# MODERN STATE AND PERSPECTIVES OF X-RAY CRYSTALLOGRAPHY OF PROTEINS

*B. K. VAINSHTEIN*

Institute of Crystallography, USSR Acad. Sci, Moscow

A quarter of a century has gone opening for us a fascinating world of protein structures, starting from the remarkable works of John Kendrew, Max Perutz and their associates on the structure of myoglobin and hemoglobin in the end of 50s and beginning of 60s. These works demonstrated a great value of information taken from the analysis of the three-dimensional structure for understanding mechanisms of protein functioning. The present-day molecular biology is unconceivable without permanent and intensive exploitment of X-ray biocrystallography data.

The knowledge of the spatial structure of various proteins provides a wealth of basic information for understanding life processes from the standpoint of the general laws of protein structure and their functions and for establishment of particular mechanisms governing the functioning of different protein molecules – enzymes, receptors, hormones, immunoglobulins, etc.

At present more than 400 protein structures and more than 1000 their varieties with different pseudosubstrates and ligands are determined.

A protein molecule is a regularly (in accordance with genetic program) built complicated physical system consisting of $10^3$ – $10^5$ atoms arranged in a definite way along a folded polypeptide chain. From the physical standpoint, a protein molecule is a small solid aggregate of atoms, but not periodic as a crystal. It has a complicated spectrum of thermal vibrations which differ for various parts of the molecule. Protein molecule may have various spontaneous or induced conformational transformations accompaned by certain changes in its chemical and electronic structure mainly of its active center.

The basic, though simplified but in many cases effective way to describe the structure and interaction of the protein molecule with the substrate or pseudosubstrate, with other molecules, e.g. DNA, or small

molecules, such as ligands, drugs, etc, is the geometrical analysis. The description is realized in geometric models or by computer graphics methods.

At present X-ray protein analysis is a field of knowledge summarizing the latest achievements in biochemistry, molecular biology, experimental physics and computer technology.

X-ray protein crystallography data, complimented, if necessary, by spectroscopy, EXAFS, NMR and other methods, can, in a number of cases, present a convincing picture of the protein molecule functioning.

Enzymatic reactions run with a rather high speed ranging from $10^{-5}$ to several seconds per catalytic act. So it is unrealistic to hope that all stages of enzyme-substrate complex conformational rearrangement may be observed by X-ray analysis (aside of some special cases of relatively slow reactions).

These difficulties may be overcome by using instead of a true substrate, pseudosubstrates or inhibitors which under certain conditions enter the active site of the enzyme and block the enzyme in some intermediate stage. Such an enzyme-inhibitor complex may imitate a certain stage of the enzymatic reaction allowing one to observe an intermediate conformation of the active site.

Great possibilities are opened by the mathematical modelling of the protein structure dynamics. It becomes possible to establish the substrates of protein molecules. These calculations, though, require very powerful computers. X-ray investigation not only gives the positions of atoms in the protein molecule, but in many cases allows one to build a model of the protein functioning and its interaction with other molecules, e.g. RNA, DNA, and small molecules.

In our laboratory we study various proteins, among them leghemoglobin, various catalases, aspartate-transaminase, several ribonucleases, pyrophosphatase, thermitase and many others. We are especially interested in the heme-containing proteins and some enzymes, mostly ribonucleases. The achieved resolution ranges from 3 Å to 1.2 Å for small proteins.

I shall start with the results on leghemoglobin.

With the formation of the oxygen atmosphere on the Earth about two billion years ago the living organisms acquired proteins whose function is to store and transfer oxygen. In animal organisms they are of the mioglobin-hemoglobin type. A similar protein –leghemoglobin– was found in the root nudels of nitrogen-fixing leguminous plants, such as soya, pea, lupine, etc. (molecular mass 17000).

We have studied leghemoglobin extracted from lupine root nudels. The monoclinic unit cell has the following parameters: a = 92.95 Å, b = 38.31 Å, c = 52.15 Å, $\beta = 98^{\circ}\ 50'$. The Lg molecule has the mioglobin-type folded chain with eight characteristic $\alpha$-helical segments. The most

important functional parts of the molecule are the heme and the amino acid residues in the heme pocket (Fig. 1)[1]. The iron atom of the heme is reversibly attached to the oxygen molecule. The attachement of the oxygen molecule is accompanied by the changes in the spin state of outer valence electrons of the Fe atom. The Fe atom is surrounded with four N atoms of the porphyrin ring, the fifth ligand is the N atom of the F 11 amino acid residue, the so-called proximal histidine. On the other, so-called distal, side of the heme an $O_2$ molecule (the 6th ligand) is attached to the Fe atom. The Fe-proximal histidine bond is weakened, whereas the Fe atom itself is displaced from the porphyrin ring plane towards the distal side. The Fe-$N_{(prox)}$ distance is different from those in other globins (2.0-2.2 Å) and equals 2.32 Å in leghemoglobin. The iron atom of Lg may be either in ferrous ($Fe^{2+}$) or in ferric ($Fe^{3+}$) state. To study the binding

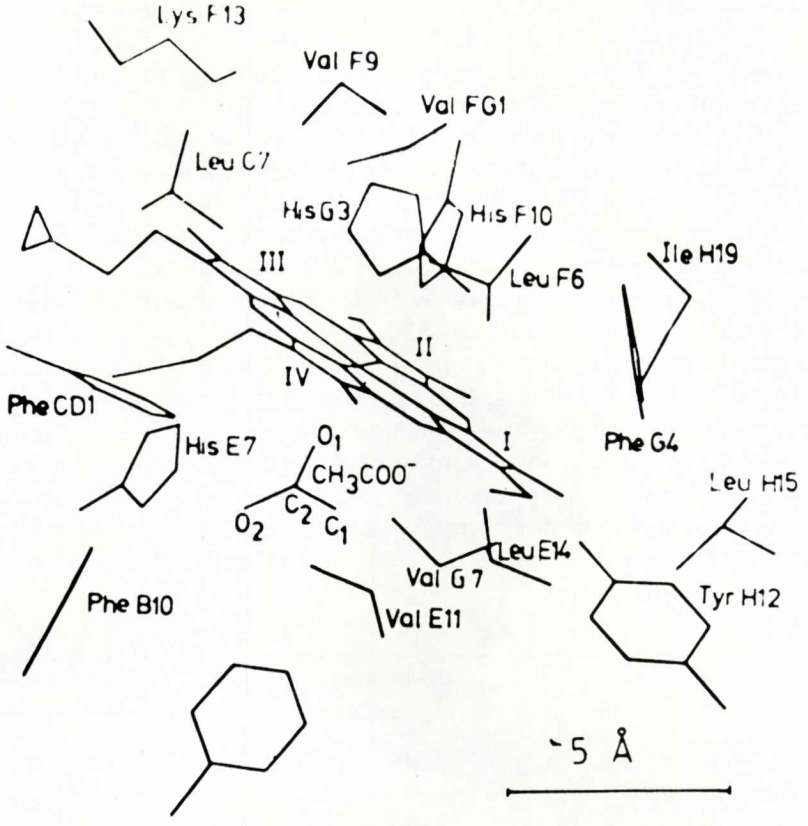


Fig. 1. Structure of the heme pocket, surroundings of the heme with the nearest amino acid residues, and position of the acetate group.

ability of the Fe atom in leghemoglobin, various molecules were used as the sixth ligand, namely, $H_2O$ (deoxyform), fluorine ion ($F^-$), acetic acid ($CH_3COO^-$), cyanide ($CN^-$), nicotinic acid ($NC_5H_4COO^-(Fe^{3+})$), and nitrobenzoate ($C_6H_5NO$ ($Fe^{2+}$)). We have also investigated oxy-Lg, i.e. Lg-$O_2$ complex[2].

In the structure of these complexes the following regularity is observed: the Fe atom displacement towards the distal side is greater with the increase of electron density transfer of the Fe atom. The deviation of the iron atom from the mean heme plane (towards the proximal side) ranges from –0.08 Å in deoxyleghemoglobin to + 0.11 Å in low spin complexes. Besides, it was established that the heme pocket in Lg is larger than in other globins, which is clearly seen from the data on the Lg complex with nicotinic acid.

On transition from deoxy– to the oxyform the iron atom position is displaced from the heme plane, the displacement Fe ($\Delta$Fe) is analogous to that in erythrocruorin (Er), but opposite to that in myoglobin (Mg), which explains the difference in kinetical constants ("on" and "off" rate constants of the oxygen complex are shown in Table 1.

Thus it was established that the main factor conditioning the high affinity of Lg to oxygen is the weakening of the Fe-proximal histidine bond and displacement of the Fe atom from the heme plane towards the ligand.

We have determined, by X-ray analysis, the structure of inorganic yeast pyrophosphatase *(Saccharomyces cerevisiae)* to a 3 Å resolution. The molecular mass of the dimer is 64000 dalton, the unit cell parameters are: a = 52.2 Å, b = 70.3 Å, c = 95.5 Å, $\gamma$ = 99.6°, space group $P2_1$. The catalytic function of the enzyme is decomposition of inorganic pyrophosphate with 3 Mg ions. It has been shown that according to the polypeptide chain packing pyrophosphatase belongs to the *($\alpha$ + $\beta$)* proteins. Most of its globule is formed of antiparallel β-strands forming two β-sheets and β-barrel type structure[3].

The active complex of pyrophosphatase with $Mg^{2+}$ ions and the inhibitor complexes with Zn, Co, Cd were studied by the difference Fourier

**Table 1.**

| | | Sperm whale-myoglobin | Erythro-cruorin | Leghemo-globin |
|---|---|---|---|---|
| Association K' ($M^{-1}$ $sec^{-1}$) | "on" | 19 | 300 | 118 |
| Dissociation K ($sec^{-1}$) | "off" | 10 | 218 | 4.4 |
| $\Delta$Fe(Å) | deoxy | 0.42 | 0.17 | 0.048 |
| | oxy | 0.18 | 0.38 | 0.30 |

technique. The non-splitting substrate analogue, Ca-pyrophosphate, was also localized in the active site region. Four metalbinding centers were found in each subunit, all of them being in the active site cavity near the substrate-binding region. These are Arg 77, Tyr 191, Lys 97, Glu 58 and some others. On the basis of these results a tentative mechanism of pyrophosphatase bond splitting by the enzyme was suggested[4].

We have also studied the structure of manganese-phosphate complex (MnP) of pyrophosphatase by the molecular replacement method. MnP crystals belong to the $P2_1\ 2_1\ 2_1$ space group and have the following unit cell parameters: 116.4 x 106.3 x 56.1 Å. Anomalous scattering data at 5 Å resolution were used to localize the binding site of the three $Mn^{2+}$ ions in the active center of each subunit in the enzyme molecule (Fig. 2). Refinement of the MnP complex spatial structure was carried out. It was shown that the active center of each subunit includes 3 $Mn^{2+}$ ions and 2 phosphate ions, held by the ligands represented by carboxyl groups of the aminoacid residues (Fig. 3). Independent crystallographic refinement of the structure of 2 subunits of inorganic yeast pyrophosphatase apoenzyme was performed[5]. At present we have the resolution of 6 –2.35 Å, the crystallographic R-factor 25.4%, correlation coefficient 87.9%, mean square deviation from standard bond lengths 0.028 Å. On the basis of these data, comparative analysis of the conformation of the apoenzyme molecule and the MnP complex was carried out. It was shown that mean square deviation of the $C_\alpha$ site for the main chain atoms amounted to approximately 0.65 Å for various subunits. We found that in the majority of cases the re-

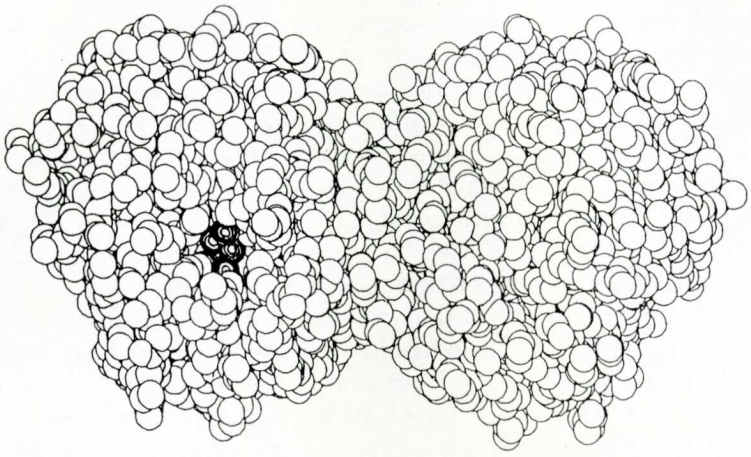

Fig. 2. Molecule of the MnP complex of inorganic yeast pyrophosphatase. In the active site of each subunit 2 Mn ions and 2 phosphate groups are observed.

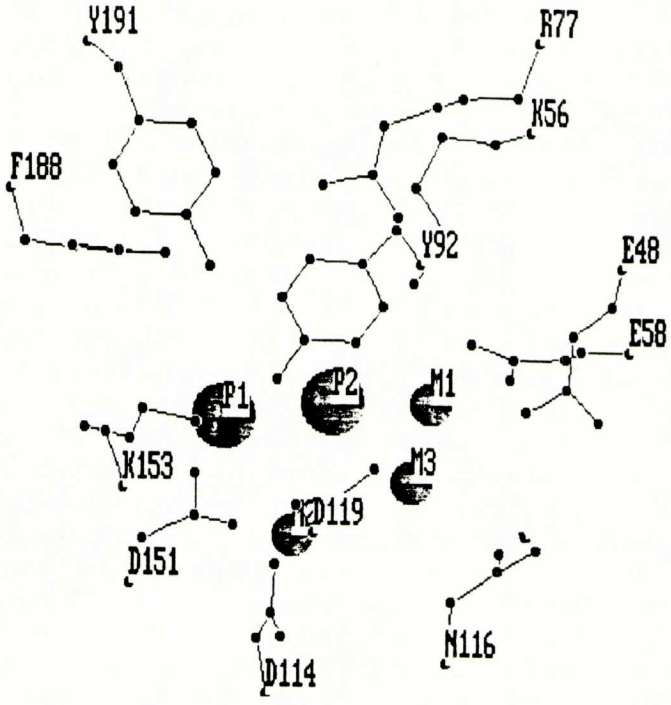


Fig. 3. Surroundings of the manganese (M) and phosphate (P) ions in the active site of inorganic yeast pyrophosphatase.

gions with different conformation coincide with the contact sites between the molecules.

There is much interest to the structural study of proteins involved in interaction with nucleic acids. We have studied a number of specific ribonucleases, which may also be included in this class of proteins.

A specific RNase recognizes the site in RNA chain where the sugar-phosphate backbone ought to be cut. The substrate is fixed by the enzyme in the active site region in such an environment of the protein catalytic groups which provides the catalytic process.

Besides the classic pancreatic RNase A, a number of microbial and fungal RNAses are actively studied in different laboratories for the last 5-7 years. Now the structure of 10 proteins of this family is known, five of them are solved in our lab. They are proteins of molecular mass 10,000 or 12,000, with 100-110 amino acid residues, with specificity to guanylic bases. These are fungal RNAses $C_2$, $Ap_1$, $Pb_1$, $Th_1$, and bacterial RNAse Bi.

Crystallographic data on these structures are shown in Table 2. We

**Table 2.**

| RNase, M. W. (D) | Space group | molecules per asym. part | Unit cell, Å | | Resolution, Å | R, % | Ref. |
|---|---|---|---|---|---|---|---|
| $C_2$ *(Asp. clavatus)* | $P2_1$ | 1 mol. | a = 30.6; c = 49.7; | b = 32.0 $\gamma = 116.1^\circ$ | 1.35 | 17.6 | [6] |
| 11 000 | $P2_1$ | 2 mol. | a = 31.6; c = 57.3; | b = 51.0 $\gamma = 93.6^\circ$ | 1.75 | 19.0 | |
| $Ap_1$ *(Asp. palidus)* 11 000 | $P2_1$ | 1 mol. | a = 30.7; c = 49.8; | b = 32.0 $\gamma = 115.8^\circ$ | 1.17 | 18.6 | |
| $Pb_1$ *(Pen. brevi-compactum)* 11 000 | I222 | 1 mol. | a = 97.8; c = 33.9 | b = 51.7 | 1.4 | 18.0 | [7] |
| + nucleotide | | | | | 1.24 | 19.4 | |
| $Th_1$ *(Trichoderma harzianum)* 11 000 | $P3_121$ | 1 mol. | a = b = 55.7 c = 80.0 | | 1.75 | 16.4 | [8] |
| Bi *(Bac. intermedius)* | B2 | 2 mol. | a = 114.5; c = 33.3; | b = 78.9 $\gamma = 119.2^\circ$ | | | [9] |
| 12 300 | $P2_12_12_1$ | 1 mol. | a = 111.4; c = 33.5 | b = 69.6 | 2.0 | 14.8 | |
| + nucleotide | | | | | 2.0 | 15.0 | |

managed to reach a rather high resolution of 1.2-2.0 A, which made it possible to distinguish individual atoms. The structures were solved by multiple isomorphous replacement (MIR) or by molecular replacement based on similarity with the structure of a previously solved protein.

The polypeptide chain course for guanyl-specific RNAses is shown in Fig. 4. These proteins are of ($\alpha + \beta$) type, i.e. they have α-helices and β-sheets. The spatial organization of β-sheet in fungal and bacterial enzymes shares many common features. It is a twisted β-sheet of 4-5 strands, screened by one or two α-helices. Active site amino acid residues are distributed in the β-sheet and the neighbouring parts of the structure.

The structure of the RNAse $Pb_1$ complex with the reaction product (guanylic nucleotide) determined and refined at a high resolution, made it

possible to reveal a pattern of interactions betwen RNAse $Pb_1$ and the guanylic base (Fig. 5). A high specificity towards guanylic bases is determined by the ability to form 5-6 specific hydrogen bonds.

It is very interesting to examine thermal vibrations in the molecule. Low values of thermal vibrations are observed for atoms in the molecule

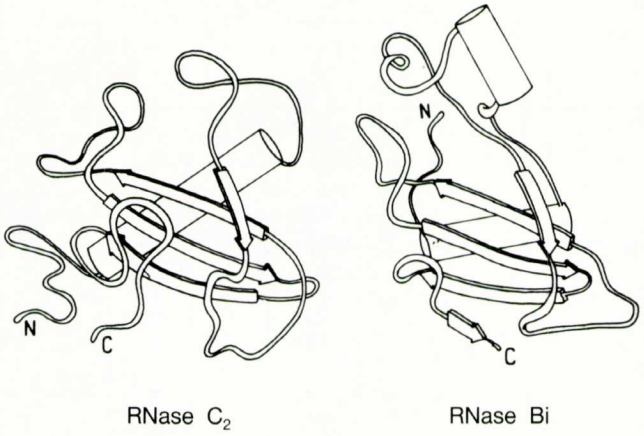


Fig. 4. The polypeptide chain course for fungal and bacterial RNases.

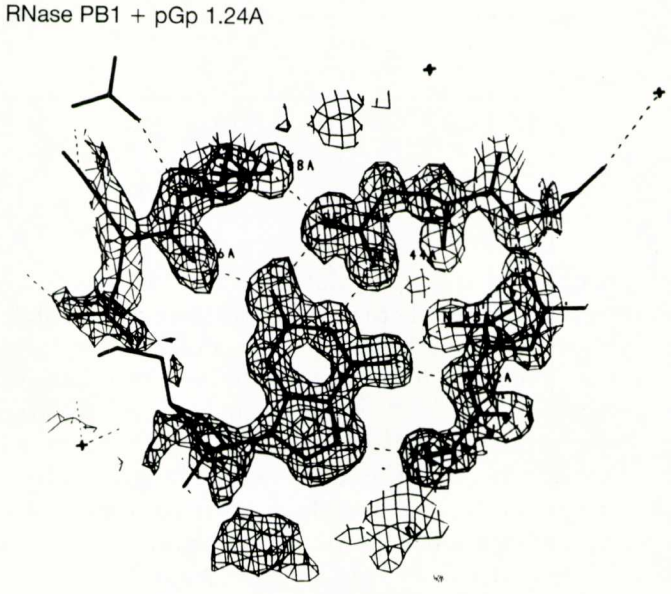


Fig. 5. Part of the electron density map of the RNase $Pb_1$–3'–GMP–complex at 1.24 Å resolution.

base or the active centre, high values – for surface or pseudosubstrate atoms (Fig. 6).

We have also studied crystals of various catalases. Table 3 presents information about crystals and structure investigations of catalases from different species, the spatial structure of which is fairly well known. The first structure to be determined in our laboratory was the structure of fun-

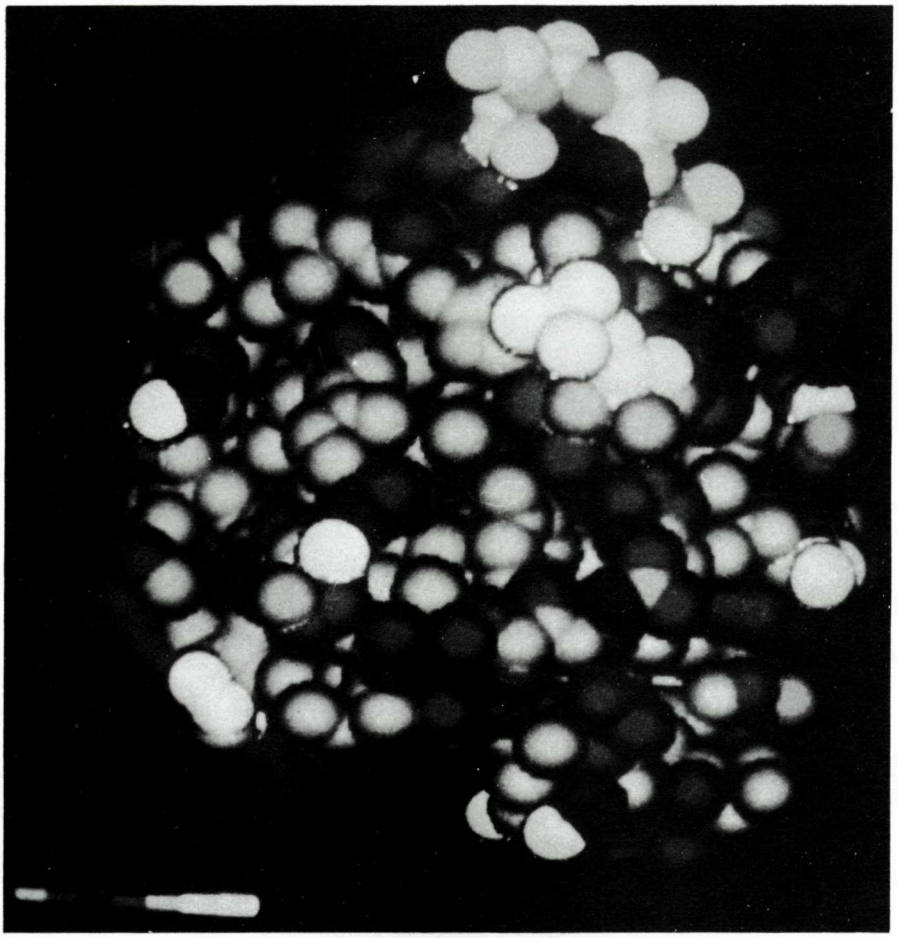

Fig. 6. Atomic model of RNase $Pb_1$. The atoms colours correspond to the individual thermal vibrations:

$\sqrt{u^2}$ = 0.4 Å – blue

$\sqrt{u^2}$ = 0.8 Å – yellow

**Table 3.**

| Catalase of | Mol. mass (kDa) | Space group | Unit cell (Å) a | b | c | Mol. in a. u. | X-ray resol. (Å) | Ref. |
|---|---|---|---|---|---|---|---|---|
| Beef liver (BLC) | 250 (4 x 62) | $P2_12_12_1$ | 88 | 140 | 233 | 1 | – | |
| | | $P3_121$ | 178 | 241 | | 1 | – | [10] |
| | | $P3_221$ | 142.3 | 104.0 | | 1/2 | 2.5 | |
| *Penicillium vitale* (PVC) | 290 (4 x 72) | $P3_121$ | 144.4 | 133.8 | | 1/2 | 2.0 | [11] |
| *Micrococcus lysodeikticus* (MLC) | 230 (4 x 58) | $P4_22_12$ | 106.6 | 106.2 | | 1/4 | 3.0 | [12] |
| *Thermus thermophilus* (TTC) | 210 (6 x 35) | $P2_13$ | 133.4 | | | 1/3 | 3.0 | [13] |

gal catalase *Penicillium vitale* (PVC) (molecular mass 290,000)[14]. Experimental data to 2.0 Å resolution were obtained with the use of synchrotron radiation source at the European Molecular Biology Laboratory.

Spatial position of the elements of the secondary structure of PVC subunit is shown in Fig. 7. The polypeptide chain which comprises the PVC subunit contains 670 residues. The first 56 residues are situated apart from the subunit globule being involved in many contacts with amino acid residues of neighbouring subunits. The polypeptide chain forms three domains. The largest domain I consists of 300 residues and contains a β-barrel of eight anti-parallel strands forming a surface of the closed near-cylindrical β-sheet, in which strands are interchanged with helical segments. Then the polypeptide chain runs as an irregular segment of about 70 residues, connecting the domain I with the smaller domain II. This domain contains 70 residues forming four α-helices, which are positioned similar to the last four helices in the globins.

The third C-terminal domain III contains 150 residues forming a sheet of five parallel β-strands and four α-helices above and below the β-sheet. Its topology is similar to that of flavodoxin.

The heme group in catalase is positioned deep inside the molecule in a large domain, so that the iron atom is at a distance of about 17 Å from the molecule surface.

The position of amino acid residues in PVC active site is shown in Fig. 8. On the proximal side a close contact with the heme iron atom is

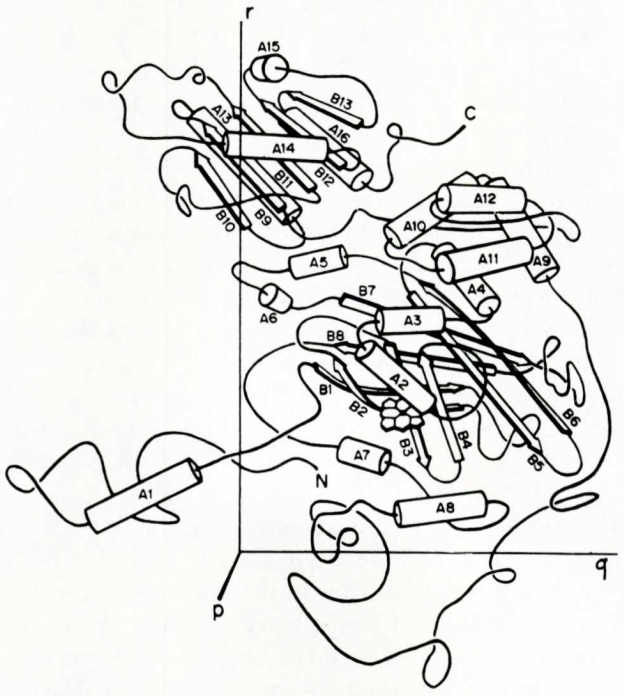


Fig. 7. Arrangement of the polypeptide chain and elements of secondary structure in a PVC subunit. α-helices and β-strands are represented by cylinders and arrows respectively. p, q and r are 2-fold axes of the molecule.

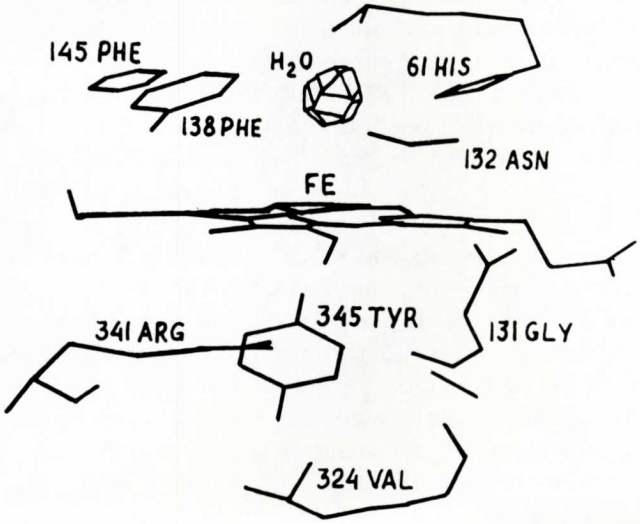


Fig. 8. Positions of aminoacid residues in PVC active site.

made by a tyrosin residue whose phenolic group occupies the fifth coordination position. On the distal side, the active site is formed by residues His 61, Phe 138, Asn 132, Ser 99. On the electron density map of PVC there is an isolated region of positive density close to the iron atom, which can be interpreted as a bound water molecule, occupying, probably, the place of a hydrogen peroxide molecule.

A comparison of the spatial structures of PVC and beef liver catalase (BLC), carried out in collaboration with M. Rossmann and his colleagues, turned out to be very interesting. Conformation of the polypeptide chain around the active site was found to be very similar, whereas the N-terminal and C-terminal parts showed differences. The most remarkable is the presence of "flavodoxin" domain III in the C-terminal part of the PVC subunit, which is completely absent in BLC[15].

Thus two catalases isolated from organisms widely diverged in the long course of biological evolution proved to have a very similar structure of domains I and II, but only PVC has domain III.

Detailed analysis of the spatial structures of the two catalases and the spacings between the atoms in the active site allowed M. Rossmann and I. Fita to suggest a tentative mechanism for the enzymatic activity of BLC[16].

Recently we obtained and investigated crystals of bacterial catalases *Micrococcus lysodeikticus* (MLC). The crystals are tetragonal, space group $P4_2\,2_1\,2$, unit cell parameters are: a = b = 106.7, c = 106.3 Å. Resolution 3 Å. Molecular mass 240,000. Three heavy atom derivatives were obtained. Analysis of the electron density maps showed that like BLC, MLC contains two domains, and, unlike PVC, contains no flavodoxin-like domain. The arrangement of the heme group and α-helices in MLC is very close to their location in PVC and BLC, the differences in atom positions being not larger than 1-2 Å. But it should be noted that in this case there is no initial α-helix (24 residues) of BLC[12].

I would like to note that we have discovered an absolutely unknown new type of catalases – *Thermus Thermophillus* catalase (TTC), which is absolutely different in structure. It turned out that the enzyme molecule consists of 6 identical subunits and has 32 symmetry. Unexpectedly, this catalase contains no heme group and the active site, located between the 4 α-helices, contains 2 Mn atoms placed at a distance of 3.6 Å[13].

The above examples illustrate the importance of the knowledge of the spatial structure for understanding the functioning of protein molecules. In many instances these data provide the interpretation of the molecule state and its conformational rearrangements in enzymatic and other reactions. It is also of interest that the active site of the molecule is built by 10-15% of all the amino acid residues of the protein. Thus it may be assumed that the remaining part of the molecule is its basement, the matrix on

which functionally active parts of the molecular robot, the protein, do work. The additional functions of the "basement" with charged ions distributed over it is to generate a certain electrostatic field at the active site. Moreover, inside the molecule body the directional system of thermal motions is formed that may promote pulling of the substrate into the active site and forcing out of it the reaction products. The mobility of some groups on the surface of a protein molecule allows one to interprete their antigene properties.

Thus a protein molecule is in fact a molecular robot, the smallest natural machine operating on stereochemical and electronic principles.

It should be kept in mind that the methods of "production" and structure formation of protein molecules were continuously refined and "polished" by natural selection for hundreds billions of years of the biological evolution on the Earth. The science penetrates the laws of the structure and synthesis of biomolecules at a much higher rate.

The detailed analysis of the protein structures permits now biochemists to substitute one or several aminoacids in the active center of natural proteins by other aminoacids, improving their functioning and making proteins more efficient. Thus the methods of gene engineering permit us not only to develop new methods of mass production of natural proteins but also methods of improving their functions, which have opened up new vistas for scientific "design" of new proteins that only recently seemed to be quite unrealistic.

Abstract

The aims and principal possibilities of X-ray analysis of proteins are considered. Detailed description of the structure of *Lupinus Luteus Leghemoglobin* and the mechanism of oxygen binding by this protein is given. Analysis of the structure of inorganic yeast pyrophosphatase *(Saccharomyces cerevisiae)* has been carried out. High-resolution structure study of five proteins in the family of specific microbial and fungal, ribonucleases has been made. Description is given of structure determination of complex proteins – catalases from various sources: *Penicillium vitale* (m. w. 290 kDa), *Micrococcus lysodeikticus* (m. w. 240 kDa), *Thermus thermophillus* (m. w. 210 kDa). The results obtained allow to explain the mechanism of enzymatic reactions catalysed by these proteins.

## REFERENCES


1. Vainshtein B. K. The structure of leghaemoglobin. In: "Structural studies of molecules of biological interest". Ed. by G. Dodson, J. P. Gluskr & D. Sayre, Clarendon Press, Oxford, 1981, pp. 310-327.
2. Obmolova G. V., Safonova T. N., Teplyakov A. V. et al. Structure of the complexes of yellow lupine ferrous leghemoglobin with CO and NO at 1.8 Å resolution. Bioorganicheskaya khimiya, 1988, v. 14, N 11, pp. 1509-1519 (in Russian).
3. Terzyan S. S., Voronova A. A., Smirnova E. A. et al. Structure of inorganic yeast pyrophosphatase at 3.0 Å resolution. Bioorganicheskaya khimiya, 1984, v. 10, N 1, pp. 1469-1482 (in Russian).
4. Kuranova I. P., Terzyan S. S., Voronova A. A. et al. X-ray structure of the active site of inorganic yeast pyrophosphatase. Bioorganicheskaya khimiya, 1983, v. 9, N 12, pp. 1611-1619 (in Russian).
5. Chirgadze N. Yu., Kuranova I. P., Strokopytov B. et al. Refined structure of MnP Complex of Yeast Inorganic Pyrophosphatase at 2.7 Å Resolution. XIIth Eur. Crystallogr. Meeting, Abstracts. Moscow, USSR, v. 2, p. 361.
6. Polyakov K. M., Strokopytov B. V., Vagin A. A. et al. Crystallographic refinement and the atomic model or ribonuclease C at 1.55 Å resolution. Kristallografiya, 1987, v. 32, N 4, pp. 918-926 (in Russian).
7. Pavlovsky A. G., Strokopytov B. V., Borisova S. N. et al. Structure of *Penicillium vrevicompactum* ribonuclease complex with 3,5-guanzine diphosphate at 1.5 Å resolution. Dokl. AN SSSR, 1987, v. 292, N 5, pp. 1253-1256 (in Russian).
8. Polyakov K. M., Strokopytov B. V., Vagin A. A. et al. In: Metabolism and Enzymology of Nucleic Acids. Plenum Press, N. Y. and London, 1988, pp. 223-225.
9. Pavlovsky A. G., Sanishvili G.G., Borisova S. N. et al. Structure of two crystal modifications of *Bacillus intermedius* ribonuclease. Kristallografiya, 1989, v. 34, N 1, pp. 137-142 (in Russian).
10. Murthy M. R. N., Reid T. J., Siciguano A., Tanaka N., Rossmann M. S. Structue of beef liver catalase. J. Mol. Biol., 1981, v. 152, pp. 465-499.
11. Vainshtein B. K., Melik-Adamyan W. R., Barynin V. V., et al. Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. J. Mol. Biol., 1986, v. 188, pp. 49-61.

12. YUSIFOV E. F., GREBENKO A. I., BARYNIN V. V. et al. Spatial strcuture of catalase from *Micrococcus lysodeikticus* at 3.0 Å resolution. Kristallografiya, 1989, v. 34, pp. 1451-1456 (in Russian).
3. VAINSHTEIN B. K., MELIK-ADAMYAN W. R., BARYNIN V. V. et al. Spatial organization of catalase proteins. Proc. Int. Biomol. Struct. Interactions, Suppl. J. Biosci., 1985, v. 8, pp. 471-479.
14. VAINSHTEIN B. K., MELIK-ADAMYAN W. R., BARYNIN V. V. et al. Three-dimensional structure of the enzyme catalase. Nature (London), 1981, v. 293, pp. 411-412.
15. MELIK-ADAMYAN W. R., BARYNIN V. V., VAGIN A. A. et al. Comparison of beef liver catalase and *Penicillium vitale* catalase. J. Molec. Biol., 1986, v. 188, pp. 63-72.
16. I. FITA, M. G. ROSSMANN. The active center of catalase. J. Molec. Biol., 1985, pp. 21-37.