ISSN 0023-1584, Kinetics and Catalysis, 2007, Vol. 48, No. 4, pp. 469–478. © MAIK "Nauka/Interperiodica" (Russia), 2007.
Original Russian Text © S.D. Varfolomeev, 2007, published in Kinetika i Kataliz, 2007, Vol. 48, No. 4, pp. 499–509.

MECHANISMS OF CATALYTIC REACTIONS

Systemic Approach in the Physical and Structural Chemistry of Enzyme Catalysis: From Primary Structure to Elementary Event

S. D. Varfolomeev

Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, 117977 Russia
Moscow State University, Moscow, 119899 Russia

Received November 1, 2006

Abstract—The state of the art in enzyme catalysis is considered in terms of physical and structural chemistry. The main chemical kinetic and structural approaches are presented that can provide detailed information concerning the elementary processes making up the multistep catalytic cycle of molecular conversion at the active site of an enzyme. It is demonstrated that knowledge of the sequence of amino acids in a protein is sufficient to reconstruct the tertiary structure of this protein, to identify the catalytic groups, and to elucidate the molecular mechanism of catalysis. This approach is based on highly efficient information and computational technologies. The architecture of the active sites of enzymes is analyzed, including geometric invariants and the characteristic bond distances and angles of catalytic groups. The template method for identifying catalytic sites in the protein 4D structure is considered. The potential of molecular mechanics in the study of active sites is illustrated by the example of computer-simulated mutagenesis. Quantum chemical calculations applied to elementary events of the catalytic cycle are considered as a physical basis for understanding the catalytic mechanism and the origin of the efficiency and specificity of enzymes.

DOI: 10.1134/S0023158407040039

It will soon be 200 years since the St. Petersburg scientist K.S. Kirkhgof discovered in 1814 that a chemical reaction can be sped up by a biological substance. Since then the notion of the nature of biological (enzyme) catalysis has undergone profound evolution. Now enzyme catalysis is properly understood from the standpoint of fundamental science and is widely used in various applications.

The capability of biomolecules to speed up chemical reactions was of interest to most of the great scientists of the 19th century. J. Berzelius, one of the founding fathers of modern chemistry and the originator of catalysis, wrote in 1836 that thousands of catalytic reactions between tissues and liquids take place in plants and animals to yield numerous products from the same starting material [1, 2].

Wilhelm Ostwald, who formulated the basic concepts of catalysis, classified the phenomena of reaction speedup under the action of a "third body" into homogeneous catalysis, heterogeneous catalysis, the effect of "nuclei," and enzyme catalysis. It has been clear since the earliest catalytic studies that enzyme catalysis has the following unique features:

(1) Enzymes are generally 10^{12}–10^{15} times as effective as classical catalysts, such as the hydrogen ion.

(2) The protein molecules that possess the property of enzymatic activity can "recognize" reactant molecules and selectively carry out reactions between molecules of certain structure. The latter property is of extreme importance for biological systems since it ensures the directivity of the chemical changes in complex, multicomponent mixtures of biological compounds.

The last 20–30 years have witnessed a historic breakthrough in the understanding of enzyme catalysis and in a wide variety of enzyme applications. The current advances in this field are due to the great amount of information now available on the protein structure, to the progress in the kinetic and mechanistic studies of enzymatic reactions, to the development of methods allowing the protein structure to be manipulated by modifying the encoding gene, and to the use of novel computer-based informatic and computational tools. The latter are of particular significance. It is the creation, development, and active use of methods for storing and processing large amounts of information and the possibility of performing extensive computations at a high count rate that have made possible the breakthrough in enzyme science and technology.

Detailed studies and use of enzymes gave birth to a variety of areas of modern science and ensured their development and success. An enzyme is a precise and highly efferent tool for carrying out complicated reactions that require accurate cleavage or establishment of certain chemical bonds. Genetic engineering, which has made possible directed genetic modifications of organisms and the synthesis of the necessary amounts

of proteins with useful properties, is based on the capability of enzymes to selectively hydrolyze certain bonds in nucleic acid molecules and, if necessary, synthesize phosphodiester bonds and link separate DNA or RNA chains.

The discovery, investigation, and use of thermally stable DNA polymerases produced a revolution in molecular biology, genetic engineering, genomics, medical analysis, and medical genetics. The ability of this class of enzymes to synthesize a complementary DNA chain enabled researchers to "breed" separate fragments of DNA in vitro, to separate out fragments of this information molecule, and to isolate protein genes.

Enzymes are the basis of many present-day technologies. The large-scale production of edible carbohydrates, the synthesis of numerous antibiotics, the fine organic synthesis of numerous drugs, and the production of polymers and biofuels now seem to be impossible without employing enzymes. The area of application of enzymes broadens steadily as new information concerning their properties and their synthesis in sufficient amounts appears. Note the marked progress recently achieved in the synthesis of new protein molecules with desired properties. The variety of new proteins and their potential have far exceeded the limits set by Nature.

Our knowledge of the molecular nature of enzyme catalysis is now becoming systematic. Knowledge of the primary structure of a protein (the sequence of amino acids in its polypeptide chain) makes it possible to reconstruct its three-dimensional structure with atomic-scale resolution. In turn, this gives the molecular active site pattern of the enzyme. If the kinetics of the reaction of interest is known, this encourages the researcher to carry out quantum chemical calculations for the elementary steps of catalysis. Here, this approach will be illustrated by several examples.

CHEMICAL KINETIC AND STRUCTURAL STUDIES AS THE KEYS TO UNDERSTANDING THE MECHANISMS OF ENZYME CATALYSIS

Enzymologists have made marked headway in understanding the functioning of enzymes. Systematic and diversified studies of enzymes have ensured a systemic approach to the investigation of the natural phenomenon of enzyme catalysis.

The present-day knowledge of the mechanisms of enzyme functioning has been provided by two basic types of studies:

(1) kinetic studies, including identification of the points at which the substrate is introduced into the catalytic cycle, detection and identification of the chemical nature of the labile intermediates, and analysis of various states of the active site;

(2) structural studies of enzymes and their active sites.

Experimentally substantiated views of catalytic mechanisms are based on structural data and on kinetic studies including identification of the intermediates involved in the process. The formal kinetic analysis of enzyme-catalyzed reactions is a matter of great interest. The fundamentals and the most important results of this analysis are presented in a number of monographs, textbooks, and manuals [3–7]. Here, we report a detailed analysis of various kinetic schemes relating the reaction rate to the concentrations of the participants of the catalytic processes (substrates, active sites, intermediates) and describe the development of the catalytic process over time. In most cases, including steady- and unsteady-state processes, significant factors simplifying the analysis are the structural homogeneity of the active sites of the enzyme and the linearity of the rate equations for the elementary steps of the process with respect to the enzyme concentration. This appreciably simplifies the formal kinetic description of the reaction and allows theoretical equations to be correctly compared to experimental data.

The main inference from formal kinetic analysis is that the catalytic cycle of an enzyme is a multistep chemical and physical process involving a large number of labile intermediate compounds, including both chemical intermediates and conformers of catalytic groups, substrates, and intermediates. The formal kinetic examination of various kinetic schemes and the corresponding experimental studies in steady- and unsteady-state regimes have provided detailed information concerning the number and nature of these intermediates and their formation and conversion kinetics [8].

The studies that have extended the area of application of the kinetic approach are exemplified by the investigation of multisubstrate multistep processes, such as the conversion of polyunsaturated fatty acids into prostaglandins, taking into account enzyme inactivation during the reaction [9–11]. Note also the mechanistic studies of the involvement of an extra nucleophile in hydrolase reactions, the process changing the hydrolytic reaction to a synthetase reaction [12–14]. The theoretical and experimental methods of kinetic analysis are steadily developing. The discovery of bioelectrocatalysis [15], which is the acceleration of electrochemical reactions by a redox enzyme under conditions of direct electric contact between a conducting material and an active site of the enzyme, added the theoretical and experimental methods of modern electrochemistry to the arsenal of tools for mechanistic studies of enzymatic action [16, 17].

Structural and kinetic data serve as the basis for deduction of schemes of molecular transformations under the action of the active sites of enzymes [8].

BIOINFORMATICS APPLIED TO THE IDENTIFICATION OF CATALYTIC SITES FROM THE SEQUENCE OF AMINO ACIDS IN THE POLYPEPTIDE CHAIN

In most cases, the active site of an enzyme, which is a set of functional radicals of amino acids, can safely be regarded as consisting of two fundamentally important sites, namely, binding and catalytic. The binding site ensures complexation between the substrate and the enzyme and is responsible for substrate selection and for the specificity of the enzyme. The catalytic site, which usually contains acids, bases, metal ions, and prosthetic groups, is responsible for other essential stages of catalysis, namely, substrate activation and conversion. Due to the extensive investigation of protein structures in the last decade, there is good reason to claim that many of the basic types of catalytic sites existing in nature have already been identified. This confidence is based on the great experience of chemical enzymology, which deals with large arrays of structural data and reveals similar catalytic site structures in many enzymes [8]. Enzymes belonging to different classes may have identical or very similar catalytic site structures.

At present, there is information concerning the primary structure of more than one million proteins, including proteins present in the genomes of humans, animals, plants, and bacteria. The 3D protein structure database includes over 30000 structures, about 5000 of which are base structures. A large fraction of these proteins are enzymes. The enzyme classifier contains some 3730 items, supporting the belief that most catalytic structures are known.

A significant role in the study of enzymes and enzymatic action is played by methods of bioinformatics. Present-day information technologies allow the structures of a large number of proteins to be intercompared, making it possible to find common, functionally significant elements, such as active sites. This approach is based on the phenomenon that the variety of catalytically active proteins, which is nearly infinite in terms of protein structure, uses a quite limited set of catalytic site structures [18].

Using computer technology, we have devised two methods for identifying the functional groups constituting the active site of an enzyme. The structure of the catalytic site of an enzyme can be derived from the primary sequence of amino acids in the protein [19–23]. On the other hand, the catalytic site of a protein can be identified using atomic-resolution 3D structure data if it has not been identified otherwise [24].

Information technologies allow the catalytic groups of an enzyme to be revealed by analysis of the primary sequence of amino acids in the protein.

This approach is based on comparing the primary structures of proteins as carriers of some enzymatic activity and on revealing common or so-called conserved positions containing the same amino acid.

Modern computer databases contain detailed information about protein and enzyme structures, including amino acid sequences in polypeptide chains and atomic-resolution protein structures.

Most data available on the subject refer to the primary structure of proteins (i.e., amino acid sequences). Genome studies have already provided a great amount of relevant information, which continues to accumulate exponentially.

The primary sequence is known to determine the entire hierarchical structure of the protein. The obvious scientific problem is to learn to predict the whole 3D structure of a protein, including the structure of its active site, knowing the amino acid sequence. At least two approaches are used now to solve this problem. The first approach is based on the theoretical calculation of the energetics of a protein globule and the determination of the minimum-energy structure [25]. This approach is computation-intensive. Furthermore, the problem has multiple solutions in most cases. However, progress in this field is expected to yield really significant results.

The second approach is homology modeling of the protein structure, which is much simpler. It is based on the fact that many proteins, particularly proteins executing identical or similar functions in spite of being obtained from different sources, are very similar and highly homologous. Since the complete 3D structures of many proteins are known, they can be used to deduce the structure of a homologue. This would markedly simplify the necessary calculations.

In order to identify the amino acids forming the active site of an enzyme and the structure-forming amino acids that assemble side radicals into a catalytically active site, we used an approach based on comparison of amino acid sequences in a large family of homologous proteins [20–23].

Intercomparing the primary structures of proteins with a high degree of homology (30% and above) is termed sequence alignment. Some protein in a sample is taken to be the base, and the other homologous proteins are computer-aligned with it to judge the degree of similarity between amino acids occupying each particular position throughout the sample. The alignment algorithm automatically reveals insertions (extra, noncommon sequences) and deletions (missing fragments of the chain). The result of alignment is a matrix whose elements are probabilities of a given amino acid appearing in a particular position of the polypeptide chain chosen as the base. This approach makes it possible to determine the positions in which a given amino acid will stay with a high probability and to find conserved positions common to all proteins in the sample. On the other hand, this approach reveals the positions that can accommodate almost any of the 20 natural amino acids. A powerful filter that can separate a signal out of statistical noise is the Shannon entropy

$$H = -\sum P_i^j \log P_i^j,$$

where P_i^j is the probability that the *i*th amino acid is in the *j*th position.

A remarkable property of this function is that, if the probability of the event is equal to unity, the Shannon entropy is zero, allowing the signal to be filtered from the noise.

Alignment data for homologous proteins are accumulated in the HSSP database [26].

We used the alignment procedure for analogous proteins, determined the probability of a given amino acid being in a given position, and calculated the Shannon entropy for each position of the sequence in order to identify the amino acids constituting the active site of an enzyme. This method enabled us to reliably identify the functional groups of the active site and the structure-forming amino acids [20, 21].

An analysis of such relationships for a large number of enzymes revealed interesting regularities.

The amino acid sequences within a protein family executing a certain biological function, for example, specific enzymatic activity, are rather diverse. However, there are nearly conserved amino acid positions; that is, some positions are occupied by a definite amino acid throughout the family (which may consist of over 2000 proteins). These positions are readily identifiable within this approach because the Shannon entropy for an amino acid in a fixed position is close to zero.

A large fraction of amino acids in the polypeptide chains of related proteins is highly variable (the corresponding Shannon entropy is 1–3). Through evolution and mutagenesis, Nature has created a wide variety of kindred structures capable of executing the same function, specifically, carrying out a certain enzymatic reaction.

The fact that, for nearly identical spatial structures of the catalytic site, there can be a wide variety of amino acid sequences in the polypeptide chain suggests that, for the formation of a catalytic site, it is essential that only a limited number of amino acids be in certain positions of the polymer chain. This natural amino acid variability is evidence that most amino acids in a polypeptide chain can be replaced with other ones. In many positions, the Shannon entropy is as high as 2.5–3, indicating that these positions can be occupied by almost any amino acid of the 20 natural ones. This would change some properties of the protein, such as stability, the isoionic point, and hydrophilicity or hydrophobicity. However, the basic properties, which are catalytic activity and catalytic site structure, are invariable throughout the family.

The amino acids forming a catalytic site are always conserved amino acids with nearly zero Shannon entropy. This group of amino acids includes aspartic acid, histidine, arginine, serine, and glutamic acid. All of these amino acids have pronounced acid–base properties and are thus capable of activating a molecule via a nucleophilic–electrophilic mechanism.

A frequency rating was established for amino acids in conserved positions. Glycine was found to be the most frequently conserved amino acid [19].

The second most frequently conserved amino acid is aspartic acid, which appears in active sites either as a nucleophile (in deprotonated form) or as an electrophile (in protonated form). Histidine, which has an imidazole group, is of great significance for hydrolases. Arginine, which has a positively charged guanidine grouping, is essential for oxide reductases.

The distribution of functions among the different types of conserved amino acids is quite clear. The carboxylic group of aspartic or glutamic acids, the imidazole group of histidine, and the guanidine group of arginine form three-dimensional catalytic structures capable of carrying out concerted nucleophilic–electrophilic catalysis. The elementary step of catalysis is associated with comparatively slight, 1- to 1.5-Å-long proton transfer along the hydrogen bond line or with bond polarization by an electrophilic agent. This enhances the reactivity of the substrate molecule or a functional group of the active site by a factor of 10^6–10^7. For example, proton transfer from the water molecule, which makes its transition-state structure similar to the structure of the hydroxide ion, enhances its reactivity in nucleophilic substitution reactions by a factor of 10^7. If the process is concerted and each substrate is activated in this way, the overall reaction would be expected to be sped up by a factor of 10^{12}–10^{15}.

The catalytically significant amino acids are distributed throughout the polypeptide chain of the protein. The folding and the formation of the 3D structure of a protein bring the catalytic functional groups closer together to form the 3D structure of the catalytic site. The factor ensuring the folding of the polypeptide chain is that the structure-forming amino acids—glycine, proline, and cysteine—are in appropriate places. Catalytically significant amino acids may appear in different orders in the polypeptide chain.

The main structural paradox of enzyme catalysis is that the colossal variety of enzymatic reactions, which are catalyzed by a nearly infinite number of proteins with various primary structures, involve only a limited set of catalytic sites.

For example, the hydrolases, which are the most numerous class of enzymes, account for about one-third of the enzymes known today (among the 3700 enzymes included in the classifier, 1100 are catalases). There is biological multiplicity behind each of these enzymes since, in biological systems, these proteins are represented by nearly infinite varieties of amino acid sequences. However, there are only five basic types of catalytic sites capable of activating substrate molecules and carrying out a catalytic cycle [22]. The variety of possible reactions (~1100 for the hydrolases) arises from the action of the binding site, which

is responsible for substrate complexation and for the orientation of the active site relative to the catalytic groups of the protein. Thus, the catalytic sites of enzymes are unique structures afforded by evolution, likely via extensive search through intermediate structures. The catalytic site structures are fixed genetically as conserved positions of the catalytically significant amino acids as well as the structure-forming amino acids, which assemble the functional groups of the former in an appropriate point in space, ensuring the necessary steric fitting.

The phenomenon of the structural uniformity of catalytic sites manifests itself when the active sites of nonhydrolase enzymes are compared to those of hydrolases.

An analysis demonstrated that the active sites of nonhydrolase enzymes are composed of the same nucleophilic and electrophilic elements as the active sites of hydrolases. Nonhydrolases activate the reaction sites of molecules via the same nucleophilic–electrophilic mechanisms, often involving the same 3D structures [8].

The structural uniformity of the catalytic sites on enzymes is still more evident when the active sites of hydrolases are compared to those of synthetases, which act to form chemical bonds and do not use a water molecule as a reactant. Passing from a hydrolytic reaction to a synthetic reaction is done by replacing water activation with, e.g., hydroxyl group activation in a carbohydrate molecule without changing the structure of the activating site [8, 22].

The above identification of catalytic and structure-forming amino acids, which is based on the primary amino acid sequence in the polypeptide chain, and the structural uniformity of enzymes from different classes have many implications.

Obviously, the catalytically active proteins can be constructed from many fewer amino acids than the twenty that occur in nature. The amino acids in most positions in the polypeptide chain can be replaced with other ones without significantly changing the catalytic function of the protein. It is essential that certain positions be occupied by crucial amino acids that either possess acid–base properties and thus form a catalytic site or are structure-forming and serve to assemble the catalytic groups in a certain place in space.

The conventional methods of modifying the properties of enzymes are site-directed mutagenesis and directed evolution [8, 27]. An analysis demonstrated that unconstrained variation of amino acids is allowable only in nonconserved positions. Otherwise, the modified protein will very likely be enzymatically inactive.

Obviously, the type of reaction catalyzed by an enzyme is determined, to a considerable extent, by the structure of the binding site, which provides a position for attack by catalytic groups. This opens up prospects for design of protein catalysts for reactions that do not take place in the biological world. Such catalysts would markedly raise the potential of enzymes for solving chemical problems. The current state of the art in chemical enzymology allows researchers to pose the problem of tailoring enzymes to desired chemical reactions, looking beyond the opportunities offered by Nature.

ATOMIC-RESOLUTION 3D STRUCTURES: ACTIVE SITES OF ENZYMES : GEOMETRIC INVARIANTS AND OBSERVED PARAMETERS

Structural characterization of enzymes has provided the researchers in catalysis with a large set of atomic-resolution structures. The protein data bank PDB, which currently contains over 30000 structures, grows exponentially at a rate of 20% annually [28]. Homology modeling using combinatorial methods is becoming an independent method of generating 3D structures of proteins and enzymes. This allows one to pose and solve general problems in order to gain insight into catalysis as a natural phenomenon. We have solved the problem of devising an automated procedure for identifying active sites as a certain atomic configuration. An associated problem is to estimate the permissible variations in interatomic distances and bond angles that will not destroy the catalytic function.

In other words, we wish to answer the following question: How rigorous are the structural requirements that an atomic configuration must meet to be catalytically active?

An answer can be obtained by statistically comparing the structures of enzymes containing the same active site but differing in origin, specificity, or the structure of the protein moiety.

In our studies, the model objects were serine hydrolases, whose catalytic site is the serine–histidine–aspartic acid triad. The enzymes of this class are best studied in terms of catalysis mechanisms, specificity, and active site structure [5, 29].

In the data bank PDB, this class of enzymes is represented by 1284 proteins, whose structures have been determined by X-ray crystallography (for 1256 proteins) or ^{1}H NMR spectroscopy (for 28 proteins).

A preliminary analysis demonstrated that the enzymes whose catalytic activity is due to the Ser–His–Asp triad may have diverse atomic configurations forming the catalytic site. Figure 1 illustrates the 3D alignment of the active sites of six different enzymes. In this computer-assisted procedure, the position of the imidazole group of histidine was fixed for all enzymes in order to follow the variation of the configurations of the two other participants of the catalytic cycle, namely, the hydroxyl group of serine and the carboxyl group of aspartic acid. Clearly, the arrangement of active site atoms can vary rather widely.

We performed a statistical analysis of permissible variations in working ferments. This was done in order to determine the limits of these variations and to devise an automated computer-assisted procedure for identifi-

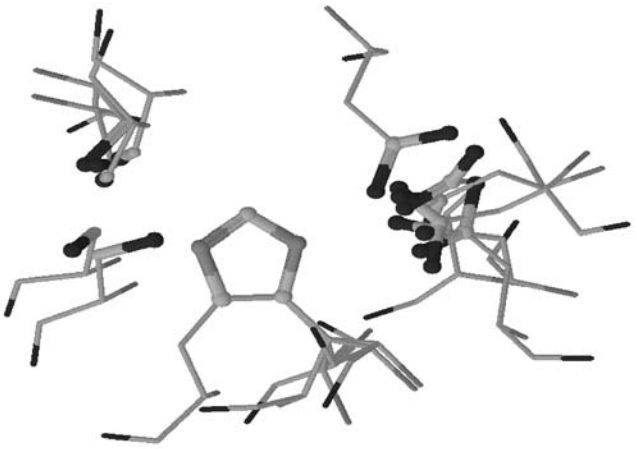

Fig. 1. Spatial alignment of six hydrolases containing the Ser–His–Asp triad: acetylcholinesterase (1FAS), trypsin (1FIZ), kexin (1OTS), acetylxylan hydrolase (1QOZ), and ATP-dependent endopeptidase (1TYF).

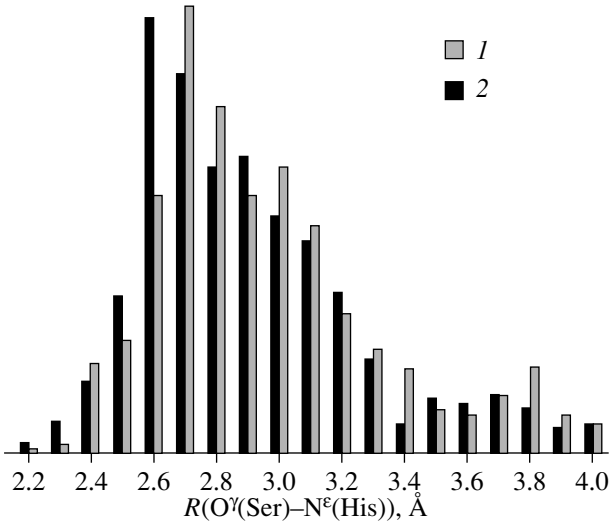

Fig. 2. Frequency diagram for the distance between the hydroxyl oxygen atom O^{γ} of serine and the imidazole nitrogen atom N^{ε} of histidine in enzymes: (*1*) free enzyme and (*2*) enzyme–inhibitor or enzyme–substrate complex.

cation of this type of catalytic site from 3D structure data for proteins. This implies generating a virtual template with allowable values of parameters used to recognize catalytic sites in the sophisticated atomic configuration of a protein molecule.

In the generation of the virtual template, we discarded the simple procedure of comparing structures in terms of rms deviations of atomic coordinates for the reason that this procedure ignores many specific structural differences associated with the spatial atomic arrangement.

We developed a geometric invariant–based procedure for revealing local atomic identities in protein structures. In this procedure, the geometric invariants are the distances between fundamentally important atoms, the plane angles defined by three atoms belonging to two or three catalytic groups, and the plane angles between two vectors each defined by one atom of one catalytic amino acid residue [24].

The "teaching" sample data set consisted of data for enzymes with a reliably determined structure. The constraints were the minimum and maximum values of the given parameters.

Table 1. Most significant constraints imposed on the structure of the Ser–His–Asp catalytic triad in the active sites of serine hydrolases

Parameter	Minimum	Maximum
$His N^{\varepsilon}$–Ser C^{β} distance, Å	2.9	4.1
$His N^{\beta}$–Asp O^{β} distance, Å	2.5	3.8
$His N^{\varepsilon}$–Ser O^{γ} distance, Å	2.4	4.2
$His C^{\beta}$–Asp $O^{\sigma 1}$–$His C^{\varepsilon}$ angle, deg	47.8	61.6
$His N^{\varepsilon}$–His C^{β}–$Ser O^{\gamma}$ angle, deg	0.9	3.6

Table 1 presents the template consisting of the five most significant parameters of the catalytic sites involving a Ser–His–Asp triad. Our automatic computer-assisted procedure allows one to reliably identify the catalytic sites of serine hydrolases. We examined all the 3D structures included in PDB and found Ser–His–Asp active sites in proteins not referred to as serine hydrolases [24].

Of great interest is our analysis of the distribution of various geometric invariants based on the entire structural data array presented in PDB. This analysis revealed several very significant facts that throw light on the nature of catalytic action. In Fig. 2, we show the frequency diagram for the distance between the hydroxyl oxygen atom of serine and the N^{ε} atom of the imidazole group of histidine. These two atoms are hydrogen-bonded, and the activation of the hydroxyl group owing to the nucleophilic effect of the imidazole group as a base is the key process of the catalytic cycle.

We examined 1180 structures, including 291 free enzymes, 515 enzyme–inhibitor (enzyme–substrate) complexes, and 374 acyl-enzymes, and obtained the following results:

(1) The distribution of $O^{\gamma}\cdots H\cdots N^{\varepsilon}$ distances is not Gaussian. It is more complicated, definitely indicating that there are two groups of active sites. For the free enzyme, the frequency distribution function peaks at 2.7 and 3.8 Å. This dual behavior of the active site was deduced for the free enzyme as well as for the enzyme–inhibitor (enzyme–substrate) complex and the acyl-enzyme.

(2) A comparison between the $O^{\gamma}\cdots H\cdots N^{\varepsilon}$ distance distributions for the free enzyme and enzyme– inhibitor (enzyme–substrate) complex indicates that the ligand somewhat squeezes the active site, bringing the active

Table 2. Root-mean-square deviations (RMSD) of atomic coordinates in the amino acid residues involved in the catalytic site for the replacement of the conserved and nonconserved Gly residues with Ala in various hydrolases

Hydrolase family	Number of conserved Gly residues	Number of nonconserved Gly residues	Catalytic site	RMSD for the replacement of the conserved residues (E_1)*	RMSD for the replacement of the nonconserved residues (E_2)*	E_1/E_2
α-Chymotrypsin	7	15	His57, Asp102, Ser195	0.40 (0.057)	0.10 (0.07)	8.1
Human pepsin	8	26	Asp 32, 215	0.62 (0.078)	0.49 (0.019)	4.1
E. coli inorganic pyrophosphatase	6	10	Glu20, Asp65, Asp70, Asp102	0.62 (0.1)	0.55 (0.055)	1.8
Alkaline phosphatase	6	39	2Zn, Ser102	A : 0.79 (0.13)	A : 1.17 (0.03)	4.3
				B : 0.36 (0.06)	B : 0.14 (0.003)	20.0
Carlsberg subtilisin	5	29	Asp32, His64, Ser221	0.55	0.26	2.1

* The numbers in parentheses are RMSD values per replaced Gly residue (RMSD/n, where n is the number of replaced Gly residues).

atoms closer together (the average distance changes from 2.7–2.8 to 2.6 Å).

(3) A comparison between the $O^{\gamma}\cdots H\cdots N^{\varepsilon}$ distance distributions for the free enzyme and the acyl-enzyme definitely indicates that, on passing to the acyl-enzyme, the active site loosens: the average distance lengthens from 2.6 to 2.9 Å. This is likely due to the steric effect of the acyl group and to the necessity to redirect the activating action of the imidazole group to water at the acyl-enzyme hydrolysis stage.

Thus, the above analysis provided fundamentally important information concerning the structure of the active sites of enzymes, including the permissible ranges of geometric invariants. Furthermore, this analysis enabled us to devise an automatic procedure for identifying active sites in protein molecules and to reveal significant features of the reorganization of active sites in catalytic cycles.

MOLECULAR MECHANICS OF PROTEINS: COMPUTER-SIMULATED MUTATIONS OF CONSERVED AND NONCONSERVED GLYCINE RESIDUES

Recent years have brought new opportunities in the investigation of the molecular mechanisms of enzyme catalysis. These opportunities are offered by methods of molecular mechanics and quantum chemistry applied to the calculation of the behavior of protein molecules during a catalysis event. Creation of efficient protein description programs in the framework of classical mechanics gave researchers an opportunity to calculate the energetics of protein molecules and to deduce the behavior of protein chains under given conditions [30–32].

It is of fundamental interest to calculate the interaction between a protein and a nonmacromolecular ligand (substrate or inhibitor), to study so-called molecular docking, and to simulate mutagenesis (that is, to calculate how a protein molecule will change upon the replacement of an amino acid with another one). Computer-simulated mutagenesis is of fundamental importance in the rational design of proteins through goal-directed changes in their properties.

The theoretical basis for biomacromolecule structure and interaction calculations is molecular mechanics. In this approach, the potential energy of the system is considered and the global minimum of the potential energy function is assumed to correspond to the equilibrium structure of the system. Various dynamic programming techniques are used to find this minimum (the potential energy of the system depends on the Cartesian coordinates of all its atoms and is generally a function of several thousand variables). The contribution from a particular atom to this function depends on the valence state and charge of this atom. The rules according to which each atom of a biomolecule is assigned some parameters characterizing its contribution to the potential of the system and the values of these parameters make up the so-called force field. At present, the standard force fields are CHARMM22 [30–32] and AMBER [33].

In order to understand the role of the conserved glycine residues in enzyme catalysis, we modeled protein molecules, replacing conserved and nonconserved glycine residues (computer-simulated mutagenesis). In a number of protein molecules, all conserved or all nonconserved glycine residues were replaced with alanine, the amino acid structurally least different from glycine. We analyzed the structural changes in the catalytically active groups (acids and bases revealed by multiple alignment) caused by the substitution of Ala for Gly [34].

Furthermore, we calculated the structural changes in enzymes caused by substituting Ala for all conserved Gly residues. In each hydrolase family to be examined, we chose a representative with a known 3D structure (Table 2). For this representative, we determined the conserved Gly numbers, using the HSSP database, and the number and positions of conserved and noncon-

served Gly residues in the enzyme chain. Three 3D structures of the enzyme, namely, the native structure, the structure with mutated conserved Gly residues, and the structure with mutated nonconserved Gly residues, were subjected to a structure relaxation procedure. This procedure consisted of 1000 steps minimizing the potential energy of the structure. This "careful" structure relaxation is quite adequate for revealing the structural rearrangements caused by Gly mutations since the steric strain induced by the substitution of Ala for Gly is weak and can be eliminated by minimization alone. With this procedure, it is possible to conserve the structural parameters of the catalytic site and, therefore, quantitatively characterize the mutation-induced changes in the 3D structure. The structures resulting from the minimization procedure were compared via the best superimposition of the atoms of the amino acid residues making up the catalytic site (Table 2). The atomic coordinate RMSD values per atom of the active site for the superimposition of the native (energy-minimized) structure on the relaxed structures in which the conserved Gly residues are replaced with Ala are listed in Table 2. The mutation of Gly, energy minimization, and the analysis of the resulting structures were carried out using the program package Insight II.

The calculated values of atomic RMSD between the mutated and native structures are a measure of the disturbance of the finely tuned active site structure caused by the replacement of the amino acids (under the assumption that the energy-minimized native structure retains the ideal mutual orientation and arrangement of active site groups). It follows from the data presented in Table 2 that the replacement of all conserved Gly residues with Ala causes much more radical changes in the catalytic site geometry than the replacement of all nonconserved Gly residues. This is particularly evident from the estimated RMSD values per replaced Gly residue (see the rightmost column of Table 2). Thus, the nonconserved Gly residues are localized in thin protein molecules that, owing to the conformational mobility of Gly, allow the atoms constituting the catalytic site to assemble in space most precisely.

QUANTUM CHEMICAL CALCULATIONS FOR ELEMENTARY EVENTS OF A CATALYTIC CYCLE

Application of methods of quantum mechanics and quantum chemistry to separate events of catalytic cycles has afforded a qualitative breakthrough in the understanding of elementary steps of enzymatic reactions. Sufficiently reliable quantum chemical calculations were made feasible by deliberately choosing catalytic sites whose behavior can really be described using quantum chemical methods. The rest of the protein molecule can be described using methods of molecular mechanics (QM–MM approximation [35, 36]). This approach makes it possible to calculate multidimensional potential energy surfaces, to identify and examine stationary points (global and local minima and transition states), and to calculate the energy profiles of chemical reactions. This has brought us to a higher level in the understanding of the molecular mechanisms of reactions.

Quantum chemical calculations have provided a key to many problems of the detailed investigation of the mechanisms of enzymatic reactions. For example, since the determination of the 3D structure of α-chymotrypsin by Blow and his colleagues, it was universally accepted until recently that the hydroxyl group of serine in the Ser–His–Asp triad is activated via the following double proton transfer process: serine hydroxyl $\longrightarrow$ histidine imidazole group $\longrightarrow$ the deprotonated carboxyl group of aspartic acid. This mechanism, which is based on the seemingly indubitable fact that the proton is transferred from the positively charged imidazole ion to the negatively charged carboxyl group of aspartic acid, was accepted to explain the action of the above triad in the serine hydrolases.

However, quantum chemical calculations of energy profiles in the QM–MM approximation taking into account the surrounding of the carboxyl group demonstrated that the serine hydroxyl group is activated via a one-proton process. The role of the carboxyl group of aspartic acid is to orient the imidazole group in the space of the active site [37].

At present, there are well-substantiated ideas of the nature of enzymatic specificity, which means the ability of an enzyme to recognize the structure of an efficient substrate or inhibitor [18]. The base physicochemical model of this phenomenon assumes that the interaction between a substrate or inhibitor and the binding area of the active site decreases the Gibbs energy of activation of the rate-limiting step of the process.

It was substantiated experimentally that the reduction of the Gibbs energy of activation is due to the reacting bond being strained. The view that the substrate-selecting properties of enzymes are associated with the conformational changes in the active site induced by the interaction between a "good" (specific) substrate and active site groups (so-called induced-fit mechanism) has also received theoretical and experimental substantiation [8].

The induced-fit mechanism, suggested by Koshland [38], has been substantiated by quantum chemical calculations [39]. As was demonstrated above for serine hydrolases (see Fig. 2), the O^{γ}–N^{ε} distance in a free enzyme is always 0.1–0.2 Å longer than the same distance in an enzyme–substrate (enzyme–inhibitor) complex. Therefore, complexation between the substrate and the active site causes some changes in the latter that advance the system along the reaction coordinate. According to quantum chemical calculations, the energy required for O^{γ}–N^{ε} proton transfer is ~3 kcal higher for the enzyme–substrate complex than for the free enzyme. Since the proton transfer energy is a component of the activation barrier (the overall activation

barrier is 9.6 kcal/mol), this result is evidence that the substrate induces structural changes in the active site that then exert a significant effect on the enzymatic reaction, raising its rate constant by a factor of ~150.

CONCLUSIONS

Note the following specific features of the state of the art in the physical and structural chemistry of enzymes:

The systemic approach to the study and understanding of enzyme catalysis is now completely formed. Knowledge of the primary structure, which is the sequence of amino acids in the polypeptide chain, provides sufficient information for higher hierarchical levels of the understanding of enzyme catalysis, including elementary events of the catalytic cycle. Advanced methods of bioinformatics are capable of identifying the functional groups constituting a catalytic site and the fundamentally important structure-forming amino acids. For a protein having homologues with known atomic-resolution structures, it is possible to obtain fairly reliable 3D structure data. This makes it possible to determine the geometry of the active site. Next, methods of molecular mechanics and quantum chemistry are applied to calculate the interaction between the catalytic groups and the substrate. This provides information concerning elementary event energetics for various possible mechanisms and participants of the catalytic process. Eventually, this approach provides insight into the molecular mechanism of catalysis and into the physical nature of reaction speedup under the action of an enzyme.

ACKNOWLEDGMENTS

The author is grateful to Prof. A.V. Nemukhin, Cand. Sci. (Phys.–Math.) I.V. Uporov, and Cand. Sci. (Chem.) I.V. Gariev for their participation in the studies used as the basis for this report.

REFERENCES

1. Solov'ev, Yu.I. and Kurennoi, V.I., *Yakob Bertselius: Zhizn' i deyatel'nost'* (Jöns Jacob Berzelius: Life and Work), Moscow: Nauka, 1980.
2. Shamin, A.N., *Biokataliz i biokatalizatory: Istoricheskii ocherk* (Biocatalysis and Biocatalysts: A Historical Outline), Moscow: Nauka, 1971.
3. Webb, J.L., *Enzyme and Metabolic Inhibitors*, New York: Academic, 1963.
4. Dixon, M. and Webb, E., *Enzymes*, London: Longman, 1979.
5. Berezin, I.V. and Martinek, K., *Osnovy fizicheskoi khimii fermentativnogo kataliza* (Fundamentals of the Physical Chemistry of Enzyme Catalysis), Moscow: Vysshaya Shkola, 1977.
6. Varfolomeev, S.D. and Zaitsev, S.I., *Kineticheskie metody v biokhimicheskikh issledovaniyakh* (Kinetic Methods in Biochemical Research), Moscow: Mosk. Gos. Univ., 1982.
7. Varfolomeev, S.D. and Gurevich, K.G., *Biokinetika* (Biokinetics), Moscow: Fair-Press, 1999.
8. Varfolomeev, S.D., *Khimicheskaya enzimologiya* (Chemical Enzymology), Moscow: Akademiya, 2005.
9. Varfolomeev, S.D. and Mevkh, A.T., *Prostaglandiny–molekulyarnye bioregulyatory* (Prostaglandins As Molecular Regulators), Moscow: Mosk. Gos. Univ., 1985.
10. Varfolomeev, S.D. and Mevkh, A.T., *Biotechnol. Appl. Biochem.*, 1993, p. 291.
11. Vrzheshch, P.V., Batanova, E.A., Mevkh, A.T., Varfolomeev, S.D., Gazaryan, I.G., and Thorneley, R.N., *Biochem. J.*, 2003, vol. 372, pp. 713–734.
12. Gololobov, M.Y., Borisov, I.L., Belikov, V.M., and Svedas, V.K., *Biotechnol. Bioeng.*, 1988, vol. 32, no. 7, p. 866.
13. Gololobov, M.Y., Borisov, I.L., and Svedas, V.K.J., *Theor. Biol.*, 1989, vol. 140, p. 193.
14. Youshko, M.I., Chilov, G.G., Shcherbakova, T.A., and Svedas, V.K., *Biochim. Biophys. Acta*, 2002, vol. 1599, nos. 1–2, p. 134.
15. Varfolomeev, S.D., Berezin, I.V., Bogdanovskaya, V.A., Tarasevich, M.R., and Yaropolov, A.I., *USSR State Register of Disciveries*, Discovery no. 306, 1984.
16. Berezin, I.V., Bogdanovskaya, V.A., Varfolomeev, S.D., Tarasevich, M.R., and Yaropolov, A.I., *Dokl. Akad. Nauk SSSR*, 1978, vol. 240, p. 615.
17. Varfolomeev, S.D., *Konversiya energii biokataliticheskimi sistemami* (Energy Conversion in Biocatalytic Systems), Moscow: Mosk. Gos. Univ., 1981.
18. Varfolomeev, S.D. and Pozheikov, A.E., *Vestn. Mosk. Univ., Ser. 2: Khim.*, 2000, vol. 41, p. 147.
19. Varfolomeev, S.D., Gurevich, K.G., Poroikov, V.V., Sobolev, B.N., and Fomenko, A.E., *Dokl. Akad. Nauk*, 2001, vol. 379, p. 548 [*Dokl. Biochem. Biophys.* (Engl. Transl.), vol. 379, p. 252].
20. Varfolomeev, S.D. and Gurevich, K.G., *Izv. Akad. Nauk, Ser. Khim.*, 2001, no. 10, p. 1629.
21. Varfolomev, S.D., *Mendeleev Commun.*, 2004, vol. 5, p. 185.
22. Varfolomeev, S.D., Uporov, I.V., and Gariev, I.V., *Usp. Khim.*, 2005, vol. 74, p. 67.
23. Varfolomeev, S.D., *Khimicheskaya i biologicheskaya kinetika. Novye gorizonty* (Chemical Kinetics and Biokinetics: New Horizons), Moscow: Khimiya, 2005.
24. Gariev, I.V. and Varfolomeev, S.D., *Bioinformatics*, 2006, vol. 22, no. 20, p. 2574.
25. Finkel'shtein, A., *Fizika belka* (Protein Physics), Moscow: Nauka, 2002.
26. ftp://ftp.embl-ebi.ac.uk/pub/database.
27. Knorre, D.G. and Myzina, S.D., *Biologicheskaya khimiya* (Biochemistry), Moscow: Vysshaya Shkola, 1992.
28. http://www.pdb.org/.
29. Antonov, V.K., *Khimiya proteoliza* (Proteolysis Chemistry), Moscow: Nauka, 1991.
30. McKerell, A.D., Wiyrkiewicz-Kuczeru, J., and Karplus, M., *J. Am. Chem. Soc.*, 1995, vol. 117, p. 11946.

31. McKerell, A.D., Bashford, D., Bellot, M., Dunbrack, K.L., Evanseck, J.D., Field, M.J., Fisher, S., Gao, J., Guo, H., Ha, S., Josph-MacCarthy, D., Kuchnir, L., Kuczera, K., Lan, F.T.K., Smith, J.K., Store, R., Straub, J., Watanobe, M., Wiyrkiewicz-Kuezera, J., Yin, D., and Karplus, M., *J. Phys. Chem.*, 1998, vol. 102, p. 3586.

32. Schelenkrich, M., Brickmann, J., McKerell, A.D., and Karplus, M., in *Biological Membranes: A Molecular Perspective from Computation and Experiment*, Merz, K., Jr. and Roux, D., Eds., New York: Springer, 1966, p. 31.

33. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.S., Fox, T., Coldwell, J.W., and Kollman, P.A., *J. Am. Chem. Soc.*, 1995, vol. 117, p. 5179.

34. Varfolomeev, S.D., Uporov, I.V., and Fedorov, E.V., *Biokhimiya*, 2002, vol. 67, p. 1328 [*Biochemistry* (Engl. Transl.), vol. 67, p. 1099].

35. Warshel, A. and Levitt, M., *J. Mol. Biol.*, 1976, vol. 103, p. 227.

36. Nemukhin, A.V., Grigorenko, B.L., Topol, I.A., and Burt, S.K., *J. Comput. Chem.*, 2003, vol. 24, p. 1410.

37. Nemukhin, A.V., Grigorenko, B.L., Rogov, A.V., Topol, I.A., and Burt, S.K., *Theor. Chem. Acc.*, 2004, vol. 111, p. 36.

38. Koshland, D.E., *Proc. Natl. Acad. Sci. U. S. A.*, 1958, vol. 44, p. 98.

39. Nemukhin, A.V., Gariev, I.V., Rogov, A.V., and Varfolomeev, S.D., *Mendeleev Commun.* 2006, vol. 16, p. 290.