Translated from *Uspekhi Khimii* **58** 1896–1937 (1989) U.D.C. 577.15.086.83:577.21

Genetic engineering and modern enzymology

G D Kutuzova, E I Dement'eva, N N Ugarova
Moscow State University

ABSTRACT. The role of genetic engineering in the solution of problems involving enzymology is demonstrated. The method for the cloning and mutagenesis of protein and enzyme genes and the possibilities of genetic engineering for the study of the structure and functions of enzymes and their active centres and for the determination of the relation between the structure and stability of protein molecules are examined. The role of directed site-specific mutagenesis in the study of the structure and mechanism of the action of enzymes as well as their stability is indicated. The investigation of the mechanism and factors determining the stability of enzymes with the aid of aminoacid substitutions has led to prospects for protein design — the creation of proteins with new specified properties and improved stability.

The bibliography includes 301 references.

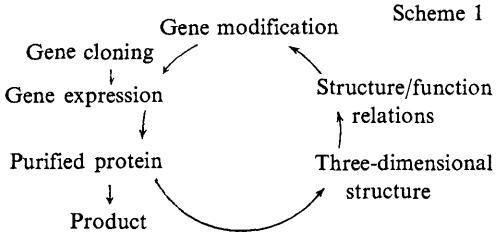
Contents

I. Introduction 1077
II. General principles of the biosynthesis and structural organisation of protein and its gene 1077
III. Methods for the cloning and mutagenesis of the genes of prokaryotic and eukaryotic proteins and enzymes 1082
IV. The study of the structure and functions of proteins by genetic engineering methods 1083
V. Determination of the relation between the structure and stability of protein molecules (by genetic engineering methods) 1090
VI. Protein design 1097

I. Introduction

The extensive development of the recombinant DNA technique has led to its penetration into the protein chemistry and enzymology and to the appearance of a new discipline – protein engineering. The 1980s have been called the "era of the second coming" of protein chemistry and have been associated with a new boom in biotechnology — protein engineering. Protein engineering is a "marriage" between the fundamental science of the structure and functions of protein macromolecules, providing a basis for the deliberate alteration of their properties, and genetic engineering methods which make it possible to achieve this. The complete cycle of protein engineering is described by the following scheme:[1]

Scheme 1



Genetic engineering provides modern enzymology with unlimited possibilities for the investigation of the properties of enzymes: (1) the possibility of the rapid determination of the primary structure of protein molecules and their genes; (2) the discovery of the principles governing the coiling of the polypeptide chain and prospects for the prediction of the secondary and tertiary structures of the protein or the protein domain; (3) the study of the structure and of the mechanism of the functioning of enzymes *in vitro* and *in vivo*; the solution of fundamental problems in enzyme catalysis: the structure of the active centres of enzymes, the mechanism of enzyme–substrate interactions, the types and strengths of the interactions, process energetics, etc.; (4) the determination of the relation between the structure and stability of protein molecules. The solution of these problems leads to splendid prospects for protein design (the synthesis of enzymes with specified properties). The following features can be altered deliberately by genetic engineering methods: catalytic parameters (k_{cat} and K_M); thermal stability and the optimum temperature for activity; stability in organic solvents; substrate specificity; the need for cofactors; the pH optima for activity and stability; resistance to proteolysis; allosteric regulation; molecular weight and subunit composition.

Finally, it has become possible to create completely new proteins by means of the *de novo* synthesis of genes and their subsequent expression.

All these possibilities in genetic engineering for the solution of problems in modern enzymology are in fact considered in the present review.

II. General principles of the biosynthesis and structural organisation of protein and its gene

Protein is a polypeptide sequence made up of 20 aminoacid residues. A definite gene in DNA is responsible for the synthesis of each polypeptide chain.[2,3] The theoretical aspects and mathematical models of protein synthesis have been examined, for example, by Heijne et al.[4] A newly synthesised polypeptide chain in the cell undergoes post-translational modification and compartmentalisation.

The modification of various aminoacid residues (phosphorylation, methylation, hydroxylation, iodination, carboxylation, glycosylation, and substitution of lipids) in the protein chain can occur both during the period when the chain synthesised is linked to a ribosome and after the end of the synthesis.[5] The specific residues undergoing modification are determined in some proteins by the aminoacid sequence and in others by the conformation. For example, the carbohydrate binding sites in proteins are Asn in the Asn–X–Ser/Thr sequence, where X is any residue other than Pro. At this stage, there is a possibility of the binding of the protein synthesised to cofactors and M^{n+} ions, of the covalent and non-covalent assembly of subunits, and of the formation of disulphide linkages between polypeptide chains and within them. In certain cases, these processes take place cotranslationally. Intramolecular disulphide linkages are not encountered in molecules of intracellular cytoplasmic proteins. They are characteristic of extracellular proteins, for which the physiological environment and physicochemical conditions can be varied greatly and can induce the uncoiling of the protein.[6]

For proteins and enzymes functioning outside the cell, a signal sequence (presequence) is split off during passage through the cytoplasmic membrane by membrane signal proteases and the protein is simultaneously glycosylated.

The properties of the side groups in the aminoacid residues constituting the protein, which interact both with one another and with the molecules of the surrounding medium, exert the main influence on the steric structure of the protein molecule. All the aminoacid residues are divided into hydrophobic, hydrophilic, and neutral (Fig. 1). The assignment of a residue to a particular group as a function of criteria varies somewhat (cf. Kutuzova et al.[7] and Dolittle[8]). The hydrophobic residues (Gly), Ala, Cys, Val, Ile, Leu, Met, Phe, and Trp are fully immersed after the coiling of the protein, while the moderately polar residues (Gly), Ser, Thr, His, and Tyr and the highly polar residues (Gly), Pro, Asp, Asn, Glu, Gln, Arg, and Lys are immersed to a lesser extent: 76 and 63% respectively.[9] The structural organisation of the protein and the mechanism of the coiling and assembly of the polypeptide chain have been described in detail in a review.[10]

In the three-dimensional steric structure of a typical globular protein, the inner region is made up mainly of β-layers, while the surface consists of α-helices. Each third or fourth position in these outer α-helices is occupied by aminoacids with non-polar side groups directed into the interior of the globule, while the side groups of the remaining aminoacids are as a rule polar and are directed towards the aqueous medium.

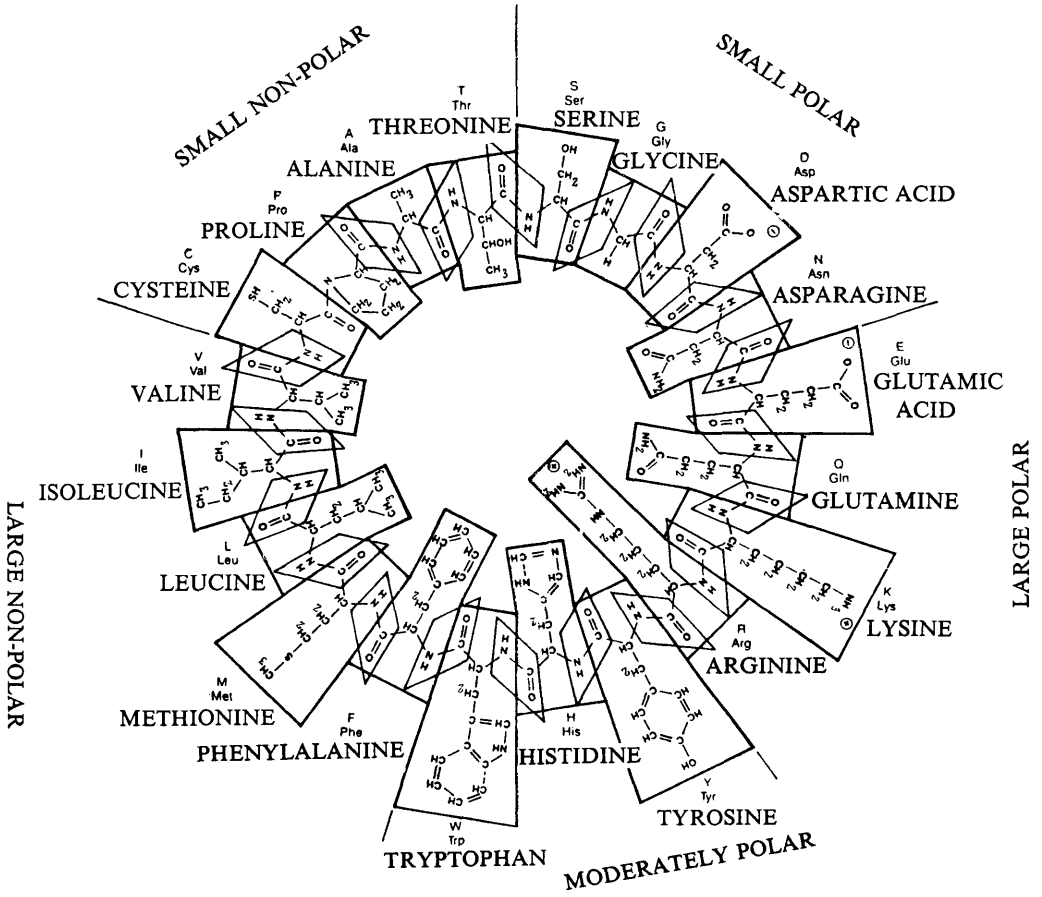


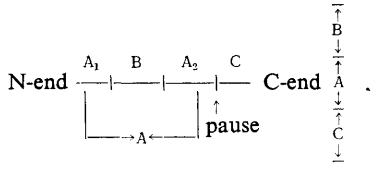
Figure 1. Classes of aminoacids based on the size and degree of polarisation of their side groups.[8]

The native protein molecule consists of several cooperative domains, each of which breaks down on denaturation as a single whole.[11] The protein consists as a rule of combinations of α- and β-structural elements, includes between 50 and 200 aminoacid residues, and is characterised by a minimal surface to volume ratio (S/V). Small proteins can contain only one domain, while larger ones contain up to tens of domains. Subdivision into domains, which is a general principle of the construction of protein molecules, ensures the necessary rigidity and relative mobility of specific groups in the protein and apparently reflects its evolution, which took place via the combination of previously independent small preproteins.[12] The existence of an independent unit of protein structure even smaller than the domain, namely the module,[13] has been postulated at the present time. The module consists of a polypeptide segment whose size is 2–5 kD (30–50 aminoacid residues) and which has a stable coiled conformation and performs a definite function in the highly specific binding of the ligand (substrate, allosteric effector, and cell components: lipids, carbohydrates, nucleic acids, etc.).[14] Thus the size of the protein may be determined by the number of different ligand-binding sections necessary for functioning *in vivo*. Knowing the size of the enzyme subunit, it is possible to estimate the number of ligand-binding sections: the size of the subunit, given by the ratio of the molecular weight to the average size of the module, is ~5.2 kD. The moduli and domains can also be visualised with the aid of crystallographic studies and micro-calorimetric investigation of denaturation processes. Proteolysis usually occurs in the interdomain region and makes it possible to divide domains with partial retention of their function, as in the subdivision of DNA-polymerase into a large and a small domain, each of which is functionally active.[15]

Many proteins consist of several polypeptide chains — subunits, having both identical and different compositions and combined into a single entity via non-covalent interactions or disulphide linkages. The size of the subunit of a eukaryotic protein varies from 10 to 480 kD (the average size is 32 kD, while that of a prokaryotic protein is 24 kD).[16] For 90% of proteins, the size of the subunit is less than 60 kD. Thus the protein consists of subunits, each of which contains domains made up of moduli.

The problem of the coiling of the polypeptide chain synthesised with formation of the three-dimensional steric structure of the protein is one of the most complex and intriguing in protein chemistry. According to modern ideas about the mechanism of the coiling of the protein, the conformational information is distributed along the polypeptide sequence non-uniformly, certain aminoacids being of fundamental importance for the correct conformation. At the beginning of the coiling process, local reversible structures (nucleation centres) are produced. These are short sections of the secondary structure of the type of α-helices or β-layers. The structure is subsequently stabilised as a result of the interaction of neighbouring sections. Coiling begins independently and more or less simultaneously in several regions of the polypeptide chain (autonomous coiling domains).[17] The problem of coiling has now acquired especially great importance in connection with the preparation of hybrid proteins in the cloning of eukaryotic genes in bacteria, where the *N*-terminal section of the required protein has been fused with part of the prokaryotic protein, which can influence the coiling process, destabilising the kinetically important intermediate states of the eukaryotic protein. Furthermore, its *N*- or *C*-terminal component can be within the globule, while the fused hybrid protein will not have the required native conformation, giving rise to insoluble derivatives.[17]

The fact that the aminoacid residues in the polypeptide chain are separated not only in space but also in time during synthesis is very important for the protein coiling process. The importance of the separation in space and time of the aminoacid residues for the structure of the whole depends on the relative rates of synthesis of the protein and of its coiling. At the beginning of the synthesis, the coiling of the segment of the polypeptide chain synthesised takes place in microseconds, since it involves the interaction of only the nearest aminoacid residues. On average, the time required for the complete translation of a protein of average size is 2 min. The coiling of the polypeptide chain to a three-dimensional structure takes place before the completion of its total synthesis and there is a possibility of the formation of disulphide linkages, which did not exist in the completed protein. The subsequent isomerisation of disulphide linkages leads to the correct conformation, which has been demonstrated for the pancreatic trypsin inhibitor.[18] Direct evidence has been obtained for the existence of breaks during the synthesis of the protein, which results in the accumulation of specific polypeptides.[19,20] In terms of the mechanism of the coiling of the polypeptide chain *in vivo*, proteins can be divided into two groups:[21] (1) proteins the correct coiling of which is unaffected by the rate of translation; in such proteins the aminoacid sequence does indeed contain all the information necessary for the correct coiling of the chain (even *in vitro* after denaturation); (2) proteins incapable of coiling effectively into the native conformation *in vitro*. These are polypeptides with alternative kinetically easy coiling pathways, which can form distorted but stable conformations. The coiling of such proteins *in vivo* is greatly influenced by the translation process itself, which takes place at different rates and with translational pauses. The presence of the pauses is determined by the accumulation in the DNA chain of rarely used codons and hence by the presence of small amounts of the corresponding tRNA, which retards significantly the supply of the corresponding aminoacids. The presence of a secondary structure in mRNA also reduces the rate of translation.[21] The above features have been discovered, for example, for yeast pyruvate kinase.[22] Each of the four subunits of pyruvate kinase consists of a short N-terminal segment and three different domains: A, B, and C:



The translational pause before the start of the synthesis of the domain C enables the two sections A_1 and A_2 to form the central domain A, containing the catalytic centre, before the synthesis of the domain C containing the allosteric centre of pyruvate kinase. This prevents anomalous interactions between the still uncoiled regions of the domains A and C. Translational pauses, which enable functionally important domains to be formed, have also been observed for the Arom-protein catalysing the biosynthesis of aromatic acids.[23] The inability of certain proteins to return to the native conformation after denaturation *in vitro* may indicate the importance of the separation in time of the processes involving the synthesis and coiling of different sections of the polypeptide chain *in vivo*.

As a result of the employment of genetic engineering methods, it has now been possible to investigate in detail the mechanism of the coiling of proteins. The construction of mutant polypeptide

segments elucidates the role of individual aminoacid residues in the coiling and assembly of enzyme subunits.[24,25] It has been shown for the heterodimer (αβ) of *bacterial luciferase* that the scission of 10–15 residues from the C-terminal section of the β-subunit does not alter the activity and stability of the enzyme, but the mutant enzyme is incapable of recoiling after treatment with urea. The αβ complex of the subunits is formed initially between incompletely coiled α and β subunits and their final coiling takes place in the heterodimeric form. If the α and β subunits have been coiled *in vivo* separately, then the formation of the complex αβ is impossible without a preliminary treatment with urea. The coiling process is unaffected by the nature of the residues in the C-terminal section, but deletions in this region reduce the ability of the β-subunit to form a dimeric structure *in vitro* and *in vivo*.[26] Even one aminoacid residue can play a key role in the formation of the structure of the active enzyme. The introduction of an additional Met group at the N-end reduced significantly the rate of coiling of chick lysozyme, while the replacement of Ala-31 even by the structurally similar Val prevented the correct coiling of the protein.[27] In the pancreatic trypsin inhibitor, the Cys residues forming one of the three S–S bonds were replaced by Ala or Thr, whereupon the rate of coiling diminished by a factor of 3 but the correctly coiled protein was obtained.[28] Thus the formation of S–S bonds reduces the activation barrier to the polypeptide chain coiling process.

For enzymes with a known three-dimensional steric structure and gene structure, important conclusions have been reached by comparing these structures.[29] It is known that the genes of eukaryotic proteins are not continuous but are made up of coding sections (exons) and non-coding sections (introns). It has been suggested that the position of the introns in the gene corresponds to the boundary of the domain in the protein. For example, the subunits of NAD-dependent dehydrogenases contain two structurally and functionally different domains. One of them binds the coenzyme, while on the other there are substrate-binding and catalytic sections. The NAD^+-binding domains of alcohol dehydrogenase (ADH), glyceraldehyde-3-phosphate dehydrogenase (GAPD), and lactate dehydrogenase (LDH) have similar structures and bind the coenzyme in a similar way, while there is no homology as regards the catalytic domains and the latter have completely different structures.[30] The structure of the domain-containing subunits and the positions of the introns for these enzymes are presented in Fig. 2. The introns 6 and 9 for ADH, 5 and 6 for GAPD, and 3 and 7 for pyruvate kinase (PK) divide the catalytic and nucleotide-binding domains. Each nucleotide-binding domain can be roughly divided into two identical units, which refer to the mononucleotide-binding region and are separated by introns. This confirms the hypothesis that[31] the mononucleotide-binding domain is the principal unit for the generation of enzymes whose function is determined by the ability to bind nucleotides.

However, it has been shown for haemoglobin,[32] carboxypeptidase,[33] and antitrypsin[34] that the positions of the introns in the genes of eukaryotic proteins do not always correspond to the boundaries between the known domains. Furthermore, introns have not been observed in the genes of prokaryotes. It has been suggested that they have been eliminated in order to economise the genetic material.[8] On the other hand, several introns can be encountered within one domain (Fig. 2). Analysis of the genes of 80 proteins[35] has shown that for 70% of them the size of the exons

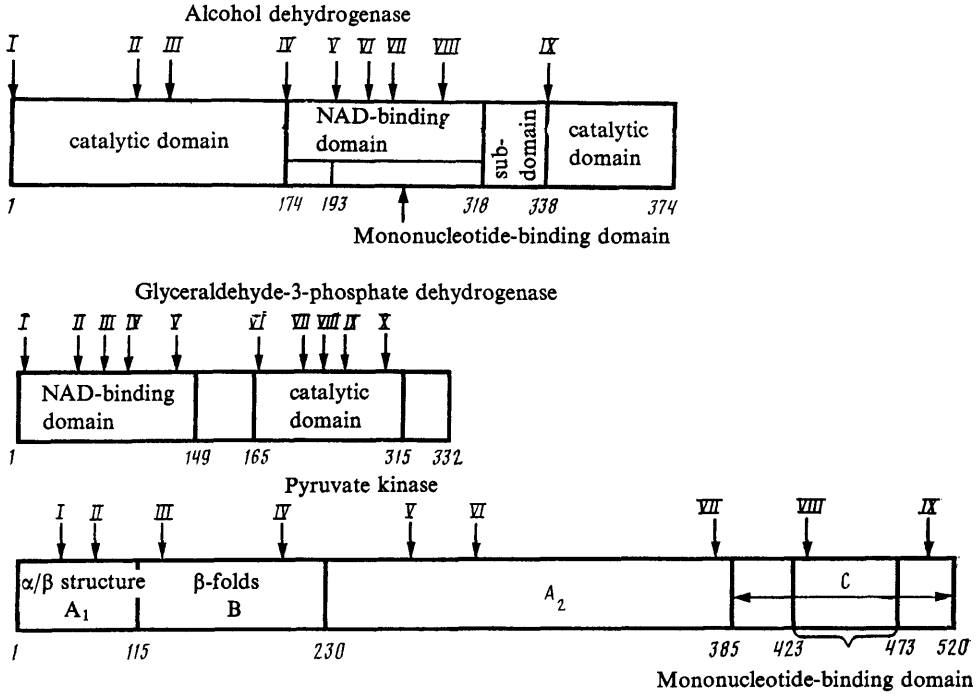


Figure 2. Diagram illustrating the positions of introns (roman numerals) and the limits of the domains along the polypeptide chains of enzyme subunits;[29] the arabic numerals indicate the positions of the residues in the aminoacid sequence.

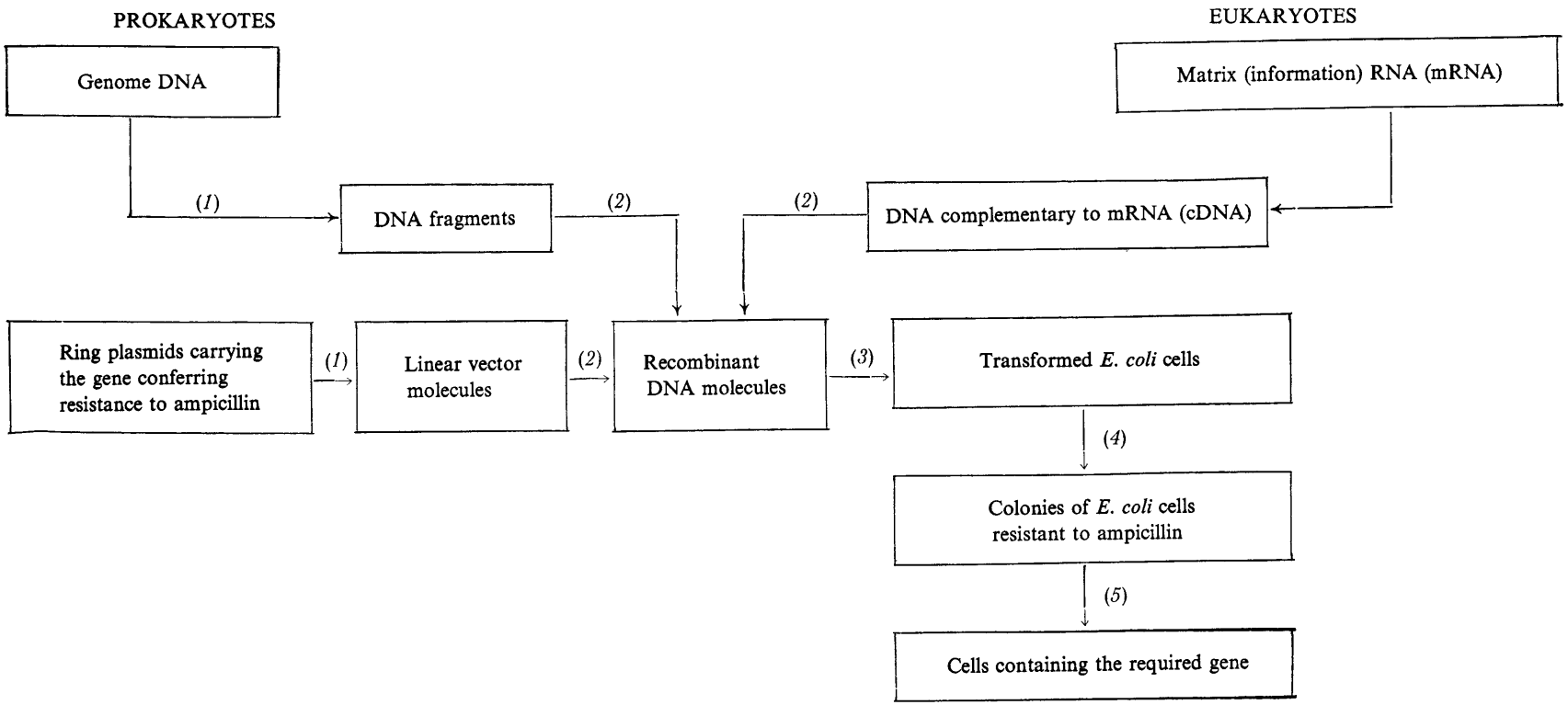


Figure 3. The principal stages in the cloning of the prokaryotic and eukaryotic DNA in the cells of *E. coli*:[3,37,38] (1) cleavage of DNA with the aid of restrictase; (2) cross-linking of DNA molecules with the aid of DNA ligase; (3) transformation of the cells of *E. coli* by the recombinant plasmids; (4) selection of *E. coli* colonies containing the recombinant plasmids on the basis of resistance to ampicillin; (5) screening of the library of clones with the aid of a radioactive probe.

varies from 52 to 223 pairs of nucleotides, which corresponds to the sizes of the moduli in the protein.[13] Introns apparently divide not only domains but also the protein moduli with definite ligand-binding functions. An algorithm has been proposed for the determination of the boundaries of the moduli and domains.[36] Many experimental facts confirm this hypothesis.

III. Methods for the cloning and mutagenesis of the genes of prokaryotic and eukaryotic proteins and enzymes

The methods for the cloning of the genes of proteins and enzymes from different sources, for the expression of genes in different host systems, have been described in detail in many textbooks and manuals.[3,37–39] The cloning of the DNA of the genes of prokaryotic proteins differs fundamentally from the cloning of the genes of the proteins from eukaryotic organisms. The DNA of the simplest organism consists of a set of genes coding for proteins and enzymes and arranged in a sequence. The first stage in its cloning is the isolation of the genome DNA, its splitting into fragments, and their incorporation into DNA vectors. The overall stages in the cloning of DNA in the plasmid vector of *E. coli* are presented in Fig. 3. The cells of *E. coli* are transformed by annular recombinant DNA molecules. The cells are then sewn in such a way that one colony, containing a population of identical daughter molecules (clones), grows from each cell. Among several thousand clones (the library), it is possible to find one or several clones containing the required incorporated DNA fragment. The colonies are screened by means of a specific radioactive probe, which makes it possible to inspect the library of clones. The probe can be pure or partly purified mRNA or a chemically synthesised oligonucleotide, which are hybridised with the corresponding gene sequence in the genome library. Antibodies to proteins, whose genes have to be detected, are also used frequently to identify the colonies.

Since the genes of the proteins of eukaryotes include non-coding sections (introns), one uses for the cloning of eukaryotic proteins cDNA — a DNA complementary to the matrix (information) RNA — or a chemically synthesised gene without introns. The first stage in the cloning of the proteins of eukaryotes is the isolation of mRNA on the basis of which the double-stranded cDNA is synthesised. Next it is incorporated in the cloning vector and is used to transform the host cells (Fig. 3).

The genetic engineering methods used for mutagenesis can be divided into two large classes. The first class includes the formation of random point mutations in the required gene. The classical method of mutagenesis with selection of mutant forms on the basis of a preselected property has been widely used hitherto for the formation of mutant enzymes differing from the wild enzymes in their catalytic activity, substrate specificity, or stability. A mutant gene is also obtained when the DNA in the form of an isolated fragment or as a component of a single-stranded or double-stranded DNA-vector is modified with the aid of chemical reagents or enzymes. In the second stage, the modified DNA is used to transform bacterial cells and individual clones are isolated. Next the mutant recombinant DNA is isolated from the set of clones and the nucleotide sequence in the mutant gene is determined. The characterised mutant gene is used to transform the host cells (bacteria, yeasts, and more rarely animal or plant cells), in which it is transcribed and translated, giving rise to the mutant protein.[40,41] The second class includes all the methods which make it possible to achieve specific preselected mutations. The modern methods for mutagenesis are classified in Table 1.

The most accurate and convenient method for the introduction of mutations into cloned DNA fragments is site-specific mutagenesis using synthetic oligonucleotides. A large number of studies have been devoted to the improvement of this method,[50–58] in particular to the increase of the yield of the mutant forms.[55,59–63] Site-specific mutagenesis has been used widely in the study of the relation between the structure and functions of proteins and their enzymes and for the creation of proteins with new properties.

Table 1. Classification of the methods for mutagenesis.

Name	Essential features of method	Remarks	Refs.
	Random mutations		
Chemical methods	Treatment of the gene by chemical mutagenic reagents (sodium bisulphite, hydrazine, nitrous acid, etc.)		[40–46]
Enzymic methods	Preparation of single-stranded DNA, construction of the second DNA chain with the aid of DNA polymerase in the presence of 4 deoxynucleotides, one of which has been replaced by its analogue	The bases of the nucleotides are modified, which leads to mutation	
	Specific preselected mutations		
Gene synthesis. Total synthesis of the gene	Construction of the gene from a set of segments of chemically synthesised oligonucleotides	The frequency of mutations approaches 100%, but the method is laborious and is nowadays rarely used	[42, 47, 48]
Partial (cassette) gene synthesis	Isolation of the gene, excision of a DNA segment at the restriction sites, and its replacement by the synthetic oligonucleotide containing the mutation	—	[40, 42, 49]
Deletion and insertion on cleavage of double-stranded DNA	Preparation of a linear DNA molecule, the removal of fragments located between restriction sites (deletions), the construction of an additional nucleotide chain (insertions)	—	[38, 42]
Site-specific oligonucleotide-directed mutagenesis	The synthetic oligonucleotide (10–20 bases) containing the mutation is hybridised with single-stranded DNA and is used as a primer for the synthesis of the second DNA chain. The additional chain is closed with the aid of DNA ligase and the resulting DNA is used to transform the *E. coli* cells, where DNA replicates with formation of two forms of the cloned gene: the wild and mutant types (Fig. 4)	The mutant form can be readily identified using the same nucleotide labelled with ^{32}P	[40–42, 47, 50–55]

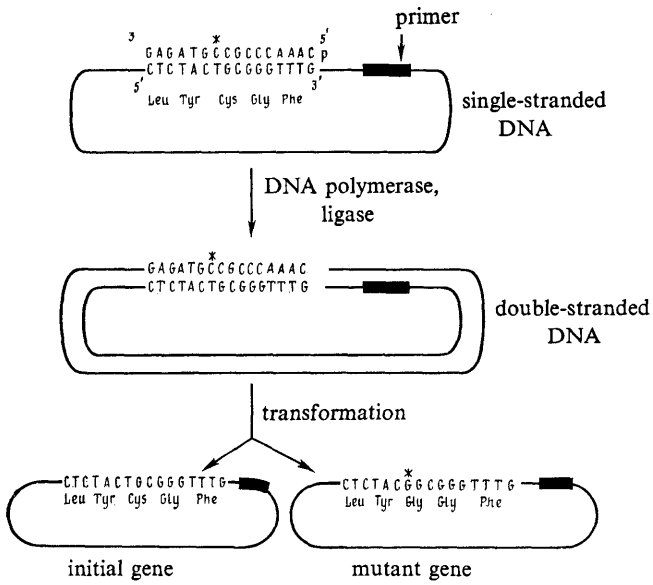


Figure 4. Schematic representation of the nucleotide-directed mutagenesis;[55] the asterisk designates the mutant nucleotide.

IV. The study of the structures and functions of proteins by genetic engineering methods

1. The primary protein structure and its evolution

The primary protein structure was determined previously solely by the sequence determination method with the aid of automatic aminoacid analysers. It was then necessary to have at ones disposal a large amount of the very pure protein, which was not always possible. With the aid of genetic engineering methods, it has become possible to isolate DNA and to determine the total nucleotide sequence in the gene coding for the given protein. The methods for the determination of the DNA sequence (the Maxan–Gilbert and Sanger methods)[3,37,38,64,65] made it possible to determine relatively simply and rapidly the sequence of nucleotide bases in the gene. With the aid of the genetic code, it is easy to establish subsequently the aminoacid sequence in the protein. The difficulty is the necessity for the error-free identification of a sequence of an enormous number of bases (a protein consisting of a thousand aminoacids is coded by 3000 bases) and also for a correct determination of the starting point of the reading of the code. The deletion of one base can lead to a shift of the reading framework and hence to an incorrect determination of the entire aminoacid sequence. In order to avoid such errors, one can, for example, determine the aminoacid sequences in short sections at the N- and C-ends of the protein by another method and to compare them with the sequence obtained from the nucleotide sequence of the gene. In order to check the correctness of the reading of the DNA base sequence, the molecular weights, measured mass-spectrometrically, of the tryptic peptides obtained from the test protein were also compared with the structures of these peptides derived from the DNA sequence.[66]

At the present time, the primary structures of more than 3700 proteins determined by different methods are known.[67] As genetic engineering methods have developed, the number of proteins and enzymes whose primary structures have been established has increased greatly (Fig. 5). In relation to certain enzymes, we shall consider what information about the structure of the enzyme itself and its active centre can be obtained solely from the analysis of the nucleotide sequence in the gene and by comparing it with the aminoacid sequence of the test protein.

The cloning and sequence determination for genes have made it possible to establish the presence of signal peptides in the proteins secreted: cytochrome c_3 from *Desulfovibrio vulgaris*,[68] protease III[69] and penicillin G-acylase[70] from *E. coli*, and chloroperoxidase.[71] The signal peptides promote the transmembrane protein transfer from the periplasm of bacteria and yeast and they split off immediately after secretion, so that the determination of nucleotide sequence in genes is a good method for the detection of signal peptides and for the investigation of their structure. Comparison of the aminoacid sequence of the protein and the nucleotide sequence of the gene shows whether the given protein undergoes post-translation modification. For example, aquarin is synthesised in the form of a precursor. During translation, seven *N*-terminal residues are split off.[72] The *N*-terminal residue in human creatine kinase M is lysine and the methionine residue is split off after translation.[73]

The determination of the aminoacid sequence makes it possible to reach certain conclusions concerning the presence of domains and the structure of the substrate-binding sections. Analysis of the structure of the penicillin acylase gene revealed the presence of four structural domains on the basis of the positions of introns: nucleotides 1–78 code for the signal peptide, nucleotides 79–705 code for the α-subunit, nucleotides 708–867 code for the peptide binding covalently two subunits, and nucleotides 868–2538 code for the β-subunit.[70] A multidomain structure has also been established for the heavy chain of calpactin (the substrate of viral protein-tyrosine kinase)[74] and calpastatin (inhibitor of the Ca^{2+}-dependent calpain).[75] The study of the structures of the genes of diglyceride synthetase and diglyceride hydrolase from *E. coli* showed that the former enzyme is an extremely hydrophobic membrane protein.[76] The latter enzyme is less hydrophobic, but

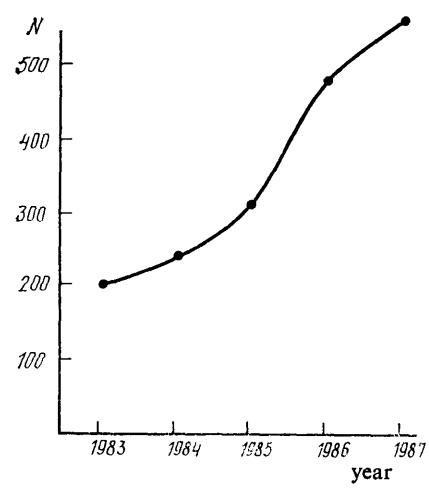


Figure 5. Number of publications (annually) on the determination of the aminoacid sequences of enzymes by genetic engineering methods (counted by the authors).

there is a very hydrophobic domain at its *N*-end, which binds the hydrolase to the membrane.[77] The composition of the bioluminescent Ca^{2+}-dependent protein aequorin incorporates three hydrophobic sections, which may be responsible for the binding of its functional chromophore, i.e. coelenterazine (luciferin), and three sections responsible for the binding of Ca^{2+} ions.[72] Analysis of the primary structure of Na^+, K^+-ATPase led to a hypothesis about the character of the disposition of this enzyme on the membrane.[78]

Comparison of the primary structure of the test enzyme with that of proteins having similar functions makes it possible to discover common structural sections. Analysis of the aminoacid sequences of aequorin and other Ca^{2+}-binding proteins (calmodulin, parvalbumin, and troponin C) revealed three Ca^{2+}-binding sections in the polypeptide chain of aequorin.[72] Important information about the structure of the active centre of bacterial luciferase has been obtained by comparing the primary structures of luciferases from different sources: *Vibrio fischeri*, *Vibrio harveyi*, *Photobacterium leiognathi*, and *P. phosphoreum*.[79] Data for the modification of the SH group of cysteine-106 in the luciferase from *V. harveyi* led to the hypothesis that this group participates in catalysis. However, the luciferases from other sources contain valine in the position 106. Thus the SH group of Cys-106 is not important for the manifestation of the activity of bacterial luciferase. The sections 115–117 and 280–282 of the α-subunit, which are the same in all the bacterial luciferases examined and are located close to one another in three-dimensional space, apparently constitute an appreciable proportion of the active centre.[39] Comparison of the primary structures of lipases from various sources (bacteria and pig pancreas) showed that Ser-83 is incorporated in the active centre.[80]

Similar aminoacid sequences have been noted for glutathione S-transferases from human and rat livers (50% homology),[81] catalase T from the yeast *S. cerevisiae*, the bacterium *Penicillinum vitale*, and ox liver (52.2% homology),[82] aromatic aminoacid hydroxylases — tryptophan, phenyl, and tyrosine hydroxylases (60% homology),[83] valyl-tRNA synthetase from *B. stearothermophilus* and isoleucyl tRNA synthetase from *E. coli*,[84] and for five adenylate kinases from various sources.[85] This means that the data for the steric structure and mechanism of action of one enzyme can be used in the study of an enzyme with a homologous sequence. The structure of the biopterin-dependent aromatic acid hydroxylases is homologous for approximately 2/3 of the molecule, starting from the C-end whereas at the N-end no homology has been observed. This confirms the hypothesis that the C-terminal domains in hydroxylases determine their catalytic properties (similar mechanism of catalysis), while 1/3 of the molecule, starting from the N-end, is made up of domains determining the substrate specificity.[83] The relation between the structure and function of human M creatine kinase,[73] Na^+, K^+-ATPase,[86] the heavy chain of calpactin,[74] glycogen phosphorylase,[87] glutathione S-transferase,[81] yeast catalase T,[82] and tryptophan hydroxylase[88] has been discussed on the basis of the homology established with other proteins.

It has become possible in recent years to investigate the evolution of protein in connection with the development of the recombinant DNA technique, the establishment of the primary structures of an enormous number of proteins, and advances in crystallography. Common domains have been discovered for a series of proteins with known crystal structures. Consequently, if any one structural element proves useful for the performance of a particular function, it is invariably used in different proteins. For example, Rossman[30] observed in a series of enzymes an invariant domain consisting of approximately 70 aminoacid residues. These enzymes had markedly different functions but had a property in common: they bound coenzymes of the type of NAD, FMN, or AMP, which had a common component — a mononucleotide. This domain is apparently a proprotein for a large number of nucleotide-binding enzymes, because its function proved to be vitally important. The first proteins capable of performing functions of any kind were small in size and had the property of being able to bind specifically other molecules. When two such proteins were combined, a primitive catalytic system might have arisen improving and developing in the course of evolution. From Gilbert's point of view,[89] exons are genome equivalents of protein domains which have been used to assemble the first proteins and which have been maintained throughout evolution in a genetically isolated state in the course of several billion years. The non-intersecting processes involving the transposition and reorganisation of domains into new combinations generated new proteins which included also useful constructions. The essential feature of the hypothesis is that the evolution of proteins was based on the sorting of exons coding for the domains.[8]

The principal mechanism of the evolution of proteins is the duplication of genes, which consists in the fact that two or more copies of the same gene arise in the cell genome. One of the copies retains the previous function, so that no changes take place in the organism, while the other copy can undergo mutations. The duplication of genes has been demonstrated by the existence of subunit enzymes with non-identical but homologous subunits (homologous proteins), containing one polypeptide chain, as in proteases of the serine–histidine–asparagine type with different substrate specificities.

Another example of gene duplication is provided by trypsin, chymotrypsin A, and elastase. Their sequences are approximately 40% homologous. They have a similar function, namely the hydrolysis of proteins, but there are differences in the type of bond hydrolysed. The residues in the active centre were different: Asp in trypsin and Ser in chymotrypsin. Analysis of the three-dimensional structures of these enzymes showed that serine proteases consist of two similarly coiled domains arranged symmetrically around a twofold axis.[90] In each case introns separate the codons of the catalytic triads (Asp, Ser, and His), but the main part of the exons is the same within this class of enzymes.

In the cAMP-, cGMP-, and Ca^{2+}-calmodulin-dependent protein kinases and phosphoryl azokinase, the domains fulfilling a similar function are homologous: for example, the cAMP- and cGMP-binding regions in the regulatory subunits are in homologous domains. The catalytic (ATP-binding) regions in all protein kinases are also encountered in homologous domains.[91] These proteins arose as a result of the duplication and fusion of the genetic elements of two related families, which determined the allosteric nature of these enzymes.[92] It may be that allosteric regulation arose initially from the random interaction of the active centre with inhibitor and the association was weakened by the covalent modification of the complex, for example, by phosphorylation, adenylation, etc.[88]

A model involving triple duplication of a precursor gene with formation of the gene of the prototype of modern calmodulin and the addition of non-coding segments to both ends of the gene has been proposed for the evolution of the calmodulin gene.[93] The last stage occurred relatively recently, since the size and nucleotide sequences of the segments of the 3′- and 5′-ends differ greatly for genes of different calmodulins.

Apart from the duplication of the genes, the latter can also fuse, as a result of which the protein becomes twice as long. The immunoglobulins, which arose from a common precursor—a polypeptide consisting of 110 aminoacid residues—may serve as an example. After duplication and merging, the different sections of the immunoglobulin chains acquired different biological functions.[2]

The evolution of proteins can be studied in terms of two aspects.

(1) Analysis of the changes in the structure of the same protein in different organisms during evolution. For example, comparison of the structures of cytochrome *c* derived from organisms ranging from yeast to man (more than 80 species) shows that closely related cytochromes hardly differ or differ in a small number of residues. The greatest differences in the aminoacid composition are characteristic of cytochromes *c* from the lowest and highest organisms. Conservative substitutions in which an aminoacid in the sequence is replaced by another aminoacid with similar properties and structure, so that the steric structure of the protein is not affected and there is no substitution in the region of the active centre and no substitution inducing pronounced conformational changes in the protein which can lead to the loss of activity, are characteristic of proteins from closely related species. There is an appreciable tolerance as regards substitutions, deletions, and insertions of aminoacids in many sections of the polypeptide chain in different proteins.[2] A homology has been observed near Cys residues among tryptophan, phenyl, and tyrosine hydroxylases from different sources. It has been suggested that they have arisen from a single gene as a result of the accumulation of point mutations.[83] On passing from the lowest to the highest organisms, the protein becomes more complex. Bacterial phosphofructokinase has a single regulatory section, while the human enzyme is larger and contains three regulatory sections and can bind in addition three different types of regulatory proteins.[13] In the limiting case, it may turn out that the aminoacid sequences of two proteins having similar steric structures differ so much that it is difficult to establish their evolutionary relation.

(2) Comparison of different proteins within the framework of a single species. A homology of the primary structure is frequently characteristic of enzymes catalysing similar reactions. For example, the homology reaches 40% for glutathione reductase and lipoamide reductase. On comparing the aminoacid sequences of any proteins, one may expect the coincidence to the extent of at least 5% on statistical grounds.

In comparing primary structures, it is essential to take into account the possibility of deletion and insertion. Cases where between 15 and 25% of coincidences are observed when account is taken of such deletions and insertions are of greatest interest. The roots of the family tree of proteins should be sought in this "intermediate" zone. For example, five proteins with different biological functions have been found to be related: antithrombin III (a protein present in blood plasma and neutralising the effect of thrombin), α_1-antitrypsin (a proteinase inhibitor), ovalbumin (a component of egg protein), angiotensin—(a precursor of the peptide hormone involved in the regulation of arterial pressure), and a barley grain protein with an unknown function.[8]

Structural similarity in the absence of homology in the sequence is exhibited by the chicken protein lysozyme and the bacteriophage T4 lysozyme, which suggests the presence of a common precursor.[94]

Studies of the evolution of proteins have given a powerful stimulus to the development of protein engineering at the level of DNA, because rearrangements, deletions, or insertions of exons can give rise to new combinations of domains and moduli, i.e. to proteins with completely new properties.

2. Prediction of the steric structure and functions of protein molecules

Comparison of the aminoacid sequences constitutes the basis for the modelling of protein at all the levels of the protein structure—the secondary structure, the structure of domains, the quaternary structure, the interaction with ligands, and the prediction of biological functions of the protein.[95–110] The approaches used to predict the protein structure on the basis of data for its primary structure can be represented schematically (Table 2).

The methods and criteria for the comparison of different proteins and for the modelling of the protein structure have been examined in a number of studies.[96,98–106]

The sequences in proteins are compared in order to discover homologous sections by means of standard programs for the analysis of aminoacid sequences and computer graphics. Special data banks of primary aminoacid sequences, for example, the data bank of The National Biomedical Research Foundation (NBRF) in the USA, have been created.

Table 2. Prediction of the structure of the protein on the basis of data for its primary structure.

Protein structure	Set of prediction methods
Aminoacid sequence	Prediction of the structures of individual section of proteins, determination of "structural motifs" (α- or β-helices, random coil, etc.)
↓ Secondary structure	The search for "structural" matrices, modelling of the disposition of structural motifs in space
↓ "Supersecondary" structure	Simplified energy calculations, minimisation of the energy, and molecular dynamics
↓ Domain	More detailed energy calculations, modelling of the disposition of the domains and subunits in space with the aid of computer graphics
↓ Set of domains, set of subunits	

There are several ways of predicting the secondary and tertiary protein structures. The first, simplest and most widely used method involves the finding of homologous proteins with a known steric structure. In order to construct a three-dimensional model of the test protein on the basis of structural homology with known proteins, special programs have been developed, for example the Frodo program[107] created by Jones as early as 1978 and the Composer program.[103] The latter is based on the simultaneous superposition of the structures of different proteins homologous with the modelled protein and on the derivation of an average structure or templet.

The following criteria serve as a basis for the comparison of the aminoacid sequences and the discovery of analogies between them: the identity of the aminoacid sequences, the identity of the nucleotide sequences, a minimal difference between the bases in the codon, the ratio of the number of modified aminoacids to the total number, and certain characteristics of the aminoacid residues, for example, their polarity, hydrophobicity, etc. Argos[100] developed a method for discovering weak analogies in protein sequences and established two sensitive criteria for their comparison: (1) the

relative molecular weight of the aminoacids replaced (the Dayhoff criterion); (2) comparison of five physical characteristics of aminoacid residues such as hydrophobicity, volume, refractive index, the location of the residue on sections corresponding to the rotation of the chain or sections anti-parallel to the chain. The author examined 30 classes of proteins, including haemoglobins, chymotrypsins, lysozymes, cytochromes *b* and *c*, triosephosphate isomerase, lactate dehydrogenase, the domains of different proteins fulfilling the same function — Ca-binding, arabinose-binding, etc. All these proteins have similar tertiary structures. If homologies with other proteins are not observed for the protein investigated, then some information about its secondary and tertiary structure can be obtained by analysing the structure of proteins close to it on the evolutionary scale. In the absence of a homology, a series of other methods are also used to establish secondary and tertiary structures. For example, the construction of a hydrophobicity profile and its comparison with the hydrophobicity profiles of other proteins permit the identification of sections forming the secondary structure. The structural profiles have been investigated for a series of proteins with different degrees of homology of the primary structures and it has been shown that identical structural profiles can imply identical three-dimensional structures even in the absence of a homology in the aminoacid sequences.[108,109] In the absence of homology, there are also other methods for predicting the tertiary structure of proteins, for example, calculations of the energy states of the protein molecule and the search for the conformation having the minimum free energy in the protein–solvent system, and also with the aid of templets: in proteins with different primary structures there is a possibility of the existence of regions which are structurally similar; thus the nucleotide-binding domain in certain proteins consists of parallel β-folds bounded by α-helices.[106]

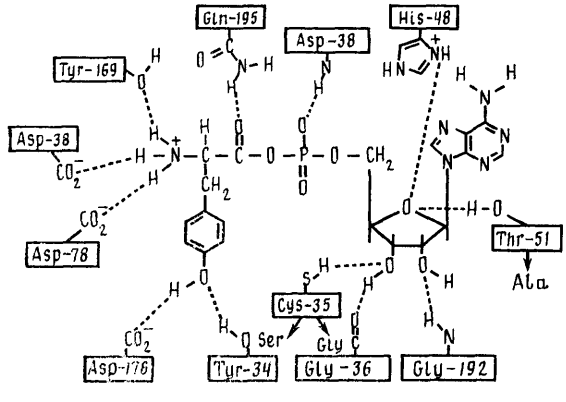


Figure 6. Schematic illustration of the active centre of tyrosyl-tRNA synthetase.[55]

Finally the primary structure permits the assignment of the protein to a particular functional class.[110] Klein analysed 2145 aminoacid sequences in the NBRF data bank and divided them into 26 functional classes. Physicochemical characteristics of the protein such as average hydrophobicity, maximum hydrophobicity, variation of the hydrophobicity of individual segments, charge, the probability of the formation of the α- or β-helix, and the periodicity in the appearance of glycine, phenylalanine, and the pentapeptide Phe-His-Phe-His-Trp were used in the analysis. It was found that 3–5 characteristics of this kind are sufficient to assign the protein with a high degree of probability to one of the 26 classes. These classes include collagen, nitrogenases, cytochromes (*b*, c_3, c^1, and *c* separately), carboanhydrases, ribonucleases, ATPases, ceratins, phospholipases, hormones, dehydrogenases, the class of prolactin and somatotropin, ferredoxins, toxins, histones, globins, the variable sections of antibodies, and the permanent sections of antibodies. A periodicity in the appearance of glycine in the primary structure is characteristic of the class of collagens, a periodicity in the hydrophobic regions is characteristic of globins, while the characteristic feature of ATPases is a low tendency towards the formation of β-folds. A characteristic feature of the variable sections of antibodies is a marked difference between the hydrophobicities of individual segments of the polypeptide chain. The knowledge of the primary and steric structures of the protein or its modelling by comparison with previously investigated proteins and data for the chemical modification of proteins and enzymes and for the activity of the enzyme centres, obtained by different methods, make it possible to proceed to a new level in the study of the active centres of enzymes with the aid of genetic engineering methods — site-specific mutagenesis.

3. The study of the active centres of enzymes with the aid of site-specific mutagenesis

Studies on proteins by the method of site-specific mutagenesis can be divided arbitrarily into two groups: the site-specific mutagenesis of enzymes with a known crystal structure and the site-specific mutagenesis of enzymes whose steric structure has not been determined. In the first case, the steric model of the enzyme serves as a basis for the specific alteration of the structure of the protein molecule and for the prediction of the results of mutation-induced changes with the aid of computer graphics. If the crystal structure of the protein is unknown, chemical modification data and comparison of the structures and properties of homologous species and species close on the evolutionary scale are used as a basis for the assessment of the mutation. In this case, the results of mutations are usually random and it is much more difficult to achieve the required alteration of the catalytic properties or stability of the test protein.

(a) Site-specific mutagenesis of enzymes with a known steric structure

Tyrosyl-tRNA synthetase[55,111–128] (from *Bacillus stearothermophilus*) catalyses the two-stage $tRNA^{Tyr}$ aminoacylation reaction:

Scheme 2

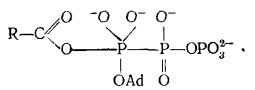
$$E + Tyr + ATP = E \cdot Tyr—AMP + PP_i \ (a),$$

$$E.Tyr - AMP = tPHK^{Tyr} = Tyr - tPHK + AMP + E \ (b)$$

Analysis of the steric structure of the complex E.Tyr–AMP showed that a hydrogen bond may exist between the SH group of Cys-35 and the 3′-OH group of the ribose ring of Tyr-AMP (Fig. 6). The replacement of Cys-35 by Ser and Gly by the method of site-specific mutagenesis led to the impairment of the binding to the ATP and a decrease in k_{cat}. This shows that the strengthening of the hydrogen bond impairs the affinity of the enzyme for the substrate and that its elimination can improve the binding of ATP.[55,111,112] The Thr-51 residue is capable of forming a weak hydrogen bond with adenylate. The removal of the hydrogen bond by means of

the replacement Thr-51 → Ala and the change in the structure of the side chain (the replacement Thr-51 → Pro) increased k_{cat} and reduced $K_{M,ATP}$ for the tyrosine activation reaction [stage (*a*), Scheme 2], while in the aminoacylation reaction the binding to ATP improved but k_{cat} fell somewhat. Overall, a single-point mutation led to an appreciable improvement in catalytic activity: in the mutation Thr-51 → Pro, the ratio $k_{cat}/K_{M,ATP}$ increased from 8400 to 208 000 litre mol^{-1} s^{-1} for stage (*a*) and from 1860 to 95 800 litre mol^{-1} s^{-1} for stage (*b*).[55,112]

Catalysis by tyrosyl-tRNA synthethase leads to the formation of a pentacoordinate intermediate compound when ATP is subjected to nucleophilic attack by tyrosine:



The γ-phosphate group of the intermediate compound can interact with Thr-40 and His-45 side chains. The site-specific mutagenesis (the replacement Thr-40 → Ala, His-45 → Gly) provided an answer to the question concerning the mechanism of the catalysis. The mutations of the γ-phosphate-binding region of the enzyme increase the energy of the complex in the transition state, while the energy levels of the enzyme–substrate complexes do not change.[113,114] The main groups participating in catalysis are Thr-40 and His-45, but certain other groups, not included in the active centre, are also important for catalysis: Tyr-34, Cys-35, His-48, Thr-51, and Tyr-169.[114–118] Comparison of the free energy profiles for the mutant and initial enzymes has shown that the change in binding energy influences the rate constant and equilibrium constants. The principal conclusion derived from the study of the kinetic properties of the mutant and from the comparison of the free energy profiles is as follows: tyrosyl-tRNA synthetase uses in catalysis the mechanisms involving the stabilisation of the transition complex (in contrast to covalent, nucleophilic, and acid–base catalysis), stabilisation occurring via hydrogen bonds and electrostatic interactions.[114–122] By substituting the Lys and Arg residues located on either side of the substrate-binding pocket, it has been shown that the structure of the enzyme changes during catalysis: the positively charged side chains neutralise the negative environment of the pyrophosphate in the transition state on formation of tyrosyladenylate.[123,124]

The method of site-specific mutagenesis has also been applied to the study of the subunit interactions in this enzyme.[125–128] It has been shown that tRNA interacts with the N-terminal (catalytic) domain of one subunit and with the C-terminal section of another. The mutation of the residue located on the subunit interaction section (Phe-164 → Asp) led to the appearance of a negative charge on the binding section and induced a pH-dependent dissociation of the dimer, while the replacement Phe-164 → Lys led to the appearance of a positive charge. The introduction of mutations into the subunit interaction zone made it possible to control the subunit interactions in tyrosyl-tRNA synthethase by controlling the subunit interactions.

Cytochromes[129–138] and cytochrome *c* peroxidase.[139,140] The cytochromes are series of structurally similar haem-containing proteins involved in electron transport over large distances (10–20 Å). The role of residues which are present in all the cytochromes in electron transport has been elucidated with the aid of site-specific mutagenesis. According to X-ray diffraction data, the Phe-87 residue is located on the surface of the globule at a distance of 5 Å from haem. This residue in yeast iso-1-cytochrome *c* has been replaced by Ser, Gly, and Tyr[129–131] and also by Leu.[132] The mutants retained 20–70% of the activity compared with the native protein but the reduction potential fell by 50 mV. Thus Phe-87 does not participate directly in electron transport but ensures a hydrophobic environment for haem and determines the high reduction potential of cytochrome *c*. The mutations of Phe-87 induced conformational changes in the protein globule or cytochrome *c*, which is indicated by changes in the CD spectra in the region of the Soret band (385–445 nm). The mutations of the conservative Glu-66 in horse mitochondrial cytochrome *c* led to the conclusion that (1) it contributes to the stabilisation of the active conformation of the protein, possibly via the formation of a salt bridge, and (2) it participates in the binding of ATP.[133] Substitutions of the Cys-102 residue on the surface of iso-1-cytochrome *c*, which is rarely encountered in other cytochromes, did not alter the properties of the protein.[135,136] The sections retained in the course of evolution are not always involved in the mechanism of electron transfer. For example, the replacement of the conservative Lys-77 by Arg in iso-1-cytochrome *c* from yeast did not influence its interaction with cytochrome b_2, cytochrome *c* peroxidase, and cytochrome *c* oxidase.[134]

In order to elucidate the role of the sections of the polypeptide chain in the functioning of cytochrome P450, chimeral proteins were obtained from cytochrome P450c by replacing the central section, or the C-terminal section, or both sections by the corresponding sections of cytochrome P450d.[137] Cytochrome P450ccd had a higher monoxygenase activity than the initial cytochrome P450c, while the P450cdd mutant had a lower activity. The substrate specificity of the mutants did not change. Hence it was concluded that the central section in cytochromes P450c and P450d participate in the binding of substrates, while the C-terminal sections play an important role in electron transport.[137] However, other workers,[138] who obtained 13 mutants with substitutions at the C-terminal section of cytochrome P450d, discovered aminoacid residues on the C-end important for the binding of the apoprotein.

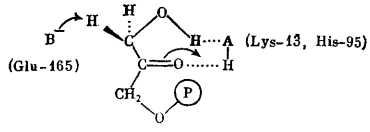
The method of site-specific mutagenesis helped to elucidate the role of Met-172 and Trp-51 in the catalytic function of **cytochrome c peroxidase**.[139,140] The interaction of the enzyme with hydroperoxide results in the formation of the intermediate (I), which contains [Fe^{IV}=O] and a free radical located on a certain hitherto unidentified group of the enzyme. It has been suggested that this group is either the indole ring of Trp-51, located 3.3 Å above and almost parallel to the plane of haem, or Met-172. However, the replacement of the indole ring by the phenylalanine ring, which is much more difficult to oxidise, did not entail a loss of activity. The Met-172 → Ser mutations also did not prevent the formation of compound (I). Thus the Trp-51 and Met-172 residues are not obligatory for the formation of compound (I) and the question of the location of the free radical remains open.

The serine proteases (trypsin[141–143] and subtilisin[144–151]) have been a satisfactory model for the construction of proteins by genetic engineering methods, since detailed information is available about their structure and kinetic and spectroscopic properties. Data on the structure of a protease from one source can be used to investigate the enzyme from another source. Although, for example, the crystal structure of trypsin from rat liver is unknown, the known structure of bovine trypsin was employed for experiments on site-specific mutagenesis having made sure that the aminoacid compositions of these enzymes are 74% identical. When Gly-166

in the substrate-binding region of the enzyme was replaced by the negatively charged Asp and Glu residues, the catalytic effectiveness (k_{cat}/K_M) of native subtilisin increased in relation to poor substrates and diminished in relation to good substrates. On replacement of Gly-166 by non-ionic residues (Ala, Leu), the catalytic activity increased for substrates containing hydrophobic groups. The change in specificity is associated mainly with the change in K_M and not in k_{cat}.[144-147]

Analysis of the crystal structures of the complexes of subtisilin with inhibitors demonstrated the presence of a hydrogen bond between the amide group of Asn-155 and the oxygen of the tetrahedral anion. Point changes in the position 155, leading to a change in the steric configuration (Asn-155 → Thr, Gln) or charge (Asn-155 → His, Asp) and in the hydrogen bond donor, diminished k_{cat} by a factor of 200–4000 and K_M by a factor of 7, which can be explained by a decrease in the stabilisation energy of the transition state by 9.2–20 kJ mol^{-1}.[148,149]

Triosephosphate isomerase.[152-155] The basic group (carboxylate ion) of Glu-165 is involved in the mechanism of catalysis by this enzyme:



Following the substitution Glu-165 → Asp, i.e. after the removal of one methylene group which shifts the basic group by 1 Å, k_{cat} diminished by a factor of 10^3 (from 4300 to 2.8 s^{-1}) with a slight alteration of K_M. Free energy calculations demonstrated that the mutation influences the free energy of the transition state in the enolisation stage.

Aspartate carbamoyltransferase[156-162] contains 12 polypeptide chains of two types: six catalytic and six regulatory. With the aid of point substitution in these chains and also on sections between them, the regions of the enzymes responsible for the regulatory properties were discovered: the loops formed by the residues 230–245 of the catalytic chains;[157] the Trp-209 residue on the section corresponding to the interaction of subunits.[156,158] The substitution Lys-83 → Gln in the catalytic chain diminishes the capacity of the enzyme for cooperation, while the substitution Gln-138 → Ala increases it, i.e. the aminoacid residues apparently determine many interchain and intrachain interactions.[159-162] Crystallographic studies on the complexes of the enzyme with substrate analogues showed that the active centre of aspartate carbamoyltransferase consists of several sections located on different catalytic subunits. A method has been proposed for the detection of the catalytic sites on oligomeric enzymes involving the construction of hybrids containing mutant monomers, for example, hybrids made up of monomers containing the Lys-84 → Glu, His-134 → Ala, and Ser-52 → His mutations or hybrids made up of trimers of the wild type and the Glu-84, Ala-134 binary mutant. The results of the mutation showed that the active species are formed only after the association of the chains with formation of an oligomer.

Staphylococcal nuclease.[163-167] Although the structure of staphylococcal nuclease has been determined by X-ray diffraction with a high resolution (1.5 Å), the detailed mechanism of the action of this enzyme has not been established. In order to elucidate the mechanism of the binding of Ca^{2+} ions, one of the residues surrounding Ca^{2+} in the protein was mutated (Asp-40 → Glu). The removal of one of the three *cis*-ligands diminished the affinity for Ca^{2+} by a factor of five and the maximum rate by a factor of 31. The mutation did not affect the binding of DNA, which agrees with the structural X-ray diffraction data, according to which Asp-40 interacts with the metal and not with the nucleotide chain.[163,164] The replacement of various Tyr residues in the substrate-binding cavity of the enzyme showed that the OH group in the Tyr residues does not play an appreciable role in catalysis. However, the mutations weakened the binding to the substrate (single-stranded DNA and RNA).[165] The mutations of Glu-43 in the active centre induce conformational changes in the protein at a distance of 15–30 Å from the residue 43.[166,167]

α-Antitrypsin[168-171]—an inhibitor of serine proteases. It has been shown by the method of site-specific mutagenesis that its inactivation is associated with the modification of Met-358 by oxidants. The Ile-356 and Pro-357 residues play an important role in the functioning of the inhibitor. The Met-358 residue in the active centre determines the specificity of the binding of antitrypsin to proteins. Native α-antitrypsin containing Met-358 recognises the Pro-Met sequence of elastase. The Ala-358, Ile-358, and Val-358 mutants proved to be effective inhibitors of neutrophilic and pancreatic elastases but not of cathepsin G; the Leu-358 and Phe-358 mutants inhibited cathepsin G preferentially.

Yeast 3-phosphoglycerate kinase[172-174] consists of two globular domains separated by a polypeptide chain loop. It has been suggested that the conformational change leading to the "merging" of the domains takes place during catalysis under the influence of the substrate. The Glu-190 and His-388 residues may interact with one another in the region between the domains (the formation of a hydrogen bond or of a salt bridge) and regulate the transition from the closed to the open form of the enzyme. The Glu-190 → Gln, Asp and His-388 → Glu mutations led to some decrease in catalytic activity and resulted in a change in the K_M with respect to both substrates. Thus the results of mutagenesis have shown that the model of the transition from the closed to the open form in catalysis requires further study.[172,173] The construction of a chimeral protein, one of the domains of which was isolated from human 3-phosphoglycerate kinase and the other from the corresponding yeast enzyme, was an interesting experiment. The hybrid retained the properties of the native proteins.[174]

Carboxypeptidases.[175-178] The three-dimensional structure of carboxypeptidase A has already been known for about 20 years, but certain details of the mechanism of its action are still not understood, for example, the role of Tyr-248 located at a distance of 3 Å from the substrate peptide bond undergoing cleavage. The results of site-specific mutagenesis have shown that this group participates in the binding of substrates and not in hydrolysis.[175,176] The OH group of Tyr-198 participates in the catalysis, stabilising the complex in the transition state.[177] The substitution Met-398 → Leu for yeast carboxypeptidase *y* altered the enzyme specificity. The mutation made the substrate-binding section more hydrophobic and the mutant therefore had an increased affinity, compared with the initial enzyme, for substrates with bulky and hydrophobic substituents.[178]

Other enzymes. The method of site-specific mutagenesis has also been used to investigate the influence of structure on the activity of dihydrofolate reductase from *E. coli* (DHFR).[179,180] The mutations of the Asp-27 residue in the enzyme active centre showed that it participates in the protonation of the substrate but not in the stabilisation of the transition state as had been postulated earlier. The dihydrofolate may be protonated directly by the asparagine group or with participation of water molecules. It has been shown for human DHFR that the Cys-6 residue partipates in the binding to NADPH and *p*-hydroxymercuriobenzoate.[181] Mutants with substitutions of aminoacids in the active centre have been obtained for human *lysozyme*.[182,183] Their analysis showed that the Asp-53 residue participates in the electrostatic stabilisation of the intermediate state in the hydrolysis of the glycosidic linkage. The existence of an aromatic residue in the position 63 and of Trp-64 is essential for the effective hydrolysis of chitin substrates but not for the lysis of bacterial substrates. The influence of insertions and deletions on the activity of ribonuclease P has been studied in detail[184] and it has been shown that, for the normal functioning of the RNA subunit of ribonuclease P from *E. coli*, a section of the secondary structure coded by the nucleotides 275–295 is essential.

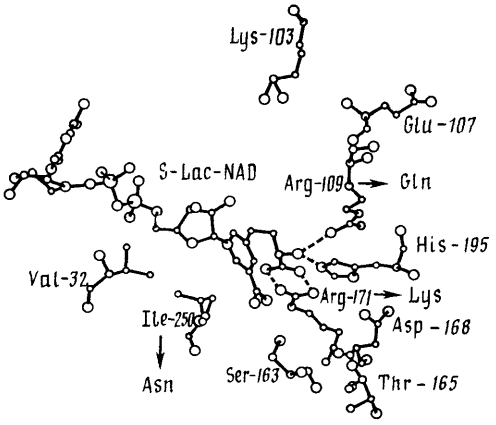


Figure 7. The active centre of native lactate dehydrogenase bound to the substrate.[188]

The sections at which the enzyme is bound to DNA have been discovered with the aid of mutagenesis for the **restriction endonuclease (restrictase)** EcoRI. The initial information about these sections was obtained from the X-ray diffraction structure of the complex EcoRI with a synthetic oligonucleotide. Point mutants (the Glu-111, Glu-144, and Arg-145 residues) hardly differed in their specificity from the initial enzyme, but their specific activity was reduced by a factor greater than 100 mainly as a result of the decrease in V.[185] Analysis of the crystal structures, especially of the geometry of the active centres, made it possible to alter with the aid of mutations the substrate specificity of *aspartate aminotransferase*,[186] *lactate dehydrogenase*,[187,188] and *alcohol dehydrogenase* [189] (see Section VI). The mutations of the Arg-109 residue in *lactate dehydrogenase* from *B. stearothermophilus*—a component of the mobile polypeptide group exposed to the solution—showed that it increases the polarisation of the carbonyl group of pyruvate and thus stabilises the transition state. The Ile-250 → Asn mutation impairs the catalytic properties of the enzyme (Fig. 7). The substitution of the catalytically important Ser-102 residue in *alkaline phosphatase*[190] by Cys altered the rate-limiting stage in the hydrolysis of esters: in contrast to the native alkaline phosphatase, the kinetics of the action of the mutant with the SH group instead of the OH group depend on the leaving group of the ester, a high catalytic constant being retained. Studies on *ATP synthetase* by the method of site-specific mutagenesis have been initiated.[191] Four residues capable of ionisation and important for the manifestation of activity have been found on the α- and β-subunits of the enzyme: Lys-175 and Asp-261 on the α-subunit and Lys-164 and Asp-252 on the β–subunit of the enzyme.

(b) The site-specific mutagenesis of enzymes with an unknown crystal structure

β-Lactamase[192–198] catalyses the opening of the β-lactam ring in penicillins and cephalosporins, ensuring the resistance of the cells to antibiotics. The class A β-lactamases contain the conservative catalytic triad (Ser-70) – (Thr-71) – (...) – (Lys-73). Point substitutions of Ser-70 permitted the conclusion that only mutants containing a nucleophilic group, for example, Cys, in the position 70 are catalytically active, although their activity is appreciably lower than that of the native β-lactamase.[192–196] Analysis of 19 mutants containing all the possible variants of the aminoacid residues in the position 71 showed that the Thr-71 residue does not participate in the binding of substrates and in catalysis but is important for the stabilisation of the protein and influences in some ways the specificity of the catalysis, determining the configuration of the β-lactam ring binding section.[197] The mutation Lys-73 → Arg led to an appreciable loss of activity. The role of the Lys-73 residue in catalysis has been discussed: the orientation of the proton moving away from the carboxy-group of the Ser-70 residue; the presence of a salt bridge with the Glu-166 residue participating in the hydrolysis of the acylenzyme.[198]

Ribulose diphosphate carboxylase.[199–202] Mutants with substitutions in the sections for the binding of CO_2, Mg^{2+} ions, and ribulose diphosphate have been obtained for the enzyme from *Rhodospirillum rubrum*. The ε-NH_2 group of the Lys-201 residue participates in the formation of the carbamate ion in the interaction with CO_2 and Mg^{2+}. The Asp-198 → Glu mutation showed that the carboxy-group of the Asp-198 residues also plays a role in the binding of the divalent metal.[199] It has been suggested that two residues (Lys-166 and His-291) participate in the enolisation of ribulose diphosphate, but the mutant in which the His-291 → Ala substitution has taken place showed a high catalytic activity. Thus the His-291 residue does not play a central role in catalysis.[200] The replacement of Lys-166 by seven different residues led to an appreciable loss of activity, but the mechanism of the participation of this group in catalysis has not been finally elucidated. The lack of crystallographic data precludes an answer to the question of the involvement of this group also in the binding of substrates.[201] The role of the small non-catalytic subunit S, point substitutions in which affected the catalytic activity but not the specificity, has been investigated by the method of site-specific mutagenesis in the enzyme from *Anacystis nidulans*. Hence it was concluded that the subunit participates in the activation of substrates but not in regulation.[202]

Lactose permease[203-212] is a hydrophobic enzyme controlling the transport of β-galactoside and of protons across a membrane. In order to investigate the role of the SH group of the active centre in catalysis, the Cys-148 → Gly[203,204] and Cys-148 → Ser mutants were obtained.[205,206] The results showed that Cys-148 is important for protecting the enzyme against inactivation (with the aid of the substrate) but is not obligatory for the transport of lactose. The mutations of other cysteine residues showed that, out of the eight residues, only Cys-154 is essential for the manifestation of activity.[207-209] According to preliminary data, the histidine residue also participates in the transport of the substrate. Each of the four histidine residues was replaced by Arg,[210] Gln, or Asp.[211] It was found that the His-322 residue is involved in the H^+ transport across the membrane via the protonation of the imidazole ring. The His-205 residue forms a hydrogen bond either with the substrate or within the protein. The Glu-325 and His-322 residues are located on the α-helix intersecting the membrane, form an ion pair, and act as components of the charge transfer system. The polarity of the residues and the distance between the imidazole and carboxy-groups therefore play an important role. The Glu-325 → Ala, Val mutations and the double mutation Glu-325 → His, His-322 → Glu lead to a partial or complete loss of activity.[212] Preliminary experiments with another membrane enzyme, melibiose permease, also demonstrated the presence of the His/Glu pair, which is important for catalysis.[212]

Other proteins and enzymes. The mutation-induced substitution of the Cys and His residues on three Ca^{2+} binding sections in the postulated active centre of the bioluminescent Ca^{2+}-dependent protein aequorin[213] showed that the binding of molecular oxygen does not proceed solely via the SH group. The His-58 residue and at least two of the four SH groups are necessary for the manifestation of activity. For another Ca^{2+}-binding protein, namely *calmodulin*,[214] a study has been made with the aid of mutagenesis of the influence of structure on the interaction with related proteins—myosin kinase, phosphodiesterase, and NAD kinase. The removal of the anionic α-helical section, which is complementary to the positively charged region of the binding protein, altered the binding specificity.

An unexpected result was obtained on substitution of the Thr–190 residue by Ile on the section close to the active centre of *malate dehydrogenase*: the catalytic activity increased mainly as a result of the increase in k_{cat}. It had been postulated earlier that the residues surrounding Asp-152, His-180, and His-187 in the active centre play a role in the binding of substrates and cofactors, but it has been found that the Thr-190 residue influences significantly also the activity.[215] The site-mutagenesis of the viral *neuraminidase*[216,217] suggested a model of the catalytic action of the enzyme involving the Glu-276 residue as a proton donor. Mainly conservative residues were used for the mutagenesis (12 residues identical for different types of viruses in 14 mutations). A series of mutants of the *RNA ligase* of the phage T4 have also been obtained.[218-220] The removal of 20% of the polypeptide chain from the C-end does not affect the RNA ligase activity. Mutations on the ATP-binding sections showed that the Lys-99 residue participates in the formation of the enzyme–adenylate complex, while the residues 100–101 do not affect this reaction. In order to elucidate the structure of the active centre of the *thiolopenicillin-binding protein* from *E. coli*, the Ser-307 and Ser-445 residues were substituted. It was found that the Ser-445 residue does not form part of the active centre, while the Ser-307 residue is responsible for the binding of penicillin in the penicillin-binding Ser–X–X–Lys section discovered in functionally similar proteins.[221] The Glu-49 residue, which is important for the manifestation of the activation of the α-subunit of *tryptophan synthetase*,[222] and the SH group important for the manifestation of the activity of *anthranylate synthetase*[223] have been identified. It has been shown for *isopenicillin N-synthetase* that the Cys-106 residue participates in the binding of the substrate but is not obligatory for catalysis.[224] The His-228 residue has been found next to the functionally important Lys-229 residue in the active centre of serine *hydroxymethyl transferase*.[225] The His-228 residue in enzymes of this type from various sources forms a hydrogen bond or a salt bridge with an aminoacid or pyridoxal phosphate. The results of mutation-induced substitutions in (or near) the phosphorylation centre of the *cAMP-dependent protein kinase* showed that the phosphorylation of the regulator subunit does not play a significant role in the regulation of the activity,[226] while the Tyr-371 residue makes a significant contribution to the manifestation of the positive cooperation between the two cAMP-binding centres of this subunit.[227]

V. Determination of the relation between the structure and stability of protein macromolecules (by genetic engineering methods)

1. Factors determining the structural stability of the protein

The difference between the Gibbs free energies of the native (N) and denatured (D) states of the cooperative domain or of a single-domain protein is a measure of the stability of its structure.[11,228]

Scheme 3

$$H \rightleftarrows D$$

$$\Delta G_{den} = G_{den} - G_{nat}.$$

For the majority of single-domain proteins, the temperature dependence of ΔG_{den} is described by a curve (Fig. 8) passing through a maximum at the temperature of the maximum stability of the protein (T_{max}) and having upper ($T_{den\,1} = T_m$) and lower ($T_{den\,2}$) temperature limits of the stability of the protein (for $\Delta G_{den} = 0$), i.e. proteins are inactivated in the cold as well as at an elevated temperature.[229] [T_m = melting point]. The temperatures T_{max} for many globular proteins are in the range from −3 to +15 °C and their ΔG_{den} in the physiological temperature range vary from 20 to 60 J mol^{-1}.[230]

The stability of the protein is characterised by the average temperature of the transition to the denatured state T_{den} (or T_m) at which 50% of the protein is denatured and $\Delta G_{den} = 0$, or by the temperature of the maximum stability of the protein T_{max} at which ΔG_{den} is a maximum (Fig. 8), or by the average concentration of the denaturing agent, etc. In the case of irreversible denaturation, the kinetic stability of proteins ($\Delta G^{\neq}_{den}$) is usually characterised by the rate constant k_{in} for the change in some kind of property of the protein (the CD or UV spectra, fluorescence, and, in the case of enzymes, their activity). The kinetic and thermodynamic parameters of the denaturation of proteins frequently vary in parallel with the variation of the denaturation conditions, i.e. the intramolecular interaction conditions, which govern the thermodynamic stability (ΔG_{den}), apparently determine their kinetic stability ($\Delta G^{\neq}_{den}$). In the general case, thermal denaturation is described by the scheme

Scheme 4

$$H \rightleftarrows D \rightarrow I,$$

where I is the irreversibly denaturing agent. When $K_{den} \geqslant 1$, the greater the stability of the enzyme in relation to reversible thermal denaturation, the greater its stability also in relation to the irreversible denaturation, but it has been assumed (without proof) that there is no such correlation for $K_{den} \ll 1$.[231] In the general case, the retention of the activity of enzymes may not in fact be identical with the stability of the entire protein globule.[232] The stability of enzymes is as a rule determined by the stability of the domains (or domain) forming the active centre and is in fact determined by the stability of the most labile domain.

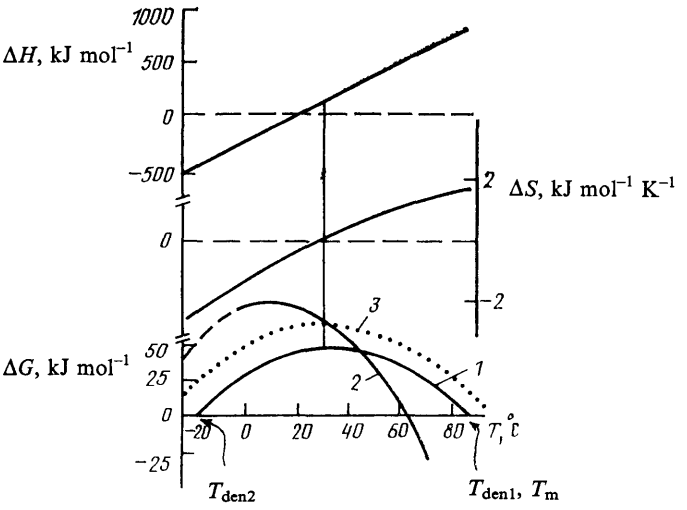


Figure 8. Temperature dependences of the enthalpy (ΔH), entropy (ΔS), and Gibbs free energy (ΔG) of the transition of proteins from the native to the denatured state: (1) myoglobin;[11] (2) α-chymotrypsin;[229] (3) hypothetical curve for the stabilised protein.

The factors responsible for the stability of the native conformation of the protein have been examined in detail in a review.[232] The main ones are as follows: (1) hydrogen bonds. (2) Increase in conformational entropy when the protein structure is uncoiled. (3) Long-range electrostatic interactions. If the charge distribution on the surface of the macromolecule is asymmetric in relation to composition and geometry, then the surface electrostatic potential in the absence of steric hindrance generates specific ion-binding regions whose saturation by the corresponding ions entails an increase in the stability of the protein.[233] (4) Hydrophobic interactions of the non-polar aminoacid residues resulting mainly from the van der Waals interactions between the molecules. The contributions of the hydrophobic interactions and hydrogen bonds to the stability of the protein structure are comparable.[228] (5) Intramolecular salt bonds (the energy of one bond is 1–3 kcal mol^{-1}) forming either single salt bridges or clusters. (6) Intramolecular –S–S– cross-links (1.8 kcal mol^{-1} per disulphide bond). (7) Formation of complexes by the protein with prosthetic groups and M^{n+} ions. The overall stability of the protein molecule is determined by the balance between the stabilising and destabilising contributions of the above interactions. The effective concentration of the interacting groups in the coiled and uncoiled states of the molecule makes a decisive contribution to the stability of the protein.[234]

The thermal stability of the protein is described by a complex relation between ΔG_{den} and temperature (Fig. 8).[11,229] The thermal stabilities of different enzymes must therefore be investigated and compared over a wide range of temperatures and not at a single temperature (cf. Lale and Klibanov[231] and Torchilin et al.[235]). The ratio of the ΔG_{den} for enzymes at a fixed temperature cannot be extrapolated to other temperatures. The formula for the calculation of small changes in ΔG_{den} from changes in T_{den} has been found for the case where the structure of one protein is altered, for example, by means of mutations.[229]

2. General relations for the specific alteration of the stability of proteins and enzymes *in vitro* and *in vivo*

(a) **Stability** *in vitro*. In order to elucidate the causes of the different stabilities of proteins, the first attempts have been made to find a correlation between the aminoacid composition of the proteins and their stability (T_{den}) and to compare the primary structures of proteins with identical functions and aminoacid sequences but having different stabilities (isolated from thermophilic and mesophilic microorganisms). These questions have been described in detail.[236] Correlations have been proposed between T_{den} and the hydrophobicity index of the residues and their average volume,[237] between T_{den} and the ratio of the areas of the surface of the protein accessible to the solvent in the native and denatured states,[238] between T_{den} and the percentage content of the stabilising and destabilising residues in the protein,[239] and between T_{den} and the aliphatic index of the protein.[240] These correlations suffer from the following disadvantages: (1) they are valid only within a limited group of proteins of a definite class; (2) T_{den}, which, as already stated above, does not characterise the stability of the protein as a whole, was adopted as a measure of stability. It is most likely that, in view of the large variety of factors responsible for the stability of the native conformation of the protein, the existence of a simple correlation is hardly possible and interest in studies of this kind has therefore diminished. When thermophilic (*t*-) and mesophilic (*m*-) proteins with known primary and three-dimensional crystal structures were compared, the structural factors responsible for the increased thermal stability of *t*-enzymes were discovered:[7,236,241,242] (1) the thermodynamically more favourable disposition of a large number of polar and charged residues on the surface and of hydrophobic residues on the inside, which is characteristic of *t*-enzymes;[241] (2) additional one–two salt and hydrogen as well as hydrophobic linkages in *t*-enzymes; (3) the most frequent occurrence of the Ser → Gly, Thr, Ala, Gly → Ala, Lys → Arg, Ala, Val → Ala, Ile, and Asp → Asn, Glu substitutions on transitions from *m*- to *t*-enzymes; an increased content of the Arg residues and a reduced content of Lys residues are characteristic of *t*-enzymes;[243] (4) *t*-enzymes have a larger number of S–S bonds, α-helices, and β-structures.[7]

Knowledge of the principal factors determining the stability of the native conformation of the protein in solution and of the characteristic features of the structures of *t*-proteins makes it possible to formulate the general principles enabling one to obtain proteins and enzymes with an increased stability: (1) the introduction of additional intramolecular and intersubunit bonds (salt and hydrogen bonds) and the enhancement of hydrophobic interactions increase ΔG_{den} for the protein, as shown in Fig. 8 (the upward shift of curve *1*). Calculation shows that a decrease in k_{in} by a factor of

10 at 70 °C requires an increase in ΔG_{den} by only 1.56 kcal mol^{-1}, while an increase in T_{den} from 35 to 45 °C for a polypeptide chain with M = 35 kD requires an increase in the number of hydrogen bonds by only 3%.[244] Therefore even one additional bond with an energy of 1–3 kcal mol^{-1}, achieved as a result of the replacement of 1–2 aminoacid residues, can increase significantly the stability of the protein. (2) The decrease in the conformational mobility of the fragments of the polypeptide chain as a result of the formation of additional intermolecular covalent or non-covalent "cross-links" in the globule; substitutions by residues with a lower conformational flexibility diminish the conformational entropy ΔS_{conf} and hence increase ΔG_{den} (and T_{den}).

The practical application of these principles to a specific protein makes it necessary to establish the position in the three-dimensional structure where the aminoacid residues are to be substituted (or modified), which particular aminoacid residues are to be replaced, what aminoacids are to replace them, and what method must be used to achieve these aims.

Up to the beginning of the 1980s, the only method for the specific alteration of the structure of aminoacid residues was the chemical modification of functional side groups in the protein. However, this method suffers from important disadvantages and limitations: it is possible to modify only residues having functional groups and accessible to the solvent under the given conditions, the volume of the residues increases after modification, and the number of modifying agents is small.[245] Site-specific mutagenesis makes it possible to replace one aminoacid residue at any position in the polypeptide chain by another, even one with a smaller bulk. It is possible to substitute only one aminoacid residue.

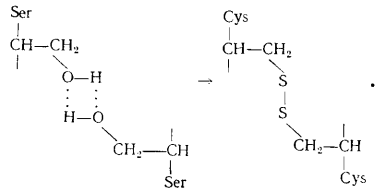
The general characteristics of the changes in the stability of proteins and enzymes as a result of chemical modification have been formulated.[236] In a coiled protein globule, there are three types of functional groups accessible to modification. *The functional residues of type I* are distributed on the surface of the globule. These are effectively hydrated, polar, or charged groups, which do not interact with one another under the given conditions, interacting only with the solvent or ions in solution; their modification does not affect the structure and stability of the protein. *The functional residues of type II* are distributed on or near the surface of the protein. In the native enzyme, they have an unfavourable environment (charged groups in a hydrophobic environment or in an environment comprising groups bearing like charges) and destabilise the protein globule. The number of groups of this type in the protein is usually small. Their modification accompanied by a change in the destabilising factor increases the stability of the protein as a function of the degree of modification and the nature of the modifying agent. *The functional residues of type III* are remote from the surface of the globule and form intramolecular salt and hydrogen bonds stabilising the enzyme. These groups are inaccessible to modification under the usual conditions and are modified for a loosened protein conformation (at an elevated temperature, at extreme pH values, or in the presence of organic solvents). The modification of type III groups diminishes the thermal stability as a result of the rupture of intramolecular bonds stabilising the native conformation of the protein. The structure of the enzyme approaches the denatured structure after such modification.

If the three-dimensional steric structure of the protein is known, it is easy to determine the aminoacid residues located on the surface of the molecule (type I groups) and within it (type III groups). The type II groups can be identified by studying the immediate environment of the given aminoacid residue. Comparison of the crystal structures of proteins of different classes (homologous proteins and their evolutionary precursors and *t*- and *m*-proteins) and data for the influence of site-specific mutagenesis on the stability of the enzymes suggest that the rules governing the influence of the chemical modification of different functional groups of the protein on its stability are on the whole valid also in cases where the stability is changed by site-specific mutagenesis. The aminoacid substitutions in homologous proteins are encountered most often on the surface of the molecule, so that the conformation of the molecule hardly changes. More radical insertions and deletions tend to be encountered in surface loops between units of the surface structure. For proteins which are more than 50% homologous, their inner region contains more than 90% common aminoacid residues.[101]

The mechanism and strategy used in altering the stability of proteins and enzymes with the aid of aminoacid substitutions. In selecting the residues in the proteins to be replaced, it is essential to take into account the following factors: (1) the replacement of type I residues by any others will hardly alter the stability of the protein, since their average environments in the coiled and uncoiled states differ little. (2) The replacement of type II residues (non-polar Leu, Ile, Val, Phe, Tyr, Trp, Met, and Pro residues), which are in contact with water, and of residues in an unfavourable electrostatic environment by small polar and charged residues can increase the stability of the protein.[246] So far, this has been confirmed only in nature. Cytochrome *c* from ox heart is more stable than the protein from horse heart by 1.1 kcal mol^{-1}, which is due, according to the authors, to the fact that the first protein contains the Gly residue, while the second contains the Lys residue exposed to water (the energy of the destabilisation of the Lys residue on contact with water is 1.5 kcal mol^{-1}). In principle, the presence of unfavourably solvated Lys, Tyr, Met, and Trp residues can be established by chemical modification under non-denaturing conditions. The replacement of exposed Lys residues by the His residues (or Arg as in *t*-enzymes) on the whole takes place with retention of the overall charge of the molecule, but the protein destabilisation energy is diminished.[246] (3) The internal region of the proteins is packed to the optimum extent and all the polar atoms form hydrogen bonds.[247] An arbitrary alteration of the size, shape, or polarity of the inner residues (type III residues) destabilises the molecule. Experiments designed to investigate proteins from temperature-sensitive strains (and their chemical modification) showed that mutations in the inner nucleus of the molecule are critical for stability, while any mutations on the surface of the globule do not alter the stability of the protein. However, it is possible to increase the thermal stability of the protein by accurate substitutions of these residues in order to form additional S–S, hydrogen, and salt bonds and by improving the hydrophobic packing, as happens in nature. The aminoacid residues to be replaced are selected with the aid of computer modelling in such a way that the steric disposition of neighbouring residues is not altered and the shortest-range intramolecular interactions are unaffected.[248]

One of the energetic approaches to the stabilisation of the globular structure of the protein involves the decrease in its *conformational entropy* ΔS_{conf} in the uncoiled state. This can be achieved, for example, by forming *intramolecular disulphide bonds* by replacing type III residues. Disulphide bonds have been formed in proteins during evolution in order to increase their stability

during functioning under altered extracellular conditions (the intracellular proteins do not have S–S bonds). The S–S bonds increase the thermodynamic stability of the protein and its resistance to irreversible thermal inactivation, diminishing its capacity for aggregation or adsorption and restricting the exposure of the inner hydrophobic regions to the solution.[249] The strategy used in introducing additional S–S bonds into proteins is as follows.
(a) The Ser or Thr residues, resembling most closely the Cys residue, are selected. For example, the Ser–Ser pair, forming hydrogen bonds, can be replaced by Cys–Cys:



If the Ser–Thr pair is replaced by Cys–Cys, then an empty space is inserted in place of the methyl group, which can be compensated by the replacement of the nearest Val residue by a Leu or Ile residue.[250] (b) For proteins with a known crystal structure, residues are selected by computer analysis in which the carbon atoms are in the same geometrical position as in the residues between which a disulphide bond exists. There are steric limitations to the introduction of an additional disulphide bond. The distance between the carbon atoms of the residues forming a S–S bond should be 4.5–6.5 Å, while the angles formed by the S–S bond should fit within the range of values established.[251] (c) The energy of the strain in the S–S bonds in native proteins varies from 0.5 to 4.7 kcal mol^{-1}. If the S–S bond produced has a strain energy >4.7 kcal mol^{-1}, then the stabilising effect of this cross-link in proteins will be abolished.[249] For this purpose, hypothetical S atoms are introduced into the model structure and candidates with unsuitable steric interactions are rejected. The selected model structures must meet yet another requirement. The cysteine mutations must be in the positions for which certain changes are observed in the course of evolution.[252]

A decrease in ΔS_{conf} is achieved also by replacing aminoacid residues having a conformational flexibility, for example, Gly, by residues containing a β-carbon atom and therefore having a smaller flexibility (Ala, Pro),[250] since

$$\Delta S_{conf/residue} = R \ln Z,$$

where Z is the number of conformations accessible to the residue in the uncoiled protein. The residues to be replaced are chosen so that steric hindrance should not arise after the substitution and the functional properties of the protein should not be altered.

The contribution of the Ala residue to ΔS_{conf} relative to the Gly residue is 2.4 e.u., while that of the Pro residue relative to Ala is 4 e.u. For example, the substitution Ala → Pro for the phage T4 lysozyme at pH 6.5 corresponds to an increase in ΔG_{den} by 1.4 kcal and to an increase in T_m by 3.5 °C.

There exist also several special approaches to increasing the stability of proteins. A higher structural stability of the α-helix in proteins can be attained following the substitution Val → Ala, while the substitution Ala → Val stabilises the β-layered protein structure. Overall, it is essential to use for the substitutions not unduly hydrophobic and bulky residues in order to preserve their structure.[253] The Ala residue is an interesting candidate for the substitution of other residues and proteins. On the one hand, its introduction frequently increases the thermal stability of proteins (see above) and its increased content is characteristic of *t*-enzymes. On the other hand, the "antifreeze" proteins (proteins preventing the freezing of cells in organisms living at temperatures below 0 °C) contain up to 60 mol. % Ala.[254] The site-specific enrichment of the proteins in Ala residues apparently increases their resistance to denaturation by the cold. The site-specific substitutions of charged residues at the ends of the α-helix (a change in the charge on the N-terminal component from +2 to –1, as in the *S*-peptide of ribonuclease) can increase the stability of the protein.[255] Somewhat different approaches to the stabilisation of proteins have been proposed in relation to tyrosinase: (1) the replacements should involve exposed residues or residues in β-coils, while the residues in the α-helix or in β-layers should not be altered; (2) the hydrophobicity of the molecule must not change; (3) the Asp → Glu, Lys → Gly, Val → Thr, Ser → Asn, Ile → Thr, Asn → Asp, and Lys → Arg substitutions are preferable (on the basis of a comparison of the structures of *t*- and *m*-enzymes).[256]

The stability of proteins in relation to oxidation and heavy metal ions can be increased by the following substitutions: Cys → Ala or Ser; Met → Glu, Val, Ile, Leu; Trp → Phe or Tyr. The bulk of the residues after the substitution hardly changes and steric hindrance does not arise, but groups sensitive to oxidation and complex formation with metal ions are substituted.[257] An increase in stability can be achieved by increasing the number of metal-binding sections in proteins by introducing additional negative charges.[1]

It has been suggested in certain studies that mutations stabilising the protein act additively, i.e. the stability of the protein can be increased significantly by accumulating a large number of stabilising mutations. An increase in the stability of the protein can also be achieved by removing (or saturating) certain uncompensated groups capable of forming hydrogen bonds and also by filling the inner cavities in the three-dimensional structure.[258] On the other hand, the substitutions of many residues can hardly act additively.[259]

Thus, although the overall theory of the prediction of the stability of a protein on the basis of its primary structure is far from completion, with the aid of graphical modelling in the presence of a three-dimensional structure it is possible to reveal the presence in the protein of residues of types I, II, and III, to select *a priori* the residues to be replaced and the substituting aminoacid residues, and to obtain the expected result.

(b) **The stability of proteins** *in vivo* in the cell is determined primarily by their resistance to intracellular proteolysis. The degradation of intramembrane proteins and long-lived cytoplasmic proteins takes place in lysosomes and in cytosol (the latter is the most selective). The rates of turnover of different proteins in the cell differ by factors up to 1000 (from 2 min to more than 200 h). The question arises what determines the half-life of the intercellular proteins ($\tau_{1/2}$). When 35 proteins with unknown crystal structure were introduced into a culture of Hela cells, it was found that their $\tau_{1/2}$ varied from 16 to 220 h and was not correlated with the size, charge, and content of hydrophobic aminoacids, the acylation of the N-terminal aminoacid, or the ratio of the α-helical and β-layered conformations.[260] Calorimetric data obtained for 22 of these proteins showed that their thermal stability *in vitro* is also not correlated with $\tau_{1/2}$. The hypothesis that intracellular proteases "recognise" more readily the inner part of proteins rather than the outer part (i.e. proteins with a greater mobility and flexibility should

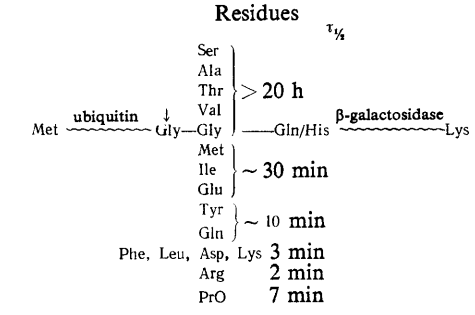
be degraded faster) has also not been confirmed, although it is known that proteins with a conformation distorted as a result of mutations are rapidly degraded.[261]

After denaturation, proteins become more accessible to proteolysis *in vitro*. However, comparison of the values of $\tau_{1/2}$ for native and denatured bovine serum albumin in cytosol and of β-lactoglobulin and cytochrome *c* in a culture of rat haepatoma cells showed that the denatured proteins are sometimes degraded several times more slowly than the native proteins (or at the same rate).[262]

Thus none of the global structural properties of the protein macromolecule which were examined above can serve as a criterion of its stability in the cell. Nevertheless, the latest data on protein sequences indicate the existence of definite signals in the sequence responsible for the degradation of the molecule. Thus proteins with $\tau_{1/2} < 2$ h contain one or more regions enriched in Pro, Glu, Ser, and Thr.[263] These regions have between 12 and 60 residues and are protected by clusters of positively charged aminoacids, which has been confirmed for a large number of intracellular proteins. There exists yet another type of signal for increased proteolysis: the S-peptide of ribonuclease — 20 N-terminal aminoacids in which (Lys-7)-Phe-Glu-Arg(Gln-11) are critical. The S-peptide "grafted" to lysozyme and insulin increased by a factor of 2 the rate of their degradation in fibroblasts.[264]

In the case of the ubiquitin-dependent non-lysosomal selective proteolysis, the state of the α-NH_2 group in the protein plays a fundamental role. The ubiquitinisation of the N-end of the protein is critical. The "N-end rule", establishing the dependence of the stability of the protein *in vivo* ($\tau_{1/2}$) on the nature of the N-terminal aminoacid, has been established.[265] With the aid of site-specific mutagenesis involving the cross-linking of β-galactosidase to the ubiquitin proteins, proteins with different aminoacids at the N-end of the molecule were obtained after the removal of the ubiquitin protein (Scheme 5). Depending on the aminoacid, $\tau_{1/2}$ varied as follows:

Scheme 5



If a Met group is placed at the ubiquitin binding site ahead of the residues destabilising β-galactosidase, then the $\tau_{1/2}$ for the enzyme increases (> 20 h). The authors suggest that, after the N-terminal peptidases have split off the N-end of the polypeptide chain synthesised, the "N-end reading" enzymes recognise specific residues at the N-end, which in fact determines their $\tau_{1/2}$. Long-lived proteins in the cell do indeed have Met, Ser, Ala, Gly, Thr, and Val residues at the end. On the other hand, eukaryotic toxins contain the destabilising Arg, Lys, Leu, Phe, and Asp residues at the N-end. Thus the rule established not only explains the metabolic stability of the protein but also suggests possible ways of regulating it. However, other workers[260] believe that this is a contranslation rule and holds immediately after translation for proteins which have still not coiled up, since $\tau_{1/2} = 1-3$ min is comparable to the protein translation time.

The oxidation of the His, Cys, and Met residues in the protein as well as the deamidation of Asn lead to its rapid degradation. The oxidised proteins are apparently recognised by special enzymes.[266] The stability of the intracellular proteins depends also on the site where they are located. The proteins forming part of multienzyme complexes and bound to membrane structures are more stable.

Thus the stability of proteins *in vitro* in solution is determined by specific intramolecular interactions. The stability of proteins *in vivo* is a more complex quantity and depends not only on the structure of the protein but also on its function — on the interaction of the protein with other proteins and cell components. The presence of specific signals in the aminoacid sequence, recognised by specific proteinases, and not the stability of the inner structure and the capacity for uncoiling when acted upon by various factors is of fundamental importance for the stability of the protein *in vivo* (in contrast to its stability *in vitro*).

3. The study of the stability of proteins with the aid of aminoacid replacements

The first studies of the influence of random aminoacid replacements on the stability of proteins were published at the end of the 1970s and concern the lysozyme of the T4 phage and the α-subunits of tryptophan synthetase (for a review of the first investigations, see Kutuzova et al.[236]). Since the proteins were isolated from mutant strains obtained in different ways, the influence of random aminoacid replacements on the stability of the proteins was also unpredictable. As the crystallographic technique and genetic engineering methods developed, it became possible to carry out specific replacements with prediction of a definite result. Examples of the study of the stability of different proteins with the aid of mutations (random and directed), which in many respects confirm and illustrate the features described above, are described below.

The presence of the type I, II, and III residues in proteins has been clearly demonstrated in a study of the lysozyme from the temperature-sensitive strains of the T4 phage, a convenient model monodomain protein, which has no S—S bond and prosthetic groups. Comparison of the crystal structures of the mutant lysozymes showed that the majority of the mutations destabilising lysozyme and encountered among residues which are inaccessible to the solvent have a rigid conformation, and are most static in the molecule. These aminoacid residues, located in the "hard" part of the molecule, form stabilising intermolecular bonds (type III residues). Mutant lysozymes, in which the inner Thr-157 residue, forming hydrogen bonds, has been replaced by the Ile, Ala, Ser, Leu, Asp, etc. residues, had T_{den} smaller than T_{den} for the native lysozyme by 9—13 °C (pH 2.0). This is caused by a dissociated and uncompensated hydrogen bond.[267—269]

Temperature-sensitive mutations are scarcely encountered among aminoacid residues which are exposed to the solvent and are mobile. For example, the replacement of the Cys-54, Glu-128, and Pro-86 residues by any others does not affect the stability of lysozyme.[270] Similar results were obtained for the staphylococcal nuclease[271] and kanamycin-nucleotidyl transferase.[272] Six mutants with one aminoacid substitution in the C-terminal region by the method of segment-directed mutagenesis were obtained for the last enzyme. The stability of all the Gly-242 → Arg, Gly-242 → Glu, Gly-204 → Asp, Phe-230 → Ser, and Pro-252 → Leu (except only Glu-210 → His) mutants diminishes in the above sequence and is smaller than the stability of the native enzyme. According to the authors, this can be explained by replacements among the residues

immersed into the hydrophobic nucleus of the molecule and depends on the nature of the substituted residue.[272] For the *protein* of the λ-phage repressor, which consists of an N-terminal DNA-binding domain (92 residues) and a C-terminal domain (144 residues), 11 single-point mutants were obtained in the region of the N-terminal domain. Analysis of the CD spectra showed that the mutations of the aminoacid residues exposed to the solvent do not alter the stability of the λ-repressor, while the mutations of the inner aminoacid residues diminish it. Just one surface mutation Gln-33 → Tyr increases the thermal stability of the λ-repressor (type II residue). It has been suggested that the Tyr-33 residue introduced interacts with the Tyr-22 residue over a distance of 25 Å, which stabilises the structure.[273]

A rare proved example of mutations of type II residues is of interest. All the mutations in subtilisin (Met-50 → Phe/Ile-107 → Val/Lys-113 → Arg), which increase its stability in the region of alkaline pH, are located near the surface of the molecule, which has been demonstrated for the three-dimensionai structure of the protein. The modelling of the Met-50 substitutions showed that the introduction of the Phe residue makes the hydrophobic interaction with the Trp-113 residue more favourable, which has in fact been observed in the crystal structure of the protein.[274]

Examples involving the introduction of additional S−S bonds. A patent has been recently obtained for the preparation of stable proteins with additional S−S bonds.[275] For this purpose, X-ray diffraction data for a specific protein are read into a computer and the sites for the introduction of Cys residues are found taking into account the necessary conditions indicated above.[39] The gene coding for the given protein with nucleotide replacements giving rise to the Cys residue is modified by the method of site-specific mutagenesis and is incorporated into the plasmid and expressed in the required culture.

A stable mutant of the T4 phage lysozyme was obtained in this way. The native enzyme has two Cys residues (54 and 97), but contains no S−S bonds. Analysis of the X-ray diffraction structure of lysozyme led to the selection of positions 3 and 97 for the formation of the disulphide bond and the Ile-3 → Cys mutant was constructed (with retention of 77% of activity) (Fig. 9). Oxidation of lysozyme *in vitro* leads to the formation of the (Cys-3)−(Cys-97) S−S bond. The stability then increases approximately by a factor of 30 (at 67 °C). Thermal inactivation in the presence of dithiothreitol (DTT) showed that the stabilisation is caused precisely by the formation of the S−S bond and not by the Ile → Cys substitution. At the same time, at an elevated temperature the Cys-54 residue is involved in the thiol−disulphide exchange, which inactivates the lysozyme as a result of the formation of intermolecular S−S bonds. Mutants were therefore obtained in which the Cys-54 residue was replaced by other residues with retention of the S−S bond in positions 3 and 97, which increased the stability several times more (the replacements of only the Cys-54 residue by Thr and Val do not alter the stability of lysozyme).[276] For a detailed study of the role of the Cys-54 and Cys-97 residues in the oxidative inactivation of the T4 phage lysozyme, the Cys-54/Cys-97 → Val/Ser double mutant was constructed. Its stability proved to be higher than that of native lysozyme. The use of such mutations made it possible to discover the mechanism of the thermal inactivation of lysozyme. The reversible uncoiling of the molecule at elevated temperatures is complicated by secondary processes, the main ones of which are oxidative inactivation with

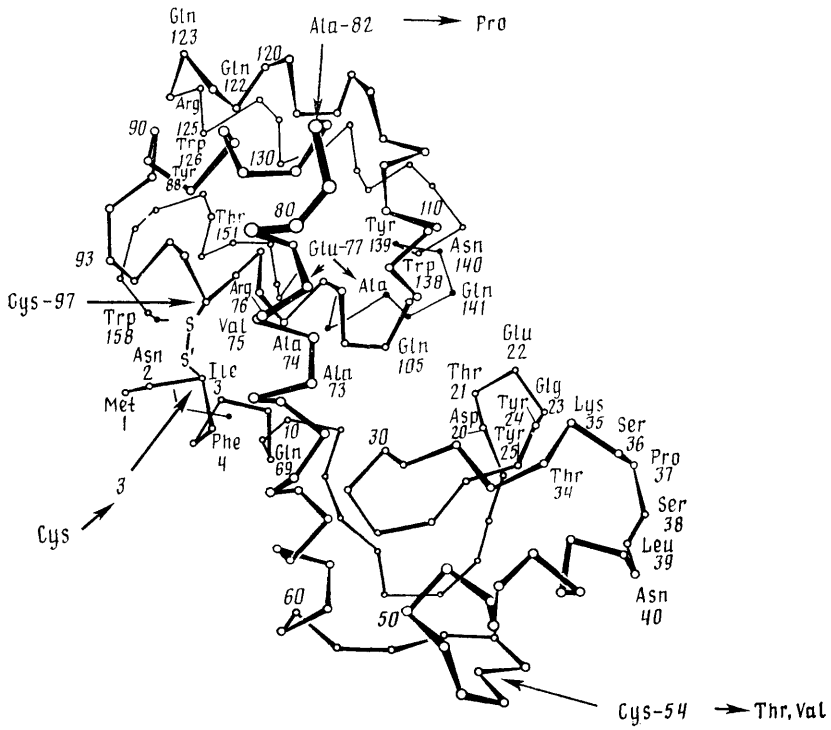


Figure 9. Steric structure of the T4 phage lysozyme.[277]

formation of oligomers linked by S—S bonds (this has been demonstrated electrophoretically). The removal of the Cys residue from lysozyme prevents the oxidative inactivation. On heating, this mutant uncoils reversibly with regeneration of the activity after treatment with guanidine hydrochloride (Gu.HCl). On the other hand, although the mutant with the 3—97 S—S bond, in which the Cys-54 → Val substitution has been carried out, does exhibit a higher stability, nevertheless the percentage regeneration of the activity after uncoiling in the presence of Gu.HCl and coiling up again is close to zero. Other proteins containing free Cys residues probably also undergo thermo-oxidative inactivation and can therefore be stabilised by similar aminoacid replacements.[277]

Dihydrofolate reductase (DFR) contains one polypeptide chain (159 residues). The Pro-39 residue is located opposite the Cys-85 residue. The distance between the α-carbon atoms is 6 Å, which is the same as in the S—S bond, and both residues are remote from the active centres. For this reason, the Pro-39 → Cys mutation leads to the formation of the 39—85 S—S bond after oxidation with dinitrobenzoate without a change in enzyme activity.[278] The DFR mutant obtained is more stable in relation to denaturation by urea and Gu.HCl to the extent of 1.8 kcal mol^{-1} and its coiling—uncoiling mechanism is altered.[279] Comparison of the crystal structures of the mutant and the native enzyme demonstrated their identity. The thermal stability of the mutant is approximately the same as that of the native DFR (T_m = 44 and 41 °C respectively).

Computer analysis for the λ-phage repressor protein[280] demonstrated the usefulness of the introduction of the (Cys-88)—(Cys-88′) S—S bond between the repressor subunits after the replacement of Thr-88 by the Cys residue.[281]

For the Thr-88 → Cys mutant, the S—S bond forms spontaneously; the mutant was more resistant to denaturation by urea than the native protein and T_m increased from 49 to 56—60 °C. On the other hand, the Tyr-85 → Cys mutant does not form a covalently bound dimer. Its chain structure is damaged and its stability is therefore lower than that of the native λ-repressor.[280]

For the subtilisin from *B.amyloliquefaciens*, which contains no S—S bonds, the Thr-22 and Ser-87 positions were selected for the production of such a bond. The Cys residues were introduced into these positions by the method of site-specific mutagenesis. They then formed spontaneously an S—S bond (either during the synthesis of the protein or during its secretion). These mutations did not affect the catalytic properties of subtilisin, while T_m increased from 58.9 to 62 °C. On the other hand, in the presence of DTT,† whereupon the S—S bond is ruptured, T_m = 56.2 °C. The rate of thermal inactivation of the Cys-22 → Cys-87 mutant is 1.5— 2 times lower, while $\tau_{1/2}$ in 8 M urea is higher by a factor of 2.[252]

Stabilisation of proteins by the replacement of residues with a high flexibility (Gly) in the uncoiled state of the protein by residues with a limited conformational freedom (Ala, Thr, Val, Ile, Pro). The replacement sites such that neither the tertiary structure of the protein nor the interactions with neighbouring residues were altered were selected for the T4 phage lysozyme by computer modelling. The selected candidates for replacement by the Ala residue were Gly-77 and Gly-113, while the Ala-82 residue was chosen for replacement by the Pro residue. Comparison of the crystal structures of the native and the Gly-77 → Ala and Ala-82 → Pro mutant lysozymes showed that they are virtually identical with the exception that the carbonyl carbon atom of the Glu-108 residue, closest to the Ala-77 residue, is displaced by 1.3 Å. Nevertheless the hydrogen bonds between the residues are unaffected. In the Ala-82 → Pro mutant, the H_2O molecule bound to the peptide nitrogen atom of the Ala-82 residue in the native protein is displaced. However, these are minor changes and do not influence the stability of lysozyme. For the Ala-77 and Pro-82 mutants, the T_m at pH 6.5 increase from 64.7 to 65.6 and 66.8 °C respectively, while the stabilisation energies are 0.4 and 0.8 kcal mol^{-1} respectively as a result of the decrease in the entropy of the uncoiling of the protein.[258] Similarly the Gly → Ala substitutions increase the stability of the *λ-repressor*[282] and of the *neutral protease from B. stearothermophilus* following the replacement Gly-144 → Ala.[283] This replacement was discovered on comparing the *t*- and *m*-enzymes. Admittedly, in such cases the increase in stability was explained by the replacement in the α-helix of the weakly helix-forming Gly residue by the effective helix-forming Ala residue.

The linear *correlation between the hydrophobicity of the substituted residues and the increase in* ΔG_{den} after point mutations was demonstrated for the α-subunit of tryptophan synthetase.[273,284] This correlation can be accounted for by the fact that the dielectric constant within the protein diminishes with increase in the hydrophobicity of the inner aminoacid residues, which intensifies the hydrophobic and electrostatic interactions. The point mutants of the α-subunit of tryptophan synthetase, isolated from mutant strains, are twice as stable as the native enzyme after the Glu-49 → Met, Gly-211 → Glu, and Gly-234 → Asp replacements and are less stable after the Phe-22 → Leu, Glu-49 → Gln, Leu-177 → Arg, Gly-211 → Arg, and Gly-213 → Val replacements. However, the tertiary structures of the mutants were not compared in these cases, which renders the results not altogether unambiguous.

Individual independent mutations can act additively on the stability of the protein and the accumulation of several such mutations with a stabilisation energy of 1 kcal mol^{-1} can give rise to a radical increase in ΔG_{den}. For example, the mutants of kanamycin-nucleotidyl transferase with the Asp-80 → Tyr and Thr-130 → Lys replacements and the Tyr-80/Lys-130 double mutant had maximum activities at 56, 60, and 62 °C respectively, while the native enzyme had a maximum activity at 52 °C. The resistance of the double mutant to proteolysis is greater than that of the point mutants and of the native enzyme.[285] The stabilisation is caused by additional hydrogen bonds and electrostatic interactions.

One of the causes of the instability of proteins is their oxidation with subsequent inactivation and denaturation. This refers particularly to proteins containing the Met, Cys, or Trp residues near the active centre. In subtilisin, the Met-50 and Met-222 residues are on the surface of the molecule, the Met-222 residue being close to the Ser-221 residue in the active centre. Following the replacement of Met-222 by the Cys-residue, the mutant had 138% of the original activity, while the replacements by the Ala and Ser residues resulted in mutants with 53% and 35% activities respectively. All other replacements led to a greater fall of the activity of subtilisin. Native subtilisin in 0.1 M H_2O_2 has $\tau_{1/2}$ = 2.5 min. The Met-222 → Ser, Ala, or Leu mutants are stable under these conditions for more than 1 h, while the

† Dithiothreitol (Translator).

Met-222 → Cys mutant is not oxidised by 0.1 M H_2O_2 at all. Thus the replacement of the Met-222 residue by residues resistant to oxidation increases the stability of subtilisin.[286]

The stability of proteins can be increased by reducing the surface area of the hydrophobic component in contact with the solvent, as shown for lactate dehydrogenase from *B. stearothermophilus*. The hydrophobic Ile-250 residue, located on the surface of the NAD binding pocket, was replaced by the hydrophilic Asn residue, whereupon k_{in} for the mutant diminished by a factor of 4.5 (at 90 °C), possibly as a result of the formation of the additional (Thr-246–(Asn-250) hydrogen bond.[287]

Examples confirming the general features of the variation of the stability of protein molecules and also examples where desirable results were achieved in the stabilisation of proteins on the basis of these features were considered above. However, many investigations showed that the influence of the aminoacid replacements on the stability is specific to each protein. As a rule, this applies to proteins with an unknown three-dimensional structure for which computer modelling of replacements and a reliable prediction of the effect are impossible.

In the monomeric form of iso-1-cytochrome *c* (the Cys-102 residue is blocked), the Pro-71 residue (the conservative residue as regards evolution) is important for the coiling of the protein and separates two short α-helical regions, defining the end of one and the beginning of the other helix. Its replacements by the Val, Thr, and Ile residues do not alter the functional properties of the structure but do alter the stability of the protein. Thus the stability of the mutants in relation to guanidine hydrochloride (ΔG_{den}, kcal mol^{-1}) diminishes in the following sequence: native; Val-71; Thr-71; Ile-71 (3.6, 2.6, 1.9, and 1.9 respectively).[288] The study of the denaturation kinetics showed that the location of the mutation in the primary structure determines the stage whose rate will change, while the nature of the substituent influences the change in the rate of the stage.[289]

The Pro-87 → Ser mutant of *adenylate kinase* from the temperature-sensitive strain of *E. coli* had k_{cat}/K_M smaller by a factor of 17 (at 27 °C) than the native enzyme and its thermal stability was lower. In the mutant adenylate kinase, the α-helix content diminished by 11%. Analysis of the CD spectra showed that the mutation broadens the structure of the enzyme. As a result, the sensitivity of the mutant to proteolysis increased. It has been suggested that the Pro-87 residue is located within the globule in the region of intensely interacting Asp-84, Phe-86, and Arg-88 residues and that its replacement by the more flexible and polar Ser residue disrupts these interactions and destabilises the structure of adenylate kinase.[290] The single point replacements in the α-subunit of *tryptophan synthetase*[291,292] and in dihydrofolate reductase[293] made it possible to discover the role of the individual aminoacid residues at each stage of the reversible denaturation of these enzymes by urea.

All the studies examined above were devoted to the stabilisation of enzymes in relation to reversible denaturation. The transition of the protein molecule to the irreversibly uncoiled state is caused by the aggregation of the protein molecules—the chemical modification of the functional groups more accessible to the solvent in the denatured state, as a result of which the molecule cannot coil up into the native conformation.

An attempt has been made to prevent the irreversible denaturation process with the aid of site-specific mutagenesis for *triose phosphate isomerase* (TPI)—an enzyme consisting of two identical subunits without S–S bonds and non-protein components. Analysis of the crystal structure established that the Asn residues located on the surface of the molecule between the subunits are deaminated in the thermal inactivation process with formation of charged Asp residues, which leads to the dissociation of the dimer into inactive monomers. The Asn-78 → Thr, Ile and Asn-14/Asn-78 → Thr/Ile mutants were obtained. Their half-lives at 100 °C increased from 13 to 17, 16, and 25 min respectively, and at the same time they had lower T_m (by 2–3 °C) than the native TPI, i.e. these mutations destabilise the enzyme in relation to reversible thermal inactivation.[294,295] This approach of the authors does not appear to be altogether correct, because the irreversible thermal inactivation of proteins involves not only (and not so much) the chemical modification of the functional groups but also physical processes leading to the aggregation of the denatured molecules after uncoiling. The stabilisation of the native conformation of the protein using the strategy described above appears more promising.

VI. Protein design

The construction of proteins with specified properties, structure, and stability, i.e. protein design, is a higher stage of development of genetic engineering. For a specific alteration of catalytic properties and substrate specificity, it is essential to investigate in detail the mechanism of the enzymic reaction and the structure of the substrate-binding section and of the active centre in order to alter the resistance to the effect of temperature or chemical reagents—factors which determine the stability of the protein, as well as the disposition of the groups, which are functionally important for catalysis, in order that the changes made in the structure to increase stability should not entail the loss of catalytic function. Protein design becomes possible provided that the X-ray diffraction structure of the protein has been obtained with a high degree of resolution. Such a structure serves as a basis for computer analysis and graphics.

For example, analysis of the steric disposition of the Met-222 group in subtilisin showed that its replacement by residues with small side chains (Glu, Ala, Ser, Cys, etc.) would not alter the conformation of the catalytic region and a high activity would be retained, while replacement by residues with long side chains would lower the activity of the enzyme. The results of mutagenesis confirmed the validity of the model. One of the mutants involving the replacement Met-222 → Cys showed a higher activity than the native subtilisin.[296]

Allowance for electrostatic effects in the active centre of subtilisin yielded mutants with an altered pH activity profile. Rules have been formulated for the construction of proteins with an altered pH dependence:[297] (1) the increase in the negative charge in the region of the active centre increases the pK_a for the acid groups, while an increase in the positive charge reduces the pK_a; (2) the change in the pH dependence is manifested at a low ionic strength; (3) the mutations should be carried out in such a way that the counterions do not concentrate in the region of the active centre. The change in the surface charge in the active centre of subtilisin (as a result of the replacement Asp-99 → Ser) (Fig. 10) led to an appreciable shift of the activity pH-profile to the acid region.[150,151,297]

The specific alteration of the substrate specificity of enzymes is achieved by altering the conformation of the substrate-binding region. The Gly-216 → Ala replacement in trypsin increased the catalytic activity in relation to arginine substrates, while the

Gly-226 → Ala replacement increased it in relation to lysine substrates.[141] In the case of lactate dehydrogenase, the active centre was reconstructed in the following way: Thr-246 → Gly (increase in the volume of the active centre), Asp-197 → Asn (neutralisation of the acid), and Glu-102 → Arg (introduction of a base to ensure the binding of a carboxylate). The ternary mutant oxidised malate 23 times more effectively than lactate and reduced oxaloacetate 500 times more effectively than pyruvate, while the native enzyme had a catalytic activity in relation to pyruvate 1000 times greater than in relation to oxaloacetate. Thus the mutant enzyme acquired the properties of malate dehydrogenase.[179]

In alcohol dehydrogenase, the long side chains of the Trp-93 and Thr-48 residues diminish the size of the substrate-binding pocket and the enzyme therefore catalyses most effectively the oxidation of alcohols with a short side chain (ethanol, 1-propanol). The replacement of these residues by residues of smaller size (Trp-93 → Phe and Thr-48 → Ser) increased the effectiveness of the oxidation of long-chain alcohols (1-octanol, 1-heptanol) compared with ethanol.[182]

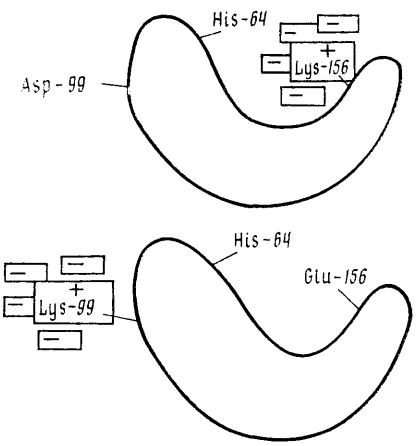


Figure 10. Schematic representation of the ionic atmosphere, made up of the Lys-99 and Lys-156 residues, of the active centre of subtilisin.[297]

The site-specific mutagenesis of haem ligands in cytochrome b_5 is an interesting example of the formation of a protein with new improved properties.[298]

The His-39 → Met replacement of the axial haem ligand greatly reduced the oxidation-reduction potential. The peroxidase activity in the reaction of the mutant protein with *o*-dianisidine increased by a factor of 30. In addition, the mutant acquired a new property—it catalysed the intermolecular oxidative demethylation of *NN*-dimethylaniline by H_2O_2. The native cytochrome b_5 is completely inactive in this reaction.

The C-terminal hydrophobic segment of penicillin-binding protein 5 from *E. coli*, which binds this protein to a membrane, was replaced by a shorter segment containing hydrophilic aminoacids, which led to the formation of a water-soluble protein with complete retention of its activity.[299]

The construction of proteins with a much higher stability by employing the general approaches and features examined in Section 5 is a brilliant example of protein design. The introduction of intramolecular disulphide bonds stabilised the T4 phage lysozyme,[276] subtilisin,[252] the N-terminal domain of the λ-repressor,[280] and dihydrofolate reductase.[180] The replacement of Met-222 in subtilisin by residues resistant to oxidation (Ser, Ala) greatly increased the stability of this enzyme.[286] The replacements of unbranched residues by branched β-carbon chains (Thr, Val, Ile) restricted the conformation of the protein and hence greatly increased the stability of the T4 lysozyme.[258]

Studies designed *to create protein-hybrids* possessing the set of properties of their constituent proteins are of interest. Thus the *E. coli* protein possessing adenylate cyclase and β-galactosidase activities has been constructed.[300] It consists of the first 556 (out of 848) aminoacid residues of adenylate cyclase (the enzyme responsible for the synthesis of cAMP), linked via the synthetic pentapeptide Val-Gly-Asp-Pro-Val to the eighth residue of β-galactosidase (M = 175 kD). The resulting chimeral protein is a tetramer. The pH optimum of the adenylate cyclase activity in the hybrid protein has been displaced to the alkaline region (from 8.5 to 9.5) and an excess of ATP does not inhibit the adenylate cyclase activity of the hybrid in contrast to the native enzyme, which can apparently be accounted for by the absence of a regulatory domain from the hybrid. The kinetic properties of the hybrid protein are close to those of both constituent proteins, but the hybrid is less stable than β-galactosidase. The hybrid of the human growth hormone and the signal peptide of the extramembrane *E. coli* protein OmpA has also been obtained by the genetic engineering method. The recombinant DNA was expressed in the cells of *E. coli* and the growth hormone synthesised was secreted into the periplasm of *E. coli*. It had the correct secondary structure and formed the required S–S bonds. Thus the addition of the signal peptide ensured the passage of the hybrid protein through the cell membrane of *E. coli* and the correct coiling of the protein synthesised with formation of the required disulphide bonds.[301]

—o0o—

The analysis of the literature described shows that the protein engineering technique is a revolutionary experimental instrument for the solution of fundamental problems in modern enzymology: the determination of the structure–function–stability of protein macromolecules relation. The study of numerous mutant proteins with the aid of computer graphics has provided an algorithm for the preparation of enzymes with properties specified beforehand in order to use them in biotechnology. The need for the employment of genetic engineering methods in modern protein chemistry is not only evident but is genuinely essential for the solution of both theoretical and applied problems.

References

1. P N Bryan *Biotechnol. Adv.*, 1987, **5** 221
2. A White, P Handler, E L Smith *Principles of Biochemistry*, McGraw Hill, New York, 1968 (Translated into Russian), Izd. Mir, Moscow, 1981, Vol. 2, Chapters 25, 26
3. J D Watson, J Tooze, D J Kurtz *Recombinant DNA: A Short Course*, Scientific American Books, New York, 1983 (Translated into Russian), Izd. Mir, Moscow, 1986

4. G Heijne, C Blomberg, H Liljenström *J. Theor. Biol.*, 1987, **125** 1
5. C Ghelis, J Yons *Protein Folding*, Acad. Press, New York, 1983
6. M T McLaman, W H Andrews, J G Files *J. Biotechnol.*, 1985, **2** 117
7. G D Kutuzova, N N Ugarova, I V Berezin *Bioorg. Khim.*, 1982, **8** 1445
8. R F Dolittle, V Mire *Nauki*, 1985, (12) 40
9. G D Rose, A R Geselonitz, G J Lesser, et al. *Science*, 1986, **229** 834
10. R Yaenicke *Prog. Biophys. Molec. Biol.*, 1987, **49** 117
11. P L Privalov *Vestn. Akad. Nauk SSSR*, 1986, (2) 72
12. P L Privalov *Biofizika*, 1985, **30** 722
13. T W Traut *Mol. Cell. Biochem.*, 1986, **70** 3
14. D B Wetlaufer *Adv. Protein Chem.*, 1981, **34** 61
15. D Ollis, P Brick, R Hamlin, et al. *Nature*, 1985, **313** 762
16. E D Kiehn, J J Holland *Nature*, 1970, **226** 544
17. E Querol, A Parrilla *Experientia*, 1987, **43** 920
18. T E Creighton, D P Goldenberg *J. Mol. Biol.*, 1984, **179** 497
19. A Protzel, A J Morris *J. Biol. Chem.*, 1974, **249** 4594
20. L L Randell, L G Josefsson, S J Hardy *Eur. J. Biochem.*, 1980, **107** 375
21. I J Purvis, A J E Bettany, T Ch Santiago, et al. *J. Mol. Biol.*, 1987, **193** 413
22. R L Burke, P Tekamp-Olson, R Najarian *J. Biol. Chem.*, 1983, **258** 2193
23. J R Coggins, D G Hardie *Multidomain Proteins*, Elsevier, Amsterdam, 1986, p. 1
24. C R Matthews, M K Hurle *Bio Essay*, 1987, **6** 254
25. C R Matthews *Methods Enzym.*, 1987, **154** 498
26. J Sugihara, T O Baldwin *Biochemistry*, 1988, **27** 2872
27. T Imoto, H Yamada, T Yasukochi, et al. *Prot. Eng.*, 1987, **1** 333
28. C B Marks, H Naderi, Ph A Kosen, et al. *Science*, 1987, **235** 1370
29. C-I Brändeu *Chem. Scr. B*, 1986, **26** 247
30. M G Rossman *Phil. Trans. Soc. London*, 1981, **293** 191
31. S T Rao, M G Rossman *J. Mol. Biol.*, 1973, **76** 241
32. S J Wodak, J Janin *Biochemistry*, 1981, **20** 6544
33. C Quinto, M Quirago, W F Swain, et al. *Proc. Natl. Acad. Sci. USA*, 1982, **79** 31
34. M Leicht, C L Long, T Chandra, et al. *Nature*, 1982, **297** 655
35. H Naora, N J Deacon *Proc. Natl. Acad. Sci. USA*, 1982, **70** 6196
36. Go H, M Nosaka *Evolution of Catalytic Function, Cold Spring Harbor Lab.*, 1987, Vol. 52, p. 915
37. T Maniatis, E F Fritsch, J Sembruk *Methods for Genetic Engineering: Molecular Cloning* (Translated into Russian), Izd. Mir, Moscow, 1984, 480 pp
38. D H Glover *Gene Cloning. Mechanics of DNA Manipulation*, Chapman and Hall, London–New York, 1984, 222 p
39. B D Hames, S J Higgins (Editors) *Transcription and Translation: A Practical Approach* (Translated into Russian), Izd. Mir, Moscow, 1987, 399 pp
40. M Smith *Philos. Trans. R. Soc. London, A*, 1986, **317**(1540) 295
41. G Dalbadie-McFarland, J H Richards *Ann. Rev. Med. Chem.*, 1983, **18** 237
42. M Smith *Ann. Rev. Genet.*, 1985, **19** 423
43. D Shortle, D Botstein *Methods Enzym.*, 1983, **100** 457
44. L A Loel *Cell.*, 1985, **40** 483
45. R Pine, P C Huang *Methods Enzym.*, 1987, **154** 415
46. R M Myers, L S Lerman, T Maniatic *Science*, 1985, **229** 242
47. K Itakura, J J Rossi, R B Wallace *Ann. Rev. Biochem.*, 1984, **53** 323
48. J Blöcker, R Frank *Biotechnology: Potentials and Limitations* (Edited by S Silver), Dahlem Konferenzen, Springer-Verlag, Berlin–Heidelberg–New York–Tokyo, 1986, p. 41
49. J A Wells, M Vasser, D B Powers *Gene*, 1985, **34** 315
50. D Botstein, D Shortle *Science*, 1985, **229** 1193
51. G Winter, P Carter, H Bedouelle, et al. *Biotechnology: Potentials and Limitations* (Edited by S Silver), Dahlem Konferenzen, Springer-Verlag, Berlin–Heidelberg–New York–Tokyo, 1986, p. 55
52. P Carter *Biochem. J.*, 1986, **237** 1
53. M J Zoller, M Smith *Methods Enzym.*, 1983, **100** 468
54. M J Zoller, M Smith *DNA*, 1984, **3** 479
55. A R Fersht, G P Winter *Ciba. Found. Symp. 111*, Pitman, London, 1985, p. 204
56. M J Zoller, M Smith *Methods Enzym.*, 1987, **154** 329
57. W Kramer, H J Fritz *Methods Enzym.*, 1987, **154** 350
58. W Mandecki *Proc. Natl. Acad. Sci. USA*, 1986, **83** 7177
59. V A Efimov, O V Mirskikh, O G Chakhmakhcheva, Yu A Ovchinnikov *Bioorg. Khim.*, 1985, **11** 621
60. V A Efimov, O V Mirshkikh, O G Chakhmakhcheva, Yu A Ovchinnikov *FEBS Lett.*, 1985, **181** 407
61. V A Petrenko, N N Pozdnyakov, S M Kipriyanov, et al. *Bioorg. Khim.*, 1986, **12** 1088
62. V A Petrenko, L N Semenova, S M Kupriyanov, et al. *Bioorg. Khim.*, 1987, **12** 1612
63. V A Petrenko, S M Kipriyanov, L N Semenova, A N Boldyrev *Bioorg. Khim.*, 1987, **13** 344
64. W J Martin, R W Davies *Biotechnology*, 1986, **4** 890
65. J C Moores *Anal. Biochem.*, 1987, **163** 1
66. B W Gilson, K Beiman *Proc. Natl. Acad. Sci. USA*, 1984, **81** 1956
67. M J Zvelebil, G J Barton, W R Taylor, M J Stenberg *J. Mol. Biol.*, 1987, **195** 957
68. G Voordouw, S Brenner *Eur. J. Biochem.*, 1986, **159** 347
69. P W Finch, R E Wilson, K Beocon, et al. *Nucl. Acids. Res.*, 1986, **14** 7695
70. G Schumaster, D Sizmann, H Haug, et al. *Nucl. Acids. Res.*, 1986, **14** 5713
71. G-H Fang, P Kenigsberg, M J Axley, et al. *Nucl. Acids. Res.*, 1986, **14** 8061
72. S Inouye, M Noguchi, Y Sakaki, et al. *Proc. Natl. Acad. Sci. USA*, 1985, **82** 3154
73. B Perryman, S A Kerner, T J Bohlmeyer, R Toberts *Biochem. Biophys. Res. Commun.*, 1986, **140** 981
74. C J M Saris, B F Tack, T Kristensen, et al. *Cell*, 1986, **46** 201
75. E Takano, M Maki, M Hatanaka, et al. *FEBS Lett.*, 1986, **208** 199
76. T Icho, C P Sparrow, C B H Raetz *J. Biol. Chem.*, 1985, **260** 12078
77. T Icho, C E Bulawa, C B H Raetz *J. Biol. Chem.*, 1985, **260** 12091
78. Yu A Ovchinnikov, N E Broude, K E Petrukhin, et al. *Biokhimiya*, 1986, **51** 1491
79. T O Baldwin, J-W Lin, L H Chen, et al. *Bioluminescence and Chemiluminescence, New Perspectives. Proceedings of the IV International Biolum. and Chemilum. Symp., Freiburg, 1986*, John Wiley and Sons, New York, 1987, p. 351

80. W Kugimiya, Y Otani, Y Hashimoto, Y Takagi *Biochem. Biophys. Res. Commun.*, 1986, **141** 185
81. C-P D Tu, B Qian *Biochem. Biophys. Res. Commun.*, 1986, **141** 229
82. A Hartig, H Ruis *Eur. J. Biochem.*, 1986, **160** 487
83. H E Grennett, F D Ledley, L L Reed, S L Woo *Proc. Natl. Acad. Sci. USA* 1987, **84** 5530
84. T Y Borgford, N Y Beand *Biochemistry*, 1987, **26** 2480
85. G E Schulz, I Shulz, A G Tomaselli, et al. *Eur. J. Biochem.*, 1986, **161** 127
86. G E Sheell, L K Lanett, J B Lingrel *Nature*, 1986, **321** 429
87. C B Newgard, K Nakano, P K Hwoeng, R J Fletterick *Proc. Natl. Acad. Sci. USA*, 1986, **83** 8132
88. E H Fischer *Bull. Inst. Pasteur*, 1983, **87** 7
89. W Gilbert *Nature*, 1978, **271** 501
90. H Neurath *Chem. Scr. B*, 1968, **26** 221
91. T Hunter, J A Cooper *Adv. Cyclic Nucleotide Res.*, 1984, **17** 443
92. K Takio, R D Wade, S B Smith, et al. *Biochemistry*, 1984, **23** 4207
93. H Nojima *FEBS Lett.*, 1987, **217** 187
94. R G Brennan, L H Weaver, B W Mattheus *Chem. Scr. B*, 1986, **26** 251
95. R Jaenicke *Progr. in Biophys. and Mol. Biol.*, 1987, **49**(2–3) 117
96. G J Barton, J E Sternberg *Prot. Eng.*, 1987, **1** 89
97. L Kelly, L A Halladay *Prot. Eng.*, 1987, **1** 137
98. P Argos, Y K Mohanakao *Methods Enzym.*, 1986, **130** 185
99. P Argos, A D Tucher, L Phillipson *Virology*, 1986, **149** 208
100. P Argos *J. Mol. Biol.*, 1987, **193** 385
101. T L Blundell, B L Sibanda, M Y E Sternberg, Y M Thornton *Nature*, 1987, **326** 347
102. T J Hubbard, T L Blundell *Prot. Eng.*, 1987, **1** 159
103. M J Sutcliffe, I Haneef, D Carney, T L Blundell *Prot. Eng.*, 1987, **1** 377
104. M J Sutcliffe, F R F Hayes, T L Blundell *Prot. Eng.*, 1987, **1** 385
105. S H Bryant, M J E Sternberg *J. Mol. Graphics*, 1987, **5** 4
106. M J E Sternberg *NATO ASI Ser, Ser. A*, 1987, **126** 141
107. T A Jones *J. Appl. Crystallogr.*, 1978, **11** 268
108. J Novotny, C Auffray *Nucl. Acids Res.*, 1984, **12** 243
109. J Novotny, C Auffray *Proc. Natl. Acad. Sci. USA*, 1986, **83** 742
110. P Klein, J A Jancquez, C Delisi *Math. Biosci.*, 1986, **81** 177
111. A J Wilkinson, A R Fersht, D M Blow, G Winter *Biochemistry*, 1983, **22** 3581
112. A R Fersht, J P Shi, A J Wilkinson, et al. *Angew. Chem.*, 1984, **96** 455
113. A J Wilkinson, A R Fersht, D M Blow, et al. *Nature*, 1984, **307** 187
114. R J Leatherbarrow, A R Fersht, G Winter *Proc. Natl. Acad. Sci. USA*, 1985, **82** 7840
115. A R Fersht, R J Leatherbarrow, T N G Wells *Nature*, 1986, **322** 284
116. T N C Wells, A R Fersht *Biochemistry*, 1986, **25** 1881
117. T N C Wells, C K Ho, A R Fersht *Biochemistry*, 1986, **25** 6603
118. K A Brown, P Brick, D M Blow *Nature*, 1987, **326** 416
119. A R Fersht, R J Leatherbarrow, T N C Wells *Philos. Trans. Soc. London, A*, 1986, **317** 305
120. H Bedouelle, P Carter, G Winter *Philos. Trans. Soc. London, A*, 1986, **317** 433
121. C K Ho, A R Fersht *Biochemistry*, 1986, **25** 1891
122. A R Fersht *Biochemistry*, 1987, **26** 8031
123. A R Fersht *Biochemistry*, 1988, **27** 1577
124. A R Fersht, J W Knill-Jones, H Bedouelle, G Winter *Biochemistry*, 1988, **27** 1581
125. D H Johnes, A J McMillan, A R Fersht *Biochemistry*, 1985, **24**(2) 5852
126. W H J Ward, D H Jones, A R Fersht *J. Biol. Chem.*, 1986, **261** 9576
127. P Carter, H Bedouelle, G Winter *Proc. Natl. Acad. Sci. USA*, 1986, **83** 1189
128. H Bedouelle, G Winter *Nature*, 1986, **320** 371
129. G J Pielak, A G Mauk, M Smith *Nature*, 1986, **313** 152
130. G J Pielak, K Oikawa, A G Mauk, et al. *J. Am. Chem. Soc.*, 1986, **108** 2724
131. G J Pielak, D W Concar, G R Moore, J P Williams *Prot. Eng.*, 1987, **1** 83
132. P B W ten Kortenaar, P J M H Adams, G I Tesser *Proc. Natl. Acad. Sci. USA*, 1985, **82** 8279
133. C J A Wallace, B E Corthesy *Prot. Eng.*, 1986, **1** 23
134. D Holzche, L Puncipio, K T Conklin, et al. *J. Biol. Chem.*, 1987, **262** 7125
135. R L Culler, G J Pielak, A G Mauk, M Smith *Prot. Eng.*, 1987, **1** 95
136. S L Mayo, W R Ellis, R J Crutchley, H B Gray *Science*, 1986, **233** 948
137. T Sakaki, M Shibata, Y Yubusaki, H Ohkawa *DNA*, 1987, **6** 31
138. T Shimizu, K Hiramo, M Takahashi, et al. *Biochemistry*, 1988, **27** 4138
139. D B Goodin, A G Mauk, M Smith *Proc. Natl. Acad. Sci. USA*, 1986, **83** 1295
140. L A Fishel, J E Villafranca, J M Mauro, J Kraut *Biochemistry*, 1987, **26** 351
141. C S Craik, C Loveqman, T Fletcher, et al. *Science*, 1985, **228** 291
142. C S Craik, S Roczniak, S Sprang, et al. *J. Cell. Biochem.*, 1987, **33** 199
143. L Craf, C S Craik, A Patthey, et al. *Biochemistry*, 1987, **26** 2616
144. D A Estell, J V Miller, T P Graycar, et al. *World Biotech. Rep.*, 1984, **2** 181
145. D A Estell, T P Graycar, J V Miller, et al. *Science*, 1986, **233** 659
146. R R Bott, M Ultsch, J Burnier *ACS Symp. Ser.*, 1987, **334** 139
147. J A Wells, D B Powers, R R Bott, et al. *Proc. Natl. Acad. Sci. USA*, 1987, **84** 31
148. J A Wells, D C Cunningham, T P Graycar, D A Estell *Philos. Trans. R. Soc. London, A*, 1986, **317** 415
149. J A Wells, D B Powers *J. Biol. Chem.*, 1986, **261** 6564
150. P G Thomas, A J Russel, A R Fersht *Nature*, 1985, **318** 375
151. A J Russel, P G Thomas, A R Fersht *J. Mol. Biol.*, 1987, **193** 803
152. D Straus, R T Raines, E Wawashima, et al. *Proc. Natl. Acad. Sci. USA*, 1985, **82** 2272
153. R T Raines, D R Straus, W Gilbert, J P Knowles *Philos. Soc. London, A*, 1986, **317** 371
154. R T Raines, E E Sutton, D R Straus, et al. *Biochemistry*, 1986, **25** 7142

155. G A Petsko, R C Davenport, D Frankel, V L Raybhandary *Biochem. Soc. Trans.*, 1984, **12** 229
156. K A Smith, S F Nowlan, S A Middleton, et al. *J. Mol. Biol.*, 1986, **189** 227
157. S A Middleton, E R Kantrowitz *Proc. Natl. Acad. Sci. USA*, 1986, **83** 5866
158. S F Nowlan, E R Kantrowitz *J. Biol. Chem.*, 1985, **260** 14712
159. E A Robey, H K Schachman *Proc. Natl. Acad. Sci. USA*, 1985, **82** 361
160. E A Robey, S R Wente, D W Markly, et al. *Proc. Natl. Acad. Sci. USA*, 1986, **83** 5934
161. H K Schachman *Biochem. Soc. Trans.*, 1987, **15** 772
162. S R Wente, H K Schachman *Proc. Natl. Acad. Sci. USA*, 1987, **84** 31
163. E H Serpersu, D Shortle, A S Mildvan *Biochemistry*, 1986, **25** 68
164. E H Serpersu, D Shortle, A S Mildvan *Biochemistry*, 1987, **26** 1289
165. E Uhlmann, J A Smith *Nucleosides Nucleotides*, 1987, **6** 331
166. D W Hibler, N J Stolowich, H A Reynolds, et al. *Biochemistry*, 1987, **26** 6278
167. J A Wilde, P H Bolton, M Dell'Acqua, et al. *Biochemistry*, 1988, **27** 4127
168. J Travis, M Owen, P George, et al. *J. Biol. Chem.*, 1985, **260** 4384
169. M Courtney, S Jallet, J-H Tissier, et al. *Nature*, 1985, **313** 149
170. S Jallet, D Carvallo, L H Tessier, et al. *Prot. Eng.*, 1986, **1** 29
171. M Courtney, S Jallet, J-H Ressier, et al. *Philos. Trans. R. Soc. London, A*, 1986, **317** 381
172. M T Mas, L E Resplandor, A D Riggs *Biochemistry*, 1987, **26** 5369
173. C A B Wilson, N Hardman, L A Fothergill-Gilmore, et al. *Biochem. J.*, 1987, **241** 609
174. M T Mas, C Y Chen, R A Helzeman, A D Riggs *Science*, 1986, **233** 788
175. S J Gardell, C S Craik, D Hilvert, et al. *Nature*, 1985, **317** 551
176. D Hilvert, S J Gardell, W R Rutter, E T Kaiser *J. Am. Chem. Soc.*, 1986, **108** 5298
177. S J Gardell, D Hilvert, J Barnett, et al. *J. Biol. Chem.*, 1987, **262** 576
178. J R Winther, M C Keilland-Brandt, K Breddam *Carlsberg Res. Commun.*, 1985, **50** 273
179. J Kraut *Science*, 1986, **231** 1123
180. J E Villafranca, E E Howell, S J Oatley, et al. *Philos. Trans. R. Soc. London, A*, 1986, **317** 405
181. N J Prendergas, T J Delcamp, P L Smith, J H Frusheim *Biochem.*, 1988, **27** 3664
182. M Michiro, Y Jigami, M Morikawa, H Tanaka *Biochem. Biophys. Acta*, 1987, **911** 376
183. M Muraki, M Morikawa, Y Jigami, H Tanaka *Biochem. Biophys. Acta*, 1987, **916** 66
184. N P Lawrence, S Altman *J. Mol. Biol.*, 1986, **191** 163
185. H Wolfes, J Alves, A Fliess, et al. *Nucl. Acids Res.*, 1986, **14** 9063
186. C N Cronin, B A Malcolm, J E Kirsch *J Am. Chem. Soc.*, 1987, **119** 2222
187. A R Clarke, D B Wigley, W N Chia, et al. *Nature*, 1986, **324** 699
188. A R Clarke, D B Wigley, D A Barstow, et al. *Biochem. Soc. Trans.*, 1987, **15** 152
189. C Murali, E H Creaser *Prot. Eng.*, 1986, **1** 45
190. S S Ghosh, S C Bock, S E Rokita, E T Kaiser *Science*, 1986, **231** 145
191. M Yohda, S Ohta, T Hisabori, Y Kagava *Biochem. Biophys. Acta*, 1988, **992** 156
192. G Dalbadie-McFarland, L W Cohen, A D Riggs, et al. *Proc. Natl. Acad. Sci. USA*, 1982, **79** 6409
193. I S Sigal, B G Harwood, R Arentzen *Proc. Natl. Acad. Sci. USA*, 1982, **79** 7157
194. I S Sigal, W F DeGrado, B J Thomas, S R Petteway *Biol. Chem.*, 1984, **259** 5327
195. G Dalbadie-McFarland, A D Riggs, J H Richards *Biochem. Soc. Trans.*, 1984, **12** 226
196. G Dalbadie-McFarland, J J Naitzel, J H Richards *Biochemistry*, 1986, **25** 332
197. P H Seeburg, W W Colly, D J Capon, et al. *Nature*, 1984, **312** 71
198. P J Madguvick, S G Waley *Biochem. J.*, 1987, **248** 657
199. S Gutteridge, I Sigal, B Thomas, et al. *EMBO J.*, 1984, **3** 2737
200. S K Niyagi, R S Foote, R J Murel, et al. *J. Biol. Chem.*, 1986, **261** 10087
201. F C Hartman, T S Soper, S K Miyogi, et al. *J. Biol. Chem.*, 1987, **262** 3496
202. G Voordouw, P A DeVries, W A M Van Den Berg, E P J De Clerk *Eur. J. Biochem.*, 1987, **163** 591
203. W R Tremble, P V Vitanen, H R Sarkar, et al. *Biochem. Biophys. Res. Commun.*, 1984, **119** 860
204. P V Vitanen, D R Menick, H R Sarkar, et al. *Biochemistry*, 1985, **24** 7628
205. J M Naehaus, J Soppa, J K Witgh, et al. *FEBS Lett.*, 1985, **185** 83
206. H R Sarhar, D R Menicj, P V Vitanen, et al. *J. Biol. Chem.*, 1986, **261** 8914
207. D R Menick, H R Sarkar, M S Poonian, H R Kaback *Biochem. Biophys. Res. Commun.*, 1985, **132** 162
208. R J Brooker, T H Wilson *J. Biol. Chem.*, 1986, **261** 11765
209. D R Menick, J A Lee, R J Brooker, et al. *Biochemistry*, 1987, **26** 1132
210. E Ragan, H R Sarkar, P V Vitanen, et al. *Proc. Natl. Acad. Sci. USA*, 1985, **82** 6765
211. I B Püttner, H R Sarkar, M S Poonian, H R Kaback *Biochemistry*, 1986, **25** 4483
212. H R Kaback *Biochemistry*, 1987, **26** 2071
213. F I Tsuji, S Inouye, T Goto, Y Sakaki *Proc. Natl. Acad. Sci. USA*, 1986, **83** 8107
214. T A Craig, D M Watterson, F G Prendergast, et al. *J. Biol. Chem.*, 1987, **262** 3278
215. M Nishiyama, V Matsubara, K Yamamoto, et al. *J. Biol. Chem.*, 1986, **261** 14178
216. M R Lentz, R G Webster, G M Air *Biochemistry*, 1987, **26** 5351
217. X Wli, M C Els, R G Webster, G M Air *Virology*, 1987, **156** 253
218. K N Rand, H Singhn, H C Thogersen, M J Gait *J. Biosci.*, 1985, (1) 89
219. M Singh, S Heapy, M J Gait *Prot. Eng.*, 1986, **1** 75
220. S Heapy, M Singh, M J Gait *Biochemistry*, 1987, **26** 1688
221. N Houla-Herin, H Hara, M Inouya, Y Hirota *Mol. Gen. Genet.*, 1985, **201** 499
222. K Yutani, K Ogasahara, T Tsujita, et al. *J. Biol. Chem.*, 1987, **262** 13429

223. J L Paluh, H Zalkin, D Belsch, H L Weith *J. Biol. Chem.*, 1985, **260** 1889
224. S M Samson, J L Chapman, R Belagaie, et al. *Proc. Natl. Acad. Sci. USA*, 1987, **84** 5705
225. S Hopkins, V Schirch *J. Biol. Chem.*, 1986, **261** 3363
226. S Yamano, K Tanaka, K Matsumoto, A Tohe *Mol. Gen. Genet.*, 1987, **210** 413
227. J Bulis, L D Saraswat, S S Taylor *Biochem.*, 1988, **27** 1570
228. P L Privalov *Biofizika*, 1987, **39** 742
229. W J Becktel, J A Shellman *Biopolymers*, 1987, **26** 1859
230. W Pfeil *Mol. Cell. Biochem.*, 1981, **40** 3
231. S E Lale, A M Klibanov *Biotechnol. Bioeng.*, 1983, **25** 2221
232. J F Brandts *Structure and Stability of Biological Macromolecules* (Edited by S H Timasheff, G D Fasman), Marcel Dekker, New York, 1969, (Translated into Russian), Izd. Mir, Moscow, 1973, p. 174
233. J B Matthew, R F N Gurd *Meth. Enzym.*, 1986, **130** 437
234. D P Goldenberg *J. Cell. Biochem.*, 1985, **29** 321
235. V P Torchilin, A V Maksimenko, V N Smirnov, et al. *Biochim. Biophys. Acta*, 1979, **567** 1
236. G D Kutuzova, N N Ugarova, I V Berezin *Usp. Khim.*, 1984, **53** 1852 [English translation in *Russ. Chem. Rev.*, 1984(11)]
237. H B Bull, R Breese *Arch. Biophys. Biochim.*, 1973, **158** 681
238. E Stellwagen, H Wilgus *Nature*, 1978, **275** 342
239. P K Ponnuswamy, R Muthusamy, P Manavalan *Int. J. Biol. Macromol.*, 1982, **4** 186
240. A Ikai *J. Biochem.*, 1980, **88** 1895
241. H Zuber *Structural and Functional Aspects of Enzyme Catalysis* (Edited by H Eggerer, R Huber), Springer Verlag, Berlin–Heidelberg–New York, 1981, p. 114
242. D P Ward, M Moo-Young *Biotech. Adv.*, 1988, **6** 39
243. D J Merkler, G K Farrington, F C Wedler *Int. J. Pep. Protein Res.*, 1981, **18** 430
244. A Atkinson *J. Appl. Chem. Biotechnol.*, 1976, **26** 577
245. G D Kutuzova, N N Ugarova, I V Berezin *Itogi Nauki i Tekhniki. Ser. Biotekhnologiya* (Advances in Science and Engineering. Biotechnology Series), Izd. VINITI, Moscow, 1986, Vol. 5, p. 5
246. E Stellwagen *Enzyme Eng.*, 1984, **234** 1
247. C Chotia *Nature*, 1975, **254** 304
248. P A Bash, U C Singh, R Langridge, P A Kollman *Science*, 1987, **236** 564
249. R Wetzel *Trends Biotechnol.*, 1987, **12** 478
250. B S Hartley *Philos. Trans. R. Soc. London, A*, 1986, **317** 321
251. J Richardson *Adv. Prot. Chem.*, 1981, **34** 154
252. M W Pantoliano, R C Ladner, P N Bryan, et al. *Biochemistry*, 1987, **26** 2077
253. P Argos *J. Mol. Biol.*, 1987, **197** 331
254. D S C Yang, M Sax, A Chakrabartty, C L Hew *Nature*, 1988, **333** 232
255. C Mitchison, R L Baldwin *Proteins: Struct. Funct. Genet.*, 1986, **1** 23
256. E Qucrol, A Parrilla *Enz. Microb. Technol.*, 1987, **9** 238
257. L J Perry, R Wetzel *Science*, 1984, **226** 555
258. B W Matthews *Biochemistry*, 1987, **26** 6885
259. K A Dill *Prot. Eng.*, 1987, **1** 369
260. M Rechsteiner, S Rogers, K Rote *TIBS*, 1987, 390
261. K V Rote, M Rechsteiner *J. Biol. Chem.*, 1986, **261** 15430
262. R Katznelson, R G Kulka *Eur. J. Biochem.*, 1985, **146** 437
263. S Rogers, R Wells, M Rechsteiner *Science*, 1986, **234** 364
264. J M Baker, J F Dice *Proc. Natl. Acad. Sci. USA*, 1986, **83** 5830
265. A Bachmair, D Finley, A Varshavsky *Science*, 1986, **234** 179
266. E E Stadman *Trends Biochem. Sci.*, 1986, **11** 11
267. M G Grütter, T M Grag, L H Weaver, et al. *J. Mol. Biol.*, 1987, **197** 349
268. T Alber, T M Gray, L H Weaver, et al. *NATO ASI Ser. A*, 1987, **126** 251
269. T Alber, M G Grütter, T M Gray, et al. *UCLA Symp. Mol. Cell. Biol. New Ser.*, 1985, **39** 307
270. T M Alber, Doo-pin Sun, J A Nye, et al. *Biochemistry*, 1987, **26**(13) 3754
271. D Shortle, B Lin *Genetics*, 1985, **110** 539
272. M Matsumura, Sh Rataoka, Sh Aiba *Mol. Gen. Genet.*, 1986, **204** 355
273. K Yutani, K Ogasahara, Y Sugino *Adv. Biophys.*, 1985, **20** 13
274. B C Cunningham, J A Wells *Prot. Eng.*, 1987, **1** 319
275. R B Wetzel *Eur. P. Appl.*, 1986, **155** 832
276. L J Perry, R Wetzel *Biochemistry*, 1986, **25** 733
277. L J Perry, R Wetzel *Prot. Eng.*, 1987, **1** 101
278. J E Villafranca, E E Howell, D H Voet, et al. *Science*, 1983, **222** 782
279. J E Villafranca, E E Howell, S J Oatley, et al. *Biochemistry*, 1987, **26** 2182
280. R T Sauer, K Hehir, R S Stearman, H A Weiss, et al. *Biochemistry*, 1986, **25** 5992
281. C O Pabo, E G Suchanek *Biochemistry*, 1986, **25** 5987
282. M H Heccht, J M Sturtevant, R T Snuer *Proteins: Struct. Funct. Genet.*, 1986, **1** 43
283. T Imanaka, M Shibazaki, M Takagi *Nature*, 1986, **324** 695
284. K Yutani, H Akutsu, K Ogasahara, et al. *Biochemistry*, 1987, **26** 5666
285. M Matsumura, S Yasumura, S Aiba *Nature*, 1986, **323** 356
286. D A Estell, T P Graycar, J A Wells *J. Biol. Chem.*, 1985, **260** 6518
287. D B Wigly, A R Clark, C R Dunn, et al. *Biochim. Biophys. Acta*, 1987, **916** 145
288. L Ramdas, F Sherman, D T Nall *Biochemistry*, 1986, **25** 6952
289. L Ramdas, B T Nall *Biochemistry*, 1986, **25** 6959
290. A Gilles, I Saint-Girons, M Monnot, et al. *Proc. Natl. Acad. Sci. USA*, 1986, **83** 5798
291. C R Matthews, M M Crisanti, J T Manz, G L Gepner *Biochemistry*, 1983, **22** 1445
292. A M Beasty, M R Hurle, J T Manz, et al. *Biochemistry*, 1986, **25** 2965
293. K M Perry, J J Onuffer, N A Touchette, et al. *Biochemistry*, 1987, **26** 2674
294. J I Casal, T J Ahern, R C Davenport, et al. *Biochemistry*, 1987, **26** 1258
295. T J Ahern, J I Casal, G A Petsko, A M Klibanov *Proc. Natl. Acad. Sci. USA*, 1987, **84** 675
296. R Bott *World Biotechnol. Rep.*, 1985, **2** 51
297. A J Russel, A R Fersht *Nature*, 1987, **328** 496
298. S G Sligar, K D Egeberg, Y T Sage, et al. *J. Am. Chem. Soc.*, 1987, **109** 7896
299. L C S Ferreira, V Schwarz, W Keck, et al *Eur. J. Biochem.*, 1988, **171** 11
300. I Crenn, D Ladant, N Guiso, et al. *Eur. J. Biochem.*, 1986, **159** 605
301. H M Hsiung, N G Mayne, G W Becker *Biotechnology*, 1986, **4** 991