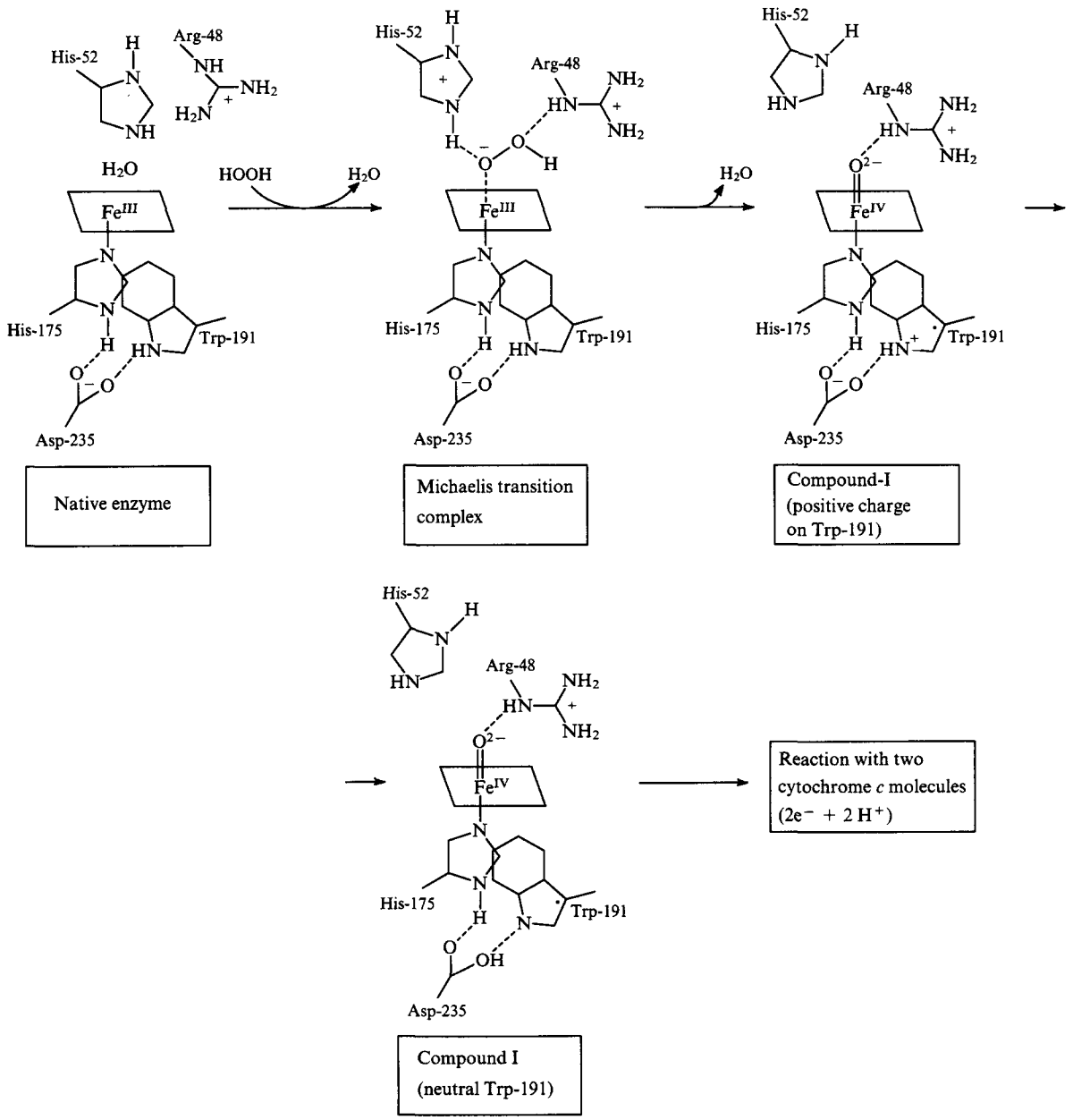
$$CCP(Fe^{III}) + H_2O_2 \rightarrow \text{compound-I } (Fe^{IV}O^{2-}, \text{Trp-191}^*) + H_2O \quad (1)$$

$$\text{Compound-I } (Fe^{IV}O^{2-}, \text{Trp-191}^*) + \text{Cyt-}c(Fe^{II}) + H^+ \rightarrow \text{compound-II } (Fe^{IV}O^{2-}) + \text{Cyt-}c(Fe^{II}) \quad (2)$$

$$\text{Compound-II } (Fe^{IV}O^{2-}) + \text{Cyt-}c(Fe^{II}) + H^+ \rightarrow CCP(Fe^{III}) + \text{Cyt-}c(Fe^{III}) + H_2O \quad (3)$$

In the first stage of the reaction of cytochrome *c* peroxidase with hydrogen peroxide, a doubly oxidised intermediate, which has been called 'compound-I', is formed. The three-dimensional structures of the native enzymes and of compound-I as well as

Scheme 9



the conformational changes in CCP, occurring on formation of the intermediate in the first stage of the catalytic reaction, have been described.[259]

Compound-I was obtained in a crystalline form from yeast cytochrome *c* peroxidase on reaction of the crystalline enzyme, placed together with the mother liquor in a flow cell, with hydrogen peroxide, also in the mother liquor. The reaction between the enzyme and H_2O_2, the occurrence of which was detected with the aid of absorption spectra in the visible region (550–750 nm), was almost completed after several minutes at 6 °C. The concentration of compound-I formed then remained virtually constant (90%–95%) for at least 30 min. The three-dimensional structures of the native enzyme and of its first intermediate were solved by three-dimensional Laue diffraction with a resolution of 2.2 Å and a divergence factor $R \sim 15.8\%$ (Table 8, No. 3). The size of the crystalline specimens was 0.4 × 4.0 × 1.0 mm. Polychromatic synchrotron radiation was employed; the duration of one exposure was ~0.3 s. Altogether, four series of independent experiments were performed on four crystals for four different orientations of the specimens (four for the native enzyme and four for compound-I).

On comparing the geometries of the observed structures, the authors[259] noted changes induced by the conversion of CCP into compound-I. They concerned only the region of the active centre and principally its distal part (Scheme 9). In the native state of the enzyme, the ligand-binding pocket over the iron atom is filled with three water molecules. On formation of the enzyme–substrate complex, two of them are eliminated, while the third is slightly displaced but still remains bound to the Trp-51 residue and the Fe-containing haem group. After the formation of compound-I and the heterolytic cleavage of the peroxide bond, the substrate oxygen atom is bound to the iron, altering its oxidation state from the high-spin ferri-state to the low-spin ferro-state ($Fe^{III} \rightarrow Fe^{IV}$). Having formed a Fe–O bond ~1.7 Å long, the iron atom is displaced by 0.3 Å from the plane of haem in the direction of the distal His-52. The Arg-47 side chain, located somewhat to one side, is displaced by 1.3 Å towards the ferro-oxygen atom and forms a hydrogen bond with the latter. The distances between the ferro-oxygen atom, on the one hand, and the N^{ε} atom of the Arg-48 residue, the $N^{\varepsilon 1}$ atom of the Trp-51 residue, and the $N^{\varepsilon 2}$ atom of the His-52 residue are 2.8, 2.7, and 3.6 Å respectively, and are too large for the formation of a hydrogen bond.

On formation of compound-I, the indole ring of the proximal Trp-191 residue is oxidised by giving up an electron to ferro-oxygen. The resulting radical-cation, bearing a charge on the indole nitrogen atom, interacts via the $N^+-H\ldots O^{\delta 1}$ hydrogen bond with the oxygen of the charged carboxy-group of the Asp-235 side chain; the second oxygen atom of this residue ($O^{\delta 2}$) is bound to the N–H group of the His-175 residue. Since there is no appreciable change in the geometry of the active centre in this region, the above considerations concerning the distribution of electron density in the proximal residues constitute merely one of the possible versions.

4. Photolysis of the myoglobin – carbon monoxide (Mb – CO) complex

Two structural studies on this complex, which in experimental terms can be regarded as the greatest advances in modern protein crystallography, have been published recently.[296,297] Both investigations were devoted to the mechanism of the photodissociation of the carbon monoxide-inhibited myoglobin.

$$Mb-C + h\nu \rightarrow Mb^*OC \rightarrow Mb + CO$$

The aim of both studies was to determine the position of the CO molecule in the transition state and to elucidate the response of the protein active centre to the dissociation of the Fe–O bond. The experimental solution of both problems required the determination of the three-dimensional structure of the short-lived intermediate Mb* and its comparison with the structures of the native myoglobin and its inhibitor complex.

Schlichting et al.[296] chose as the object of study a mutant myoglobin forming a hexagonal lattice. The specimen was cooled to ~20 K and irradiated with visible light from the usual source to initiate and maintain the photolysis; the X-ray diffraction pattern was obtained with the aid of synchrotron radiation (Fig. 21). The second experiment mentioned was formulated in accordance with a similar scheme.[297] The differences consisted in the employment in the latter case of native myoglobin (from whale sperm), the crystals of which have a monoclinic and not a hexagonal lattice; the specimen was cooled to ~40 K and was irradiated with light from a powerful laser source.

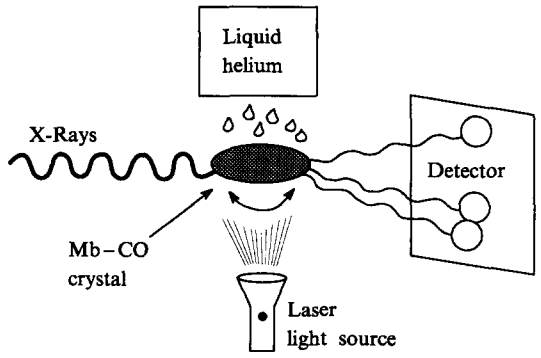


Figure 21. Schematic illustration of the experimental apparatus for X-ray diffraction by Mb–CO crystals during photolysis.[296–298]

Apart from certain quantitative discrepancies in the estimation of the conformational changes in the myoglobin active centre, the results of both investigations agreed. It was found in both cases that, on photodissociation of the complex, the CO molecule is removed from the Fe atom but remains in the active centre and is located parallel to the plane of the haem group at a distance of ~3.6 Å. The Fe atom shifts simultaneously in the same direction over a distance of 0.2 Å (data of Schlichting et al.[296]). In addition, on photolysis, changes were noted in the positions of the side chains of the distal His and Arg residues and of the proximal His residue and also of the α-helix F. Teng et al.[297] did not observe such changes, which apparently cannot be accounted for by the different packing of the crystals in the native and mutant myoglobin. It had been shown earlier that the mutation of the specimen employed by Schlichting et al.[296] does not entail conformational changes in the three-dimensional structure of the protein and does not affect it physiological properties.[299] The structural changes noted on photolysis agree with spectroscopic data[300,301] and the results of a molecular dynamics calculation.[302,303]

The results of the X-ray diffraction studies on the mechanisms of the functioning of proteins, carried out using synchrotron radiation and published towards the end of 1994, have been examined in this section. So far there have been few such investigations, but there is no doubt that they reflect the fundamental trend in the development of the method. Whereas only a few years ago, its purpose might have been formulated as concerning only the study of the three-dimensional structures of protein molecules and their stable complexes, nowadays the three-dimensional structures of the intermediates at the same atomic level have become accessible to investigation. Thus X-ray diffraction analysis has passed from the study of the molecular morphologies of the starting materials and final products of physiological reactions to the molecular morphologies of the intermediates. This finding denotes a jump in the development of the method — a qualitative change in its possibilities for the solution of an enormous number of new problems in molecular biology.

VIII. Multiwave anomalous diffraction

The most complex problem in X-ray crystallography is undoubtedly the phase problem. It has in fact held back the elucidation of diffraction patterns and the determination of three-dimensional molecular structures.

In the case of proteins, a solution to the phase problem was found by introducing into the X-ray diffraction analysis isomorphous derivatives — nonvalence complexes of the protein with heavy atoms (Au, Pb, Hg, Ag, Tl, etc.), the addition of which should not distort the three-dimensional structure of the protein molecule and the crystal lattice. The differences between the intensities of the beams diffracted by the crystals of the native protein and its derivatives can be used to determine the phases, provided that one establishes beforehand the positions of the heavy atoms. There are several different ways of solving this problem.[5,304]

The method for the determination of phases with the aid of heavy atoms, which has been called multiplet isomorphous substitution (MIS), was first tested in 1954 in a study of the crystal structure of haemoglobin.[305] Since then and almost until the very latest time, it remained virtually the only approach to the calculation of the required phase angles. However, throughout all these years the search for an alternative approach did not cease, because the MIS method is far from perfect.

Another solution to the phase problem became possible by employing the anomalous X-ray scattering effect. The classical theory of this phenomenon was developed by Honl as early as 1933.[306]

According to diffraction theory, the atomic scattering amplitude f is a complex quantity and can be formulated as follows:

$$f = f^0 + f^\Delta e^{ik} = f^0 + f' + f'' ,$$

where f^0 is the amplitude of the elastic 'normal' scattering, whilst $f^\Delta e^{ik}$ is the 'anomalous' correction formulated as the sum of the real dispersion (f') and imaginary absorption (f'') components. The correction f^Δ depends greatly on the radiation wavelength λ, whereas the usual scattering amplitude (f) is largely a function of the Bragg scattering angle θ.

The absolute value of f'' is proportional to the absorption coefficient. It therefore acquires great importance when the frequency of the X-ray quantum is close to the absorption edge of the atom.

The crystals of compounds containing no anomalous centres scatter X-rays in the normal manner. There is no anomalous scattering in this case, since the frequencies of the intrinsic vibrations of the H, C, N, O, and S atoms are far too remote from the frequency of the K emission for resonance to occur. The overall diffraction by a crystal (including a protein crystal) having anomalously scattering centres is described for each measured reflection by the structure factor ${}^\lambda F_T$. It includes the normal component ${}^0F_T$, consisting of the scattering components due to all the atoms in the protein structure, and the anomalous component. The latter is equal to the product of the normal components for the anomalously scattering centres ${}^0F_A$ and the ratio of the anomalous and normal scattering factors $f^\Delta(\lambda)/f^0(\theta)$. Consequently

$${}^\lambda F_T = {}^0F_T + {}^0F_A \cdot f^\Delta(\lambda)/f^0(\theta) .$$

Karle in 1980 [307] and later Hendrickson [308] showed that precise measurements of the intensities of all the diffraction reflections from a single crystalline protein specimen with anomalous scattering, obtained for several (3–5) X-ray wavelengths, make it possible to calculate from the set of experimental ${}^\lambda I$ data the real components of the factors $|{}^0F_T|$, $|{}^0F_A|$, and $\Delta\Phi^0 = \Phi_T^0 - \Phi_A^0$. The values of $|{}^0F_A|$ are then used to determine the positions in space of the anomalous scattering centres. Initially the values of Φ_A^0 and then, knowing $\Delta\Phi^0$, those of Φ_T^0 are calculated on this basis.

Thus the phase problem may also be solved without resorting to the multiplet isomorphous substitution method. However, the cost of such a solution remained for a long time too high and at least two conditions, fairly difficult to fulfill, had to be observed. Firstly, the test specimen must contain anomalously scattering atoms, which may not be present in the native protein. Secondly, in order to optimise the anomalous scattering effect, it is necessary to have a source of X-rays, such as the radiation from a synchrotron, continuously tuned to different wavelengths with a high accuracy and over a wide spectral range. Currently, the above difficulties of the multiwave anomalous diffraction (MAD) method have been largely overcome and its evident advantages over the traditional approach to the solution of the phase problem (MIS) have begun to play a decisive role.

The monopoly of the isomorphous substitution method in the X-ray crystallography of proteins was first broken in 1981 [309] when it became possible to establish the three-dimensional atomic structure of crambin from the diffraction pattern of a single crystalline specimen. The phase problem was solved by the MAD method employing three wavelengths of the synchrotron radiation and the anomalous scattering by the sulfur atoms in the methionine residues forming part of the aminoacid sequence of the protein. It became evident that synchrotron radiation may render MAD a powerful instrument in the study of protein structures. However, the use of sulfur as the anomalously scattering centre was not entirely satisfactory. The dispersion (f') and absorption (f'') components of the anomalous scattering factor for this atom are much smaller than the normal component of the structure factor and are many times smaller than the corresponding values of the components of the anomalous scattering by heavier atoms. In the subsequent X-ray diffraction analysis of proteins with the aid of MAD, the use of sulfur as the anomalously scattering atom had to be abandoned. Proteins containing metal atoms (Fe, Zn, Cu, etc.) in the native state and isomorphous heavy atom derivatives of proteins became the objects of study, as in the MIS method. After crambin, the structures of parvalbumin,[310] cytochrome c',[311] azurin,[312] haemoglobin,[313,314] ferroxin,[314] selenolanthionine,[315] and selenobiotinyl-streptavidin [314] were solved by using anomalous scattering.

The experience gained in studies involving MAD made it possible to improve many experimental and mathematical procedures of the resonance method,[316] as a result of which the calculations of the phases with the aid of MAD are just as accurate as in the multiplet isormorphous substitution procedure. The solution to the phase problem based on the anomalous scattering effect has become universal and is applicable to any crystalline protein specimens containing anomalously scattering centres. The method for the introduction of diffraction labels, necessary in connection with the use of MAD, has also become general.[317]

The accuracy of the X-ray diffraction data obtained with the aid of MAD and MIS evidently improves the closer the isomorphism of the protein derivatives with resonance centres (for MAD), and with heavy atoms (for MIS). In this respect, the situation is preferable in the case of anomalous scattering, because here all the necessary data for the determination of the three-dimensional structure of the protein molecule can be obtained by irradiating a single crystal at several wavelengths. In order to make use of this advantage, it is necessary to have a method which would guarantee the reduction to a minimum of the probability of the breakdown of iso-morphism when resonance labels are employed in situ, which would be sufficiently simple, and which could be used in the X-ray diffraction analysis of any protein. The method used to obtain selenomethionine protein preparations, giving rise to anomalous scattering, is extremely close to satisfying these requirements.[317]

As early as 1957, it was established that auxotrophic [auxoheterotrophic? (Ed. of Translation)] *E. coli* cells, which have lost the ability to synthesise methionine, can develop successfully and can give rise to at least 100 generations after the addition of selenomethionine to the nutrient medium.[318]

It was shown thereby that the proteins required for the maintenance of the life of *E. coli*, in which all the Met residues have been replaced by Se-Met, continue to function normally. Later these results were confirmed by the deliberate creation of a clone auxotrophic in relation to methionine by damaging the MetA gene.[319] It became possible to synthesise the aminoacid sequences of *E. coli* in which all the sites designated for the Met residues contain the Se-Met residues. Developing further this aspect of the research, Hendrickson et al.[317] devised a method for the incorporation of Se-Met residues into recombinant proteins produced by the *E. coli* plasmids. The authors' preceding studies on the three-dimensional structures of the selenolanthionine[315] and selenobiotinylstreptavidin[314] molecules with the determination of the phases by the MAD method showed that the Se atom is an extremely effective anomalous scattering centre. One can therefore assume that the proposed method for the incorporation of selenomethionine residues in the methionine positions will become a universal approach for proteins, especially those which do not contain natural resonators in the native state.

The authors of the method used it to obtain two thioredoxins of the T4 cells of *E. coli* with 100% replacement of methionine by selenomethionine. The crystals of [Se-Met]-substituted thioredoxins proved to be fully isomorphous with the crystals of the natural recombinant proteins. The diffraction measurements were performed at three selected synchrotron radiation wavelengths. One wavelength corresponded to minimum anomalous scattering (f'_{min}), the second corresponded to the maximum absorption energy (f''_{max}), and the third wavelength was remote from the band due to the electronic transitions in the Se atom and hence corresponded to the normal scattering factor (f^0). Fig. 22 presents the dependence of the anomalous scattering factors f' and f'' on the irradiation wavelength obtained for a single orientation of the *E. coli* [Se-Met]-thioredoxin crystal. The high sensitivity of f' and f'' to changes in the wavelength in the region of 0.98 Å and the distinct edge effect indicate the considerable possibilities of the MAD method in the solution of the phase problem: the phases are initially found for several anomalously scattering centres and then, on this basis, for all the nonhydrogen atoms of the protein (for further details, see Hendrickson[320]).

Thus, towards the beginning of the 1990s the internal problems of the multiwave anomalous diffraction method were largely solved. Hendrickson's method for the preparation of selenomethionine derivatives of proteins rapidly gained recognition. In the studies of recent years, the crystals of such proteins have frequently served for the measurement of the diffraction patterns arising from both anomalous (using the resonance wavelengths) and normal scattering (considering selenium as a heavy atom). The study of Ramakrishnan et al.,[273] in which the three-dimensional structure of the globular domain of the linker histone H5 was determined with a resolution of 2.5 Å, is precisely of this kind.[273] The [Se-Met]-derivative of the domain was synthesised by genetic engineering methods. The crystals grown from it were fully isomorphous with the crystals of the native protein. The phases were determined by means of the MAD and MIS methods. The authors noted that treatment of the results by the traditional method combined with the recently proposed procedure for the refinement of the heavy atom parameters, called the maximum entropy or maximum similarity method,[321,322] led to somewhat preferable electron density maps compared with the procedure called the algebraic method.[316]

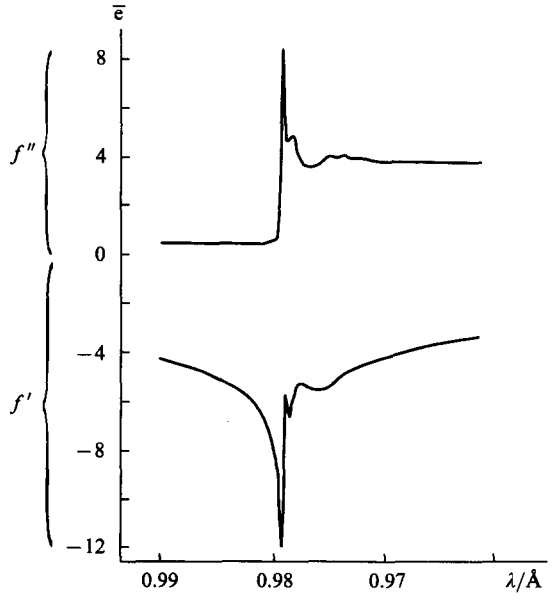


Figure 22. Anomalous scattering factors f' and f'' for crystals of *E. coli* selenomethionine thioredoxin measured in the vicinity of the absorption edge of selenium.

Newman et al.[323] established with a resolution of 1.95 Å the crystal structure of restriction endonuclease Bam HI, consisting of 213 aminoacid residues. The authors resorted to the use of the selenomethionine version after the failure of attempts to interpret the Patterson map constructed from measurements for several heavy atom protein derivatives. The [Se-Met]-endonuclease was expressed in the auxotrophic *E. coli* cells grown in a medium incorporating selenomethionine instead of methionine together with 19 standard aminoacids.[317] The object of study thus obtained had the same activity as the native enzyme and its crystals were fully isomorphous with those of the latter. The novel feature of the above investigation is the fact that the three-dimensional structure of the endonuclease was determined by the combination of the phase information, contained in the anomalous diffraction data, obtained by the maximum entropy and algebraic methods.[316,321,322] Both methods lead to approximately the same phase errors in relation to the phases of the refined model. The authors suggest that the averaged phases are appreciably closer to reality than those obtained by one of the methods. The [Se-Met]-BamHI crystals were irradiated at 5 °C at three wavelengths corresponding to the Se absorption edge (0.979 Å) and peak (0.978 Å) and the remote region of the spectrum (0.900 Å).

A human embryo gonadotropin, the structure of which has been described,[324] is a glycoprotein of the placenta hormone (237 aminoacid residues) stimulating the secretion of the steroid progesterone, which maintains the pregnancy. Like the recombinant [Se-Met]-protein, it was synthesised in the cells of the hamster ovary. This somewhat unexpected finding shows that the eukaryotic mammalian cells can be just as highly tolerant as the prokaryotic *E. coli* cells as regards the complete substitution of methionine by selenomethionine. The level of substitution in the above study was 92%. The structure of gonadotropin after partial deglycosylation was identified with a resolution of 2.6 Å. Diffraction data obtained by the MAD method were employed. The crystalline specimen was irradiated at four synchrotron radiation wavelengths. Two of them (~0.98 Å) were at the Se absorption edge and peak, whilst the other two were to one side of the absorption band. The approximate values of the anomalous scattering components f' and f'' were found for each wavelength.

The MAD method has been used[276] to determine the three-dimensional structure of the allosteric regulator enzyme glutamine–5-phosphoribosyl-1-pyrophosphate (PRPP) amidotransferase.[276] The protein consists of 465 aminoacid residues and an Fe atom, which is an anomalously scattering centre on irradiation at three wavelengths from a synchrotron (1.7425, 1.7390, and 1.5000 Å). The crystalline form of PRPP amidotransferase was obtained in the form of a homotetramer with a molecular mass of ~200 kDa. In the preceding X-ray diffraction analyses by the anomalous diffraction method, the objects of

study were proteins whose molecular masses did not exceed 30 kDa. The above study indicates the major possibilities of MAD in the determination of the three-dimensional structures of cumbersome protein complexes.

IX. Conclusion

The studies examined clearly leave no doubt that during the 1990s the X-ray diffraction analysis of proteins was applied to the solution of new problems of primary interest for molecular biology whilst retaining the previous high rate of extensive development. The main cause of the qualitative change which has arisen in the possibilities of the crystallography of macromolecules is associated with the employment of synchrotron radiation and new methods for the measurement of the diffraction pattern. Firstly, the transition to a new source of X-rays weakens the requirements which must be met by the size of the crystals, which is especially important in the structural analysis of high-molecular-mass proteins and their complexes with large unit cells. Secondly, the continuous synchrotron radiation spectrum and the ease of selecting any wavelength of the monochromatic radiation permitted a new approach to the solution of the phase problem and the development of the multiwave anomalous diffraction method, in which only one crystalline specimen is required for the determination of the phases. The procedure for the recombinant formation in auxotropic cells of proteins in the aminoacid sequences of which all the methionine residues have been replaced by selenomethionine supplemented significantly the MAD method. The use of [Se-Met]-proteins not only frees the X-ray diffraction analysis from the prolonged routine procedure involving the preparation of several isomorphous heavy atom protein derivatives but virtually eliminates the very problem of isomorphism.

Finally, the most important feature—the application in protein crystallography of synchrotron radiation and of new coordinate detectors—reduces the exposure time by at least 2–3 orders of magnitude. It is difficult to overestimate the importance of this fact, since the determination of the three-dimensional structures at the atomic level of intermediates in the processes involved in vital activity becomes possible. Whereas until recently, X-ray diffraction analysis was in fact the only direct method for the investigation of the native conformations of protein molecules with atomic resolution, nowadays, literally in front of our very eyes, it is becoming a source (also the only one available) of just as detailed information about the three-dimensional structures of intermediates.

References

1. H W Wyckoff, K D Hardman, N M Allewell, T Inagami, D Tsernoglou, L N Honson, F M Richards *J. Biol. Chem.* **242** 3749 (1967)
2. G Kartha, J Bello, D Harker *Nature (London)* **213** 862 (1967)
3. B W Matthews, B Sigler, R Henderson, D M Blow *Nature (London)* **214** 652 (1967)
4. G N Reeke, J A Hartsuck, M L Ludwig, W N Lipscomb *Proc. Natl. Acad. Sci. USA* **58** 2220 (1967)
5. T L Blundell, L N Johnson *Protein Crystallography* (London: Academic Press, 1976)
6. *Protein Data Bank, Int. Union Cryst. Quarterly Newsletter* **1** 67 (1994)
7. J C Boyington, B J Gaffney L M Amzel *Science* **260** 1482 (1993)
8. A Helenius, K Simons *Biochim. Biophys. Acta* **415** 29 (1975)
9. R Henderson, N T Unwin *Nature (London)* **257** 28 (1975)
10. J Deisenhofer, O Epp, K Miki, R Huber, H Michel *J. Mol. Biol.* **180** 385 (1984)
11. J Allen, G Feher, T Yeats, H Komiya, D C Rees *Proc. Natl. Acad. Sci. USA* **84** 5370 (1987)
12. M S Weiss, A Kreusch, E Schiltz, U Nestel, W Welte, J Weckesser, G E Schulz *FEBS Lett.* **280** 379 (1991)
13. A M de Vas, M Ultsch, A A Kossiakoff *Science* **255** 306 (1992)
14. M V Milburn, G G Prive, D L Milligan, W K Herber, P L Darke, N E Kohl, A F Dixon *Science* **252** 1342 (1991)
15. A F Russo, D E Koshland *Science* **220** 1016 (1983)
16. A Krikos, N Mutoh, A Boyd, K Kendall, M I Simon *Cell* **33** 615 (1983)
17. J Jancarik, W G Scott, D L Milligan, D E Koshland, S-H Kim *J. Mol. Biol.* **221** 31 (1991)
18. W L Smith, P Borgeat, F A Fitzpatrick, in *Biochemistry of Lipids, Lipoproteins and Membranes* (Eds D F Vance, J Vance) (New York: Elsevier, 1991) p. 297
19. I J Marnett *Environ. High Perspect.* **88** 5 (1990)
20. A N Makheja *Mol. Cell. Biochem.* **111** 137 (1992)
21. S Abramson *Curr. Opin. Rheumatol.* **3** 336 (1991)
22. D Picot, J Loll, R M Garavito *Nature (London)* **367** 243 (1994)
23. G Blobel *Proc. Natl. Acad. Sci. USA* **77** 1496 (1980)
24. M F Greaves, J J Owen, M C Raff *T u B Lymphocytes: Origins, Properties and Roles in Immune Responses* (Amsterdam: Excerpta Medica, 1973)
25. B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson *Molecular Biology of the Cell* Vol. 5 (New York: Garland, 1983)
26. F M Burnet *The Clonal Selection Theory of Acquired Immunity* (Nashville, TN: Vanderbilt University Press, 1959)
27. B E Biere, B Sleckman, S E Ratnofsky, S J Burakoff *Annu. Rev. Immunol.* 7 579 (1989)
28. H Cantor, R K Gershon *Fed. Proc.* **38** 2058 (1979)
29. R M Zinkernagel *Hosp. Pract.* **13** 83 (1978)
30. R Billingham, W Silvers *The Immunobiology of Transplantation. Foundations of Immunology Series* (Englewood Cliffs, NJ: Prentice-Hall, 1971)
31. J L Strominger, in *Immunobiology* (Ed. M E Dorf) (New York: Garland STPM, 1981) p. 115
32. R M Zinkernagel, C Doherty *Adv. Immunol.* **27** 51 (1979)
33. I Roitt, in *Essential Immunology* (Oxford: Blackwell, 1980)
34. F Ramsdell, B J Fowlkes *J. Immunol.* **143** 1467 (1989)
35. J C Zuniga-Pflucker, L A Jones, D L Longo, A M Kruisbeek *J. Exp. Med.* **171** 427 (1990)
36. W Fung-Leung, M W Schilham, A Rahemtulla, T M Kundig, M Vollenweider, J Potter, M Swijk, T Mak *Cell* **65** 443 (1991)
37. S L Swain *Immunol. Rev.* **74** 129 (1983)
38. Z Dembic, W Haas, R Zamoyska, J Parnes, M Steinmetz, H Boehmer *Nature (London)* **326** 510 (1987)
39. J Gabert, C Langlet, R Zamoyska, J Parnes, A Schmitt-Verhulst, B Malissen *Cell* **50** 545 (1987)
40. A M Norment, D R Littman *EMBO J.* 7 3433 (1988)
41. Y Rosenstein, S Ratnojsky, S J Burakoff, S H Herrmann *J. Exp. Med.* **169** 149 (1989)
42. R D Salter, R J Benjamin, K Wesley *Nature (London)* **345** 41 (1990)
43. S K Sanders, R O Fox, P Kavathas *J. Exp. Med.* **174** 371 (1991)
44. A Veillete, M A Bookman, E M Horak, J B Bolen *Cell* **55** 301 (1988)
45. D R Littman, Y Thomas, P Maddon, L Chess, R Axel *Cell* **40** 237 (1985)
46. A M Norment, D R Littman *EMBO J.* 7 3433 (1988)
47. D J Leahy, R Axel, W A Hendrickson *Cell* **68** 1145 (1992)
48. J Bjorkman, M A Saper, B Samraoui, W S Bennett, J L Strominger, D C Wiley *Nature (London)* **329** 506 (1987)
49. R D Salter, R J Benjamin, K Wesly, S E Buxton, T P J Garrett, C Clayberger, A M Krensky, A M Norment, D R Littman, P Parham *Nature (London)* **345** 41 (1990)
50. J H Brown, T Jardetsky, M A Saper, B Samraoui, P J Bjorkman, D C Wiley *Nature (London)* **332** 845 (1988)
51. G Lange, S J Lewis, G N Murshudov, G G Dodson, P C Moody, J P Turkenburg, A N Barklay, R L Brady *Structure* **2** 469 (1994)
52. J Wang, J Liu, D W Rodgers, R L Garlick, G E Torr, Y Hussain, E L Reinherz, S L Harrison *Nature (London)* **338** 411 (1990)
53. T J Garrett, J Wang, Y Yan, J Liu *J. Mol. Biol.* **234** 763 (1993)
54. R L Brady, D G Lange, S J Davis, A F Williams, A N Barklay *Science* **260** 979 (1993)
55. G Cammarota, A Scherle, B Takas, D M Doran, R Knorr, W Bannwarth, J Guardiola, F Sinegaglia *Nature (London)* **356** 799 (1992)
56. T A Potter, T V Rajan, R F Dick, J A Bluestone *Nature (London)* **337** 73 (1989)

57. Y Rosenstein, S Ratnofsky, S J Burakoff, S H Hermann *J. Exp. Med.* **169** 149 (1989)
58. J Maddon, D R Littman, M Godfrey, D E Maddon, L Chess, R Axel *Cell* **42** 93 (1985)
59. D J Capon, S A Marsters, T Gregory, H Mitsuya, R A Birn, C Lucas, F M Wurm, J E Groopman, S Broder, D H Smith *Nature (London)* **337** 525 (1989)
60. D Klatzmann, D H Crawford, M F Greaves, R Weiss *Nature (London)* **312** 767 (1984)
61. A G Dalgleish, C L Beverley, R Clapham, D H Crawford, M F Greaves, R A Weiss *Nature (London)* **312** 763 (1984)
62. A F Williams, S J Davis, Q He, A N Barklay *Cold Spring Harbor Sym. Quant. Biol.* **54** 637 (1989)
63. D Kwong, S-E Ryu, W A Hendrickson, R Axel, R M Sweet, G Folena-Wasserman, P Hensley, R W Sweet *Proc. Natl. Acad. Sci. USA* **78** 6423 (1990)
64. S J Davis, H A Ward, M J Puklavec, A C Willis, A F Williams, A N Barklay *J. Biol. Chem.* **265** 10410 (1990)
65. D F Hunt, J Shabanowitz, A L Cox, K Sakaguchi, E Appella, H M Grey, A Sette *Science* **256** 1817 (1992)
66. R M Chicz *J. Exp. Med.* **178** 27 (1993)
67. J L Tiwari, P I Terasaki *HLA and Disease Associations* (Berlin: Springer, 1985)
68. J A Todd, H Acha-Orbea, J I Bell, N Chao, Z Fronek, C O Jacob, H O McDevitt *Science* **240** 1003 (1988)
69. J C Gorga, J H Brown, T Jardetzky, D C Wiley, J L Strominger *Res. Immunol.* **142** 401 (1991)
70. J H Brown, T Jardetzky, C Gorga, L J Stern, R G Urban, J L Strominger, D C Wiley *Nature (London)* **364** 33 (1993)
71. J Bjorkman, W C Bennett, L Strominger, D C Wiley *Nature (London)* **329** 506 (1987)
72. J H Brown, T Jardetzky, M A Saper, B Samranoui, P J Bjorkman, D C Wiley *Nature (London)* **332** 845 (1988)
73. R Konig, L-Y Huang, R N Germain *Nature (London)* **356** 796 (1992)
74. L Jorgensen, A Reay, E W Ehrich, M M Davis *Annu. Rev. Immunol.* **10** 835 (1992)
75. R N Germain, D H Margulies *Annu. Rev. Immunol.* **11** 403 (1993)
76. K Falk, O Rotzschke, S Stevanovic, G Jung, H-G Rammensee *Nature (London)* **351** 290 (1991)
77. D F Hunt *Science* **356** 1261 (1992)
78. T S Jardetzky, W S Lane, R A Robinson, D R Madden, D C Wiley *Nature (London)* **353** 326 (1991)
79. A Y Rudensky, P Preston-Huriburt, B K Al-Ramadi, J Rothbard, C A J Janeway *Nature (London)* **359** 429 (1992)
80. R M Chicz, C Gorga, L J Stern, D A A Vignali, J L Strominger *Nature (London)* **358** 764 (1992)
81. J Hammer, D Bolin, J Higelin, B Takacs, F Sinigaglia *Cell* **74** 197 (1993)
82. A Sette *J. Immunol.* **151** 3163 (1993)
83. D R Madden, J C Gorga, J L Strominger, D C Wiley *Nature (London)* **353** 321 (1991)
84. D R Madden, J C Gorga, J L Strominger, D C Wiley *Cell* **70** 1035 (1992)
85. M L Silver, H-C Guo, J L Strominger, D C Wiley *Nature (London)* **360** 367 (1992)
86. D R Madden, D N Garboczi, D C Wiley *Cell* **75** 693 (1993)
87. W Zhang, A C M Young, M Imarai, R G Nanni, T Li, G F Arnold, D A Oren, A Jacobo-Molina, R L Williams, G Kamer, D A Rubenstein, Y Li, E Rozhon, S Cox, P Buontempo *J. Connell. Acad. Sci. US* **89** 8403 (1992)
88. A C M Young, W Zhang, J C Sacchettini, S G Nathenson *Cell* **76** 39 (1994)
89. D H Fremont, M Matsumara, E A Stura, P A Petersen, I A Wilson *Science* **257** 919 (1992)
90. D R Madden, D C Wiley *Curr. Opin. Struct. Biol.* **2** 300 (1992)
91. M Matsumura, D H Fremont, A Peterson, I A Wilson *Science* **257** 927 (1992)
92. H C Guo, D R Madden, M L Silver, T S Jardetzky, J C Gorga, J L Strominger, D C Wiley *Proc. Natl. Acad. Sci. USA* **90** 8053 (1993)
93. L J Stern, J H Brown, T S Jardetzky, J C Gorga, R J Urban, J L Strominger, D C Wiley *Nature (London)* **386** 215 (1994)
94. N S Andreeva, A E Gustchina, A A Fedorov *Adv. Exp. Med. Biol.* **95** 23 (1977)
95. N S Andreeva, A A Fedorov, A E Gustchchina *Mol. Biol. (Moscow)* **12** 922 (1978))]
96. N S Andreeva, A S Zhdanov, A A Fedorov *Mol. Biol. (Moscow)* **18** 313 (1984)
97. A R Sielecki, A A Fedorov, A Boodhoo, N S Andreeva, M N J James *J. Mol. Biol.* **214** 148 (1990)
98. J B Cooper, G Khan, G Taylor, I J Tickle, T L Blundell *J. Mol. Biol.* **214** 199 (1990)
99. N S Andreeva, A S Zhdanov, A E Gustchina, A A Fedorov *J. Biol. Chem.* **259** 11353 (1984)
100. M H G James, A R Sielecki *Nature (London)* **319** 33 (1986)
101. J A Hartsuck, S J Remington, in *Abstracts of the 18th Linderstrom-Lang Conference on Aspartic Proteinases* (Elsinore, 1988) p. 40
102. M G Safro, N S Andreeva, A S Zhdanov *Mol. Biol. (Moscow)* **19** 400 (1985)
103. M Newman, M Safro, C Frazao, G Khan, A S Zhdanov, I J Tickle, T L Blundell, N S Andreeva *J. Mol. Biol.* **221** 1295 (1991)
104. G L Gilliland, E L Winborne, J Nachman, A Wlodaver *Proteins* **8** 82 (1990)
105. A R Sielecki, E Hayakawa, M Fuginaga, M E P Murphy, M Fraser, A K Muir, C T Carilli, J A Lewicki, J D Buxter, M N G James *Science* **241** 1346 (1989)
106. J Rahuel, J Priestle, M G Grutter *J. Struct. Biol.* **107** 227 (1991)
107. V Dharanaj, C G Dealwis, C Frazao, M Badasso, B L Sibanda, I J Tickle, J B Cooper, H P C Driessen, M Newman, C Aguilar, S P Wood, T L Blundell, P M Hobard, K F Geoghegan, M J Ammirati, D E Danley, B A O'Connor, D J Hoover *Nature (London)* **357** 466 (1992)
108. C G Dealwis, C Frazao, M Badasso, J B Cooper, I J Tickle, H Driessen, T L Blundell, K Murukami, H Mijazaki, J Suerias-Diaz, D M Jones, M Szelke *J. Mol. Biol.* **236** 342 (1994)
109. F Cumming, D Le-Nguen, B Castro, J Menard, P Carvol *Biochim. Biophys. Acta* **913** 10 (1987)
110. I N Hsu, L T J Delbaere, M N G James, T Hoffman *Nature (London)* **266** 140 (1977)
111. M N G James, A R Sielecki *J. Mol. Biol.* **163** 229 (1983)
112. M N G James, A R Sielecki *Biochemistry* **24** 3701 (1985)
113. M N G James, A R Sielecki, F Salituro, D H Rich, T Hoffmann *Proc. Natl. Acad. Sci. USA* **79** 6137 (1982)
114. M N G James, A R Sielecki, T Hoffmann, in *Aspartatic Proteinases and Their Inhibitors* (Ed. V Kostka) (Berlin: Walter de Gruyter, 1985) p. 163
115. E Subramanian, I D A Swan, M Li, D R Davies *Proc. Natl. Acad. Sci. USA* **74** 556 (1977)
116. Chi-h Wong, T J Lee, T-H Lu *Biochemistry* **18** 1638 (1979)
117. T Blundell, A Jenkins, B T Sewell, L H Pearl, J B Cooper, S P Wood, B Veerapandian *J. Mol. Biol.* **211** 919 (1990)
118. J B Cooper, S Foundling, A Hemminge, T L Blundell, D M Jones, A Hallett, M Szelke *Eur. J. Biochem.* **169** 215 (1987)
119. A Sali, B Veerapandian, J B Cooper, S I Foundling, D J Hoover, T L Blundell *EMBO J.* **8** 2179 (1989)
120. J B Cooper, S Foundling, J Boger, R A Jupp, J Kay, T L Blundell *Biochemistry* **28** 8596 (1989)
121. E Subramanian, M Lin, I D A Swan, D R Davies *Adv. Exp. Biol. Med.* **95** 33 (1977)
122. K Suguna, R R Bott, E A Padlan, E Subramanian, S Sheriff, G A Cohan, D K Davies *J. Mol. Biol.* **196** 877 (1987)
123. R Bott, E Subramanian, D Davies *Biochemistry* **28** 6956 (1982)
124. K Suguna, E A Padlan, C M Smith, W D Carlson, D R Davies *Proc. Natl. Acad. Sci. USA* **84** 7009 (1987)
125. M E Fraser, K Hayakawa, Y Yoshinaka *Viral Proteinases as Targets for Chemotherapy* (New York: Cold Spring Harbor Laboratory, 1989)
126. M Miller, M Jaskolski, M J K Rao, J Leiss, A Wlodawer *Nature (London)* **337** 576 (1989)
127. M A Navia, M D Fitzgerald, B M McKeever, J P Springer, M Poe, J Bodger, K Hoogsteen *Nature (London)* **337** 615 (1989)
128. A Wlodawer, M Miller, M Jaskolski, B K Sathyanarayana, E Baldwin, I T Weber, I M Selk, I Clawson, I Schneider, S B H Kent *Science* **245** 616 (1989)

129. M Jaskolski, A G Tomasselli, T K Sawyer, D S Staples, R L Hendrikson, J Sneider, S B H Kent, A Wlodawer *Biochemistry* **30** 1600 (1991)
130. A L Swain, M Miller, J Green, D H Rich, J Schneider, S B H Kent, A Wlodawer *Proc. Natl. Acad. Sci. USA* **87** 8805 (1990)
131. M Miller, J Schneider, B K Sathyanarayana, M V Toth, G R Marshall, L Clawson, L Selk, S B H Kent, A Wlodawer *Science* **246** 1149 (1989)
132. T N Bhat, E T Baldwin, B Liu, Y S Cheng, J W Erickson *Nature Struct. Biol.* **1** 552 (1994)
133. A Sali, B Verapandian, J B Cooper, D S Moss, T Hofmann, T L Blundell *Proteins* **12** 158 (1992)
134. V K Antonov, L M Ginodman, L D Rumsh *Eur. J. Biochem.* **117** 195 (1981)
135. V K Antonov *Adv. Exp. Med. Biol.* **95** 179 (1977)
136. V K Antonov, L M Ginodman, Yu V Kapitannikov *FEBS Lett.* **88** 87 (1978)
137. V K Antonov, L P Yavashev, L I Volkova, V L Sadovskayach, L M Ginodman *Bioorg. Khim.* **5** 1427 (1979) [*Sov. J. Bioorg. Chem.* **5** (1979)]
138. H Mitzusuya, R Yarchoan, S H Broder *Science* **249** 1533 (1989)
139. A G Tomasselli, W F Hove, T K Sawyer, A Wlodawer, R L Hendrikson *Chim. Oggi* 6 (1991)
140. N E Kohl, E A Emini, W A Schleif, L J Davis, J C Heimbach, R A F Dixon, E M Scolnick, I S Sigal *Proc. Natl. Acad. Sci. USA* **85** 4686 (1988)
141. J R Huff *J. Med. Chem.* **34** 2305 (1991)
142. L Pearl, W Taylor *Nature (London)* **329** 351 (1987)
143. Y S E Cheng, F H Yin, S Foundling, D Blomstrom, C A Kettner *Proc. Natl. Acad. Sci. USA* **87** 9660 (1990)
144. C L Dianni, L J Davis, M K Holloway, W K Herber, P L Darke, N E Kohl, R A F Dixon *J. Biol. Chem.* **265** 17348 (1990)
145. S Seelmeier, H Schmidt, V Turk, K von der Helm *Proc. Natl. Acad. Sci. USA* **85** 6612 (1988)
146. S F J Le Grice, J Mills, J Mous *EMBO J.* **7** 2547 (1988)
147. D D Leob, R Swanstrom, L Everitt, M Manchester, S E Stamper, C A Hutchison *Nature (London)* **340** 397 (1989)
148. X Lin, R N S Wong, J Tang *J. Biol. Chem.* **264** 4482 (1989)
149. L M Babe, S Pichuantes, C K Craik *Biochemistry* **30** 106 (1991)
150. R H Jacobson, X-J Zhang, R F Du Bose, B W Matthews *Nature (London)* **369** 761 (1994)
151. C G Dealwis, C Frazao, M Badasso, J B Cooper, I J Tickle, H Driessen, T L Blundell, K Murukami, H Miyazaki, D Sneiras-Diaz, D M Jones, M Szelke *J. Mol. Biol.* **236** 342 (1994)
152. G R Delpierre, J S Fruton *Proc. Nat. Acad. Sci. USA* **54** 1161 (1965)
153. M N G James *Can. J. Biochem.* **58** 3701 (1980)
154. M N G James, I N Hsu, L T J Delbaere *Nature (London)* **267** 808 (1977)
155. J S Fruton *Adv. Enzymol.* **33** 401 (1970)
156. T L Blundell, H B Jones, G Rhan *Proc. FEBS Meet.* **60** 281 (1980)
157. N S Andreeva, A E Gushchina, A A Fedorov *Dokl. Akad. Nauk SSSR* **259** 1284 (1981)
158. B Foltmann *Essays Biochem.* **17** 52 (1981)
159. L Pearl *FEBS Lett.* **214** 8 (1987)
160. W P Jencks *Adv. Enzymol.* **43** 219 (1975)
161. E M Popov *Strukturno-funktsional'naya Organizatsiya Belkov* (Structural Functional Organisation of Proteins) (Moscow: Nauka, 1992) p. 358
162. B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson *Molecular Biology of the Cell* Vol. 2 (New York: Garland, 1983)
163. R D Kornberg *Science* **184** 868 (1974)
164. R D Kornberg, J O Thomas *Proc. Natl. Acad. Sci. USA* **72** 2626 (1975)
165. R T Simpson *Biochemistry* **17** 5524 (1978)
166. A Klug, D Rhodes, J Smith, J T Finch, J O Thomas *Nature (London)* **287** 508 (1980)
167. P Oudet, M Gross-Bellard, P Chambon *Cell* **4** 281 (1975)
168. F Thoma, T Koller, A Klug *J. Cell. Biol.* **83** 403 (1979)
169. I Isenberg *Annu. Rev. Biochem.* **48** 159 (1979)
170. J Baldwin, G Boseley, E M Bradbury, K Ibel *Nature (London)* **253** 245 (1975)
171. L Perez-Grau, J Bordas, M N J Koch *Nucleic Acids Res.* **12** 2987 (1984)
172. G A Bentley, A Lewit-Bentley, J T Finch, A D Podjarny *J. Mol. Biol.* **176** 55 (1984)
173. J T Finch, L C Lutter, D Rhodes, R S Brown, B Rushton, M Levitt, A Klug *Nature (London)* **269** 29 (1977)
174. J T Finch, R S Brown, D Rhodes, T Richmond, B Rushton, L C Lutter, A Klug *J. Mol. Biol.* **145** 757 (1981)
175. T J Richmond, J T Finch, B Rushton, D Rhodes, A Klug *Nature (London)* **311** 532 (1984)
176. R W Burlingame, W E Love, E N Moudrianakis *Science* **223** 413 (1984)
177. R W Burlingame, W E Love, Bi-Ch Wang, R Hamlin, N H Huong, E N Moudrianakis *Science* **228** 546 (1984)
178. Bi-Ch Wang, J Rose, G Arents, R W Burlingame *J. Mol. Biol.* **236** 179 (1984)
179. G Arents, R W Burlingame, Bi-Ch Wang, E N Moudrianakis *Proc. Natl. Acad. Sci. USA* **101** 48 (1991)
180. J C Wang *J. Mol. Biol.* **55** 523 (1971)
181. R J DiGate, K Marians *J. Biol. Chem.* **264** 17924 (1989)
182. J W Wallis, G Chrebet, G Estesky, M Rolfe, R Rothstein *Cell* **58** 409 (1989)
183. R A Kim, J C Wang *J. Biol. Chem.* **267** 17178 (1992)
184. A Kikuchi, K Asai *Nature (London)* **309** 677 (1984)
185. F Confalonieri, C Elie, M Danal, C B de La Tour, P Forterre, M Duguet *Proc. Natl. Acad. Sci. USA* **90** 4753 (1993)
186. A Kornberg, T Beker *DNA Replication* (New York: Freeman, 1992)
187. Ch D Lima, J C Wang, A Mondragon *Nature (London)* **367** 138 (1994)
188. V-C Tse-Dinh, R K Beran-Steez *J. Biol. Chem.* **263** 15857 (1988)
189. R Lynn, J C Wang *Proteins* **6** 231 (1989)
190. J C Wang *Sci. Am.* **247** 94 (1982)
191. C D Lima, J C Wang, A Mondragon *J. Mol. Biol.* **232** 1213 (1993)
192. G E Schulz, R H Schirmer *Principles of Protein Structure* (Berlin: Springer, 1979)
193. B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson *Molecular Biology of the Cell* Vol. 1 (New York: Garland, 1983)
194. J J Champoux, in *DNA Topology and its Biological Effects* (Eds N R Cozzarelli, J C Wang) (New York: Cold Spring Harbor Laboratory, 1990) p. 217
195. E M Popov *Mol. Biol. (Moscow)* **11** 5 (1977)
196. R H Jacobson, X-J Zhang, R F DuBose, B W Matthews *Nature (London)* **369** 761 (1994)
197. M Ring, R E Huber *Arch. Biochem. Biophys.* **283** 342 (1990)
198. C G Cupples, J H Miller, R E Huber *J. Biol. Chem.* **265** 5512 (1990)
199. R D Vale, L S B Goldstein *Cell* **60** 883 (1990)
200. H E Huxley *Biochim. Biophys. Acta* **12** 387 (1953)
201. H E Huxley, J Hanson *Nature (London)* **173** 973 (1954)
202. V A Engelhardt, M N Ljubimova *Nature (London)* **144** 663 (1939)
203. B Katz *Nerve, Muscle and Synapsis* (New York: McGraw-Hill, 1966)
204. H E Huxley *Science* **164** 1356 (1969)
205. H E Huxley *Reflections on Muscle* (Princeton: Princeton University Press, 1980)
206. J Squire *The Structural Basis of Muscle Contraction* (New York: Plenum, 1981)
207. R Henderson, J M Baldwin, T A Ceska, F Zemlin, E Beckmann, K H Downing *J. Mol. Biol.* **213** 899 (1990)
208. W Kuhlbrandt, D N Wang, Y Fugiyoshi *Nature (London)* **367** 614 (1994)
209. K C Holmes *Structure* **2** 589 (1994)
210. A J Fisher, J E Johnson *Nature (London)* **361** 176 (1993)
211. R H Cheng, V S Reddy, N H Olson, A J Fisher, T S Baker, J E Johnson *Structure* **2** 271 (1994)
212. G Wang, C Porta, Z Chen, T S Baker, J E Johnson *Nature (London)* **355** 275 (1992)
213. K C Holmes, D Popp, W Gebhard, W Kabsch *Nature (London)* **347** 44 (1990)
214. W Kabsch, H C Mannherz, D Suck, E F Pai, K S Holmes *Nature (London)* **347** 37 (1990)
215. M Lorenz, D Popp, K S Holmes *J. Mol. Biol.* **234** 826 (1993)

216. E M Popov *Strukturnaya Organizatsiya Belkov* (Structural Organisation of Proteins) (Moscow: Nauka, 1989)
217. A T Brunger, M Kuriyan, M Karplus *Science* **235** 458 (1987)
218. K C Holmes, M Tirion, D Popp, M Lorenz, W Kabsh, R A Milligan *Adv. Exp. Med. Biol.* **332** 15 (1993)
219. R A Milligan, M Whittaker, D Safer *Nature (London)* **348** 217 (1990)
220. S Lowey, H S Slayter, A G Weeds, H Baker *J. Mol. Biol.* **42** 1 (1969)
221. A G Weeds, S Lowey *J. Mol. Biol.* **61** 701 (1971)
222. S S Margossian, W F Stafford, S Lowey *Biochemistry* **20** 2151 (1981)
223. H E Huxley *J. Mol. Biol.* **7** 281 (1963)
224. B Alberts, D Bray, J Lewis, M Raff, K Roberts, J D Watson *Molecular Biology of the Cell* Vol. 3 (New York: Garland, 1983)
225. H Muller, S V Perry *Biochem. J.* **85** 431 (1962)
226. D Wagner, E Giniger *Nature (London)* **292** 560 (1981)
227. S Citi, J Kendrick-Jones *Bio Essays* **7** 155 (1987)
228. J S Sellers *Curr. Opin. Cell Biol.* **1** 98 (1991)
229. A Elliott, G Offer *J. Mol. Biol.* **123** 505 (1978)
230. D A Winkelmann, H Mekeel, I Rayment *J. Mol. Biol.* **181** 487 (1985)
231. D A Winkelmann, T S Baker, I Rayment *J. Cell Biol.* **114** 701 (1991)
232. I Rayment, W R Rypniewski, K Schmidt-Base, H M Holden *Science* **261** 50 (1993)
233. H W White, I Rayment *Biochemistry* **33** 9859 (1993)
234. I Rayment, H M Holden *TIBS* **19** 129 (1994)
235. I Rayment, H M Holden, M Whittaker, C B Yohn, M Lorenz, K C Holmes, R A Milligan *Science* **261** 58 (1993)
236. F F Flicker, T Wallimann, P Vibert *J. Mol. Biol.* **169** 723 (1983)
237. R G Yount, C R Cremo, J C Grammer, B A Kerwin *Philos. Trans. R. Soc. London, B* **336** 55 (1992)
238. E F Pai, U Krengel, G A Petsko, R S Goody, W Kabsh, A Wittinghofer *EMBO J.* **9** 2351 (1990)
239. C W Muller, G E Schulz *J. Mol. Biol.* **224** 159 (1992)
240. R W Lymn, E W Taylor *Biochemistry* **10** 4617 (1971)
241. M A Geeves *Biochem. J.* **274** 1 (1991)
242. E W Taylor *CRC Crit. Rev. Biochem.* **6** 103 (1979)
243. M G Hibberd, D R Trentham *Annu. Rev. Biophys. Biochem.* **15** 119 (1986)
244. I Wakabayashi, M Tokunaga, I Kohno, Y Sigimoto, T Hamanaka, Y Takesawa, T Wakabayashi, Y Amemiya *Science* **258** 443 (1992)
245. K A Taylor, L A Amos *J. Mol. Biol.* **147** 297 (1981)
246. J A Spudich *Nature (London)* **372** 515 (1994)
247. K M Ruppel, T Q Uyeda, J A Spudich *J. Biol. Chem.* **269** 18773 (1994)
248. J T Finer, R M Simmons, J A Spudich *Nature (London)* **368** 113 (1994)
249. L G Parratt *Rev. Sci. Instrum.* **30** 297 (1959)
250. R Goodwin *Springer Tracts Modern Phys.* **51** 1 1969
251. G Rosenbaum, K C Holmes, J Witz *Nature (London)* **230** 434 (1971)
252. L J Barrington, G Rosenbaum *J. Appl. Crystallogr.* **7** 117 (1974)
253. T D Mokul'skaya, M A Mokul'skii, A A Nikitin *Dokl. Akad. Nauk SSSR* **218** 824 (1974)
254. H W Wyckoff, in *Research Applications of Synchrotron Radiation* (Eds R W Watson, M L Perlman) Publ. BNL No 5038 (New York: Brookhaven National Laboratory, 1973) p. 133
255. S C Harrison, in *Research Applications of Synchrotron Radiation* (Eds R W Watson, M L Perlman) Publ. BNL No 5038 (New York: Brookhaven National Laboratory, 1973) p. 105
256. J C Phillips, A Wlodawer, M M Yevitz, K O Hodgson *Proc. Natl. Acad. Sci. USA* **73** 128 (1976)
257. S J Watowich, J J Skehel, D C Wiley *Structure* **2** 719 (1994)
258. T R Schneider, K S Wilson, in *American Crystallography Association Annual Meeting, Atlanta, GA, 1994* Vol. 22, p. 52
259. M Fulop, R Phizackerley, S M Soltis *Structure* **2** 201 (1994)
260. M Nikkola, Y Lindqvist, G Schneider *J. Mol. Biol.* **238** 387 (1994)
261. T Singer, A Smalas, R Carty, W E Mangel, R M Sweet *Science* **259** 669 (1993)
262. R H Jacobson, X-J Zhang, R F Du Bose, B W Matthews *Nature (London)* **369** 761 (1994)
263. J John, S J Crennell, D W Hough, M J Dason, G L Taylor *Structure* **2** 385 (1994)
264. C G Dealwis, C Frazao, M Badasso *J. Mol. Biol.* **236** 342 (1994)
265. K H G Verschueren, F Seljee, H J Rozeboom, K H Kalk, B W Dijkstra *Nature (London)* **363** 693 (1993)
266. M H Ultsch, W Somers, A A Kossiakoff, A M de Vos *J. Mol. Biol.* **236** 286 (1994)
267. H Lauble, M C Kennedy, H Beinert, C D Stout *J. Mol. Biol.* **237** 437 (1994)
268. Y A Muller, G Schumacher, R Rudolph, G E Schulz *J. Mol. Biol.* **237** 315 (1994)
269. M M Teeter, S M Roc, N H Heo *J. Mol. Biol.* **230** 292 (1993)
270. D Housset, C Habersetzer-Rochat, J-P Aster, J C Fontecilla-Camps *J. Mol. Biol.* **237** 88 (1994)
271. F R Ahmed, D R Rose, S V Evans, M E Pippy, R To *J. Mol. Biol.* **230** 1216 (1993)
272. U Schulze-Gahmen, J M Rini, L A Wilson *J. Mol. Biol.* **234** 1098 (1993)
273. V Ramakrishnan, J T Finch, V Graziano, P L Lee, R M Sweet *Nature (London)* **362** 219 (1993)
274. M Newman, T Strzelecka, L F Dorner, I Schildkraut, A K Aggarwal *Structure* **2** 439 (1994)
275. H Wu, J W Lustbader, Y Lin, R E Canfield, W A Hendrickson *Structure* **2** 545 (1994)
276. J L Smith, E J Zaluzec, J-P Wery, L Niu, R L Switzer, H Salkin, Y Satow *Science* **264** 1427 (1994)
277. J M Thomas *Nature (London)* **364** 478 (1993)
278. S M Franken, H J Rozeboom, K H Kalk, B W Dijkstra *EMBO J.* **10** 1297 (1991)
279. W P Jencks *Chemistry and Enzymology* (Translated into Russian; Moscow: Mir, 1972)
280. A R Fersht *Enzyme Structure and Mechanism* (Translated into Russian; Moscow: Mir, 1980)
281. V K Antonov *Khimiya Proteoliza* (The Chemistry of Proteolysis) (Moscow: Nauka, 1991)
282. R M Stroud, L M Ray, R E Dickerson *J. Mol. Biol.* **83** 185 (1974)
283. J L Chambers, R M Stroud *Acta Crystallogr., Sect. B, Struct. Crystallogr.* **33** 1824 (1977)
284. J L Chambers, R M Stroud *Acta Crystallogr., Sect. B, Struct. Crystallogr.* **35** 1861 (1979)
285. R Huber, W Bode *Acc. Chem. Res.* **11** 114 (1978)
286. H Fehlhammer, W Bode *J. Mol. Biol.* **98** 683 (1978)
287. W Bode, P Schwager *J. Mol. Biol.* **98** 693 (1978)
288. A Ruhlmann, D Kukla, A Schwager, D M Blow *J. Mol. Biol.* **77** 417 (1973)
289. R Huber, D Kukla, W Bode, P Schwager, K Bartels, J Deisenhofer, W Steigemann *J. Mol. Biol.* **89** 73 (1974)
290. R M Sweet, H T Wright, J Jonin, C H Chothia, D M Blow *Biochemistry* **13** 4212 (1974)
291. M Krieger, L M Kay, R M Stroud *J. Mol. Biol.* **83** 209 (1974)
292. G Lang, K Spartalian, T Yonetani *Biochem. Biophys. Acta* **451** 250 (1976)
293. M Chance, L Powers, T L Poulos, B Chance *Biochemistry* **25** 1266 (1986)
294. S L Edwards, N H Xuong, R C Hamlin, J Kraut *Biochemistry* **26** 1503 (1987)
295. H Pelletier, J Kraut *Science* **258** 1748 (1992)
296. I Schlichting, J Berendzen, G N Phillips, R M Sweet *Nature (London)* **371** 808 (1994)
297. T-Y Teng, V Srajer, K Moffat *Nature Struct. Biol.* **1** 701 (1994)
298. G A Petsko *Nature (London)* **371** 740 (1994)
299. G N Phillips *Biophys. J.* **57** 381 (1990)
300. S Dasgupta, T G Spiro *Biochemistry* **25** 5941 (1986)
301. M R Chance, J L Parkhurst, G L Woolery, B Chance *J. Biol. Chem.* **261** 5689 (1986)
302. Q H Gibson, R Regan, R Elber, J S Olson, J E Carver *J. Biol. Chem.* **268** 17908 (1992)
303. K Ruszera, J C Lambry, J-L Martin, M Karplus *Proc. Natl. Acad. Sci. USA* **90** 5805 (1993)
304. D E McRee *Practical Protein Crystallography* (New York: Academic Press, 1993)
305. D W Green, V M Ingram, M F Perutz *Proc. R. Soc. London, A Math. Phys. Sci.* **225** 287 (1954)
306. H Honl *Z. Phys.* **84** 1 (1933)

307. J Karle *Int. J. Quantum Chem. Symp.* **7** 357 (1980)
308. W A Hendrickson *Trans. Am. Crystallogr. Assoc.* **21** 11 (1985)
309. W A Hendrickson, M M Teeter *Nature (London)* **290** 107 (1981)
310. R Kahn, R Fourme, R Bosshard, M Chaimdi, J L Risler, O Didiberg, J P Wery *FEBS Lett.* **179** 133 (1985)
311. S Harada, M Yasui, K Mirikawa, N Kasai, Y Satow *J. Appl. Crystallogr.* **19** 448 (1986)
312. Z R Korzun *J. Mol. Biol.* **196** 413 (1987)
313. W A Hendrickson *Acta Crystallogr., Sect. A, Fund. Crystallogr.* **40** 3 (1984)
314. J L Smith, A Pahler, H M K Murthy, W A Hendrickson *Acta Crystallogr., Sect. A, Fund. Crystallogr.* **43** 10 (1987)
315. W A Hendrickson, J M Troup, N Swepston, G Zdansky *Abstracts of the American Crystallographic Association., Ser. 2* **14** 48 (1986)
316. W A Hendrickson, J L Smith, R Phizackerley, E A Meritt *Proteins* **4** 77 (1988)
317. W A Hendrickson, J R Horton, D M Le Master *EMBO J.* **9** 1665 (1990)
318. D B Cowie, G N Cohen *Biochim. Biophys. Acta* **26** 252 (1957)
319. D M Le Master *J. Virol.* **59** 759 (1986)
320. W A Hendrickson *Science* **254** 51 (1991)
321. Z Otwinowski, in *Isomorphous Replacement and Anomalous Scattering* (Eds W Wolf, R Evans, A G W Lestie) (Daresbury, UK, 1991) p. 80
322. G Bricogne, in *Isomorphous Replacement and Anomalous Scattering* (Eds W Wolf, R Evans, A G W Lestie) (Daresbury, UK, 1991) p. 60
323. M Newman, T Strzelecka, L F Dorner, I Schildkraut, A K Aggarwal *Structure* **2** 439 (1994)
324. Wu Hao, J W Lustbader, Lin Yee, R E Canfield, W A Hendrickson *Structure* **2** 545 (1994)