

## Postgenomic chemistry: New problems and challenges\*

Sergey D. Varfolomeyev<sup>‡</sup>, Teimur K. Aliev, and Elena N. Efremenko

*The Lomonosov Moscow State University, Faculty of Chemistry, Chemical Enzymology Department, Lenin's Hills, 1/11, Moscow 119992, Russia*

*Abstract:* New trends in chemistry induced by genetic engineering and genomic information are analyzed. Genomic information and bioinformatics make the identification of all molecular participants of biological processes possible. The methods of identification of enzyme-active sites from sequence data were demonstrated. The problems of “chemical proteomics” were formulated and analyzed. New biocatalytic processes based on a new generation of enzymes, including CO<sub>2</sub> reduction and enzyme fuel cell construction, are demonstrated. It was shown that in many reactions the enzymes are substantially more efficient in comparison with classical chemical catalysts. The creation of recombinant proteins from unnatural amino acids is demonstrated, and organisms containing such proteins are described. The new trends of proteomics and bioanalytical chemistry in the postgenomic era are discussed.

### INTRODUCTION

The past decade has been characterized by the dynamic integration of chemistry and biology. Chemistry as the science investigating structures and properties of molecules provides the fundamental basis for the development of biochemistry, molecular biology, and biotechnology. Chemical science actively investigates and employs the principles of molecular organization of biological processes in living systems. Chemistry and chemical technology peculiarly use biological materials and catalysts to create steadily increased variety of useful molecules. Genomic research also has a great impact on related fields of modern science, including chemistry. The creation of a holistic system of knowledge of genetic nature, genetic materials, and structures of genomes has influence on the development of chemistry, providing it with evolutionary new representatives, methods, and molecular structures. In addition, the challenge of identification of chemical structure and function of all proteins in living organisms links chemical and biological science. The complete elucidation of the human genome and genomes of many other organisms is a remarkable scientific breakthrough that will affect our lives for decades to come. The content of structural information generated as a result of genomics studies is huge. More than 500 genomes are already partially or completely sequenced. Genomes of various microorganisms, plants, animals, as well as the human genome, are among them. These studies allowed identification of 1.1 million nucleotide sequences that code for proteins.

However, there are certain limitations that need to be overcome before one can exploit successfully the wealth of information from genomic studies. Although not at the forefront during the past genomics efforts, chemistry will undeniably continue to play a crucial role in the biological sciences in the future. Therefore, one can predict a bright future for chemistry in the postgenomic era. For example, developments in physical, surface, and analytical chemistry will surely provide the basis of novel,

---

\*Plenary lecture presented at the XVII Mendeleev Congress on General and Applied Chemistry, Kazan, Tatarstan, Russia, 21–26 September 2003. Other presentations are published in this issue, pp. 1605–1798.

<sup>‡</sup>Corresponding author: E-mail: sdvarf@enz.chem.msu.ru

improved detection methods that will speed up and facilitate diagnostics. Organic chemistry will influence biological sciences even more. Genomics provides chemistry with new possibilities, tasks, and challenges.

The functioning of a living organism is far more complex than the genome may suggest. Therefore, profound studies of proteome, metabolome, cellome, and physiome are under way. Another important feature is the multidisciplinary nature of problems, which can be solved only by the combined and concerted efforts of scientific teams, including experts from different disciplines (chemists, biologists, physicians, mathematicians, and others).

There is a need to discuss the directions of future developments in various fields of chemistry in view of and with respect to the developments in genomics. Obviously, the achievements of genomics and proteomics could considerably influence the classical chemistry as well as result in new scientific directions (IUPAC Project # 2001-005-1-300 "Postgenomic chemistry").

## **IDENTIFICATION OF ALL MOLECULAR PARTICIPANTS OF BIOLOGICAL PROCESSES**

The fundamental and philosophical result of the genomic project is that all molecular representatives of the world of biological diversity can be identified.

### **Trace element-containing protein**

Selenoproteins (SPs) are responsible for most biomedical effects of dietary applicable selenium and are important for mammals because selenium plays a significant role in many physiological processes, namely, in male reproduction, aging, immune function, and cancer prevention [1]. The information about the human SPs potentially can be used for the further systematic clarification of mammalian SPs and their functions. It was established that SPs contain selenium in the form of selenocysteine (Sec), which was named as the 21<sup>st</sup> amino acid.

There are three terminating codones in the universal genetic code. It was discovered that one of them (UGA) has dual function, signaling termination of protein synthesis as well as incorporation of such amino acid as Sec [2]. Since then, classical methods of biochemistry and a variety of biophysical techniques are not enough to analyze the correct function of UGA. Most known selenoprotein genes have homologues, in which Sec is replaced with cysteine (Cys). Using that, specific approach was suggested for identification of SPs [3]. It implies the search for the Sec/Cys-containing protein pairs of homologs. In addition, to identify mammalian SPs, human and mouse (or human and rat) genomes were searched in parallel to identify pairs of Sec insertion sequence (SECIS) elements present in the selenoprotein genes. This method supported by the application of specially developed computer programs [4] provided independent verification of a number of SPs in various genomes. Subsequent investigations of new SPs revealed examples of proteins with expression patterns limited to embryos or testes, as well as proteins with novel subcellular distributions. It was established that all characterized mammalian SPs are globular proteins. Sec is inserted into polypeptide chains during ribosome-based protein synthesis. The expression of fusion SPs, containing a C-terminal green fluorescent protein (GFP) tag, and their following detection with antibodies to GFP by means of electron microscopy helped in revealing the first known plasma membrane SPs [3]. It was found that the human genome has 25 selenoprotein genes and the mouse genome 24 ones. The further understanding of mechanisms of Sec insertion can assist in revealing approaches to the targeted insertion of this residue into proteins. These examples emphasize the power of bioinformatics in establishing of deep interrelation between genomics and chemistry.

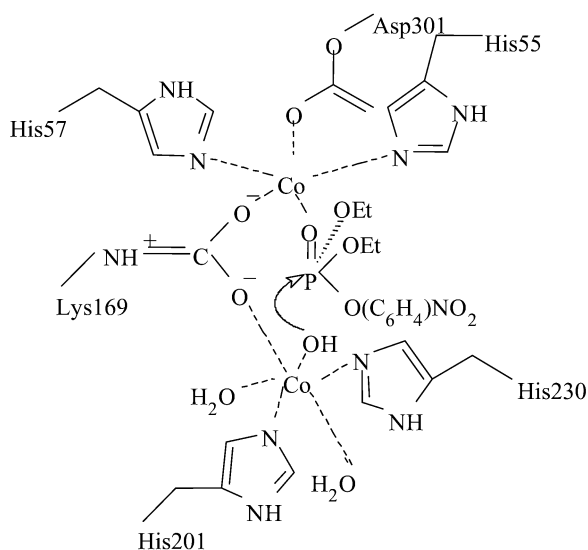
Identification and characterization of new metal-containing proteins in humans and other organisms should set up new challenges and highlight the need for investigation of genomic and postgenomic chemistry of biological trace elements. Understanding of the identities and functions of trace element-containing proteins can also provide new tools for nonspecific incorporation of these elements into proteins.

### Phosphotriesterases in all known genomes

In recent years, of notable interest to us has been the study of triesterases, the enzymes catalyzing hydrolysis of triesters of phosphoric acid [5–11]. The enzymes of this class can accelerate the hydrolysis of a wide group of phosphoric acid derivatives including such supertoxic compounds as Sarin, Soman, Vx, many organophosphorus pesticides, and toxins [5]. The enzymes of this class are applied as detectors of toxic organophosphates in biosensors [8,9] and considered as major components in new protective systems against weapons of mass destruction.

We offered the mechanism of catalysis [7] including the steps of the electrophilic activation of substrate reaction site by a complex with one metallic ion (CoI) in enzyme-active site and of the electrophilic activation of water molecule by a complex with a second metallic ion (CoII) (Fig. 1). The application of modern computer methods for comparison of sequence of nucleic acids coding protein synthesis allows the researchers to answer the question of existence of homologous proteins possessing phosphoesterase activity [12]. Twenty-five proteins detected were highly homologous with the best-studied enzyme isolated from *Flafobacterium* sp. cells (Fig. 2). The figure shows that the proteins of this family are present in cells of diverse biological species, including prokaryotes, eukaryotes, and the human genome. The procedure for construction of the phylogenetic tree includes the combination sequences in clusters with the follow-up comparison of proteins by the principle of minimal deviations. Structurally, the proteins unite in “branches”; the more similar the primary protein structures are, the closer they are located to one another on the phylogenetic tree.

The necessary step in the active site formation and the catalytic activity manifestation is the presence of Lys169 in protein, the carbamylated form of which “collects” cobalt ions into a common active site (Fig. 1) and provides a necessary spatial closeness between substrate and water activation sites. It follows from Fig. 2 that only a limited number of proteins have the amino acid lysine in the proper positions. These proteins are marked with dots in Fig. 2. It shows that, according to the presented data, human proteins and *E. coli* cells do not possess the organophosphate-related enzymatic activity. Proteins of some microorganisms, such as *Mycobacterium*, *Sulfolobus*, *Mycoplasma*, and *Lysteria*, show the potential enzymatic activity. This opens notable possibilities in search of new enzymes of this class.



**Fig. 1** Scheme of organophosphate hydrolase action with paraoxon as substrate in the active site.

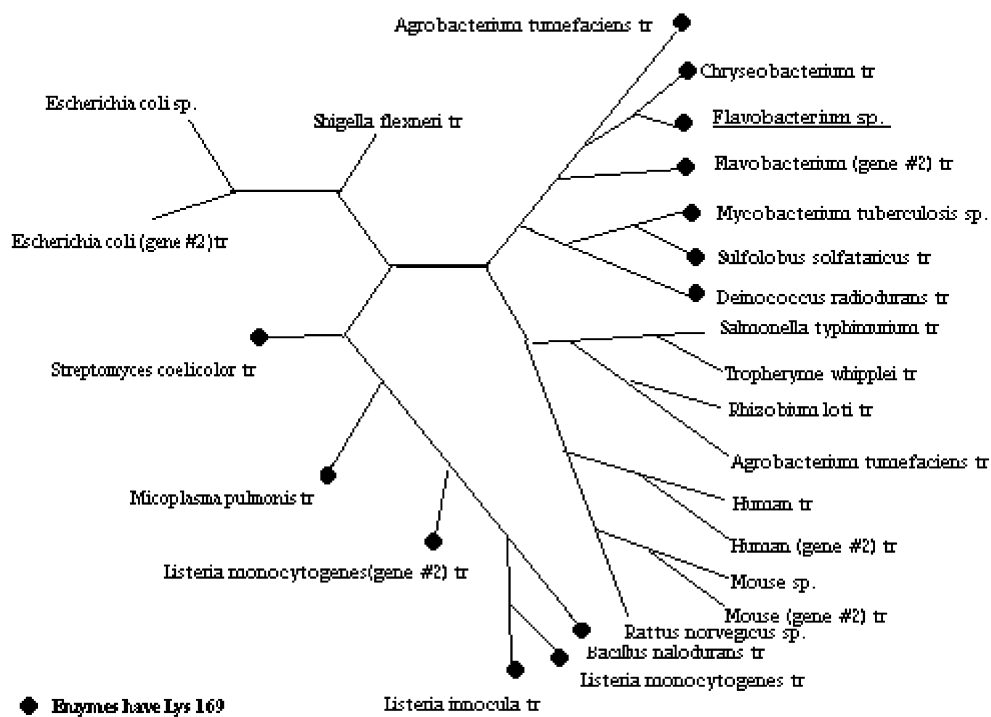


Fig. 2 The phylogenetic three of phosphotriesterases.

## CHEMICAL PROTEOMICS: MANAGEMENT OF BIOSYSTEMS BY SELECTIVE LIGANDS

The regulation of many processes taking place in the biological systems can be organized through application of effectors special for some key biomolecules. Thousands of unknown proteins have been identified through genomic and proteomic research. It is important to discover specific molecules that can interact with them selectively to understand the role of newly discovered proteins and to regulate their activity. The main goal is to find one low-molecular-weight compound able to block or stimulate the activity of one protein in a highly selective way. This area can be identified as “chemical proteomics”. The goal of the chemical proteomics is the creation of “chemical keys” for all proteins existing in nature. Special interest to these tasks is dictated presumably by pharmacological importance. More than 95 % of all human diseases are associated with defective gene expression or post-translation protein modification. Since that, specific molecular ligands can be very useful in explanation and restoration of the functions of biological macromolecules. They also can be used for construction of novel affinity tags to facilitate purification and detection of target proteins or polypeptides. The ligands helpful in control (suppression or activation) of gene expression have high importance. Specific binding between proteins and their ligands is the molecular basis of all essential biological processes. The ways used in identification and characterization of protein sites interacting with other molecules are based on different strategies. One of the modern suggested ways is the following. Selection of appropriate site on the target protein, design of a complementary ligand compatible with the 3D structure of the site, construction of a limited solid-phase combinatorial library of near-neighbor ligands and solution synthesis of the hit ligand, immobilization, optimization, and application of the adsorbent for the purification of the target protein [13]. This strategy is directed to the understanding of affinity and selectivity for enzyme inhibitors uses information from ligands and 3D structures of target proteins [14].

Combinatorial chemistry has emerged as a set of novel strategies for the synthesis of large sets of compounds (combinatorial libraries) for biological evaluation [15]. Combinatorial approaches capable of generating libraries of different proteins allow the directed screening for a particular biological activity. Analysis of proteins in complex with their ligands provides structural information on the amino acid residues involved in ligand binding. Synthetic chemistry enables one to create a large number of different molecules, including simple and more complex peptides, which may contain different amino acids, including unnatural ones, as well as non-peptide small molecules. The design and generation of synthetic molecules that can interfere with protein–ligand interactions represents a promising strategy for the explanation of protein structure and function. In addition to their basic significance, such synthetic proteomimetics are also useful tools for a range of biomedical applications.

A new very fast developing approach to the investigation of protein–ligand interactions is aimed at synthesis of molecules, which, owing to their specific molecular structure, are capable of mimicking the binding sites of natural proteins. Various protein-binding sites are composed of parts that are remote in the amino acid sequence, but brought into proximity by protein folding. They can be mimicked through scaffold peptides, in which the peptide fragments making up the binding site are presented through a molecular scaffold in a nonlinear, discontinuous fashion [16].

This task was initially accomplished by using short synthetic peptides to establish the linear binding sites of proteins. Such peptides, representing overlapping fragments of the protein, were tested individually for binding to the respective ligand in order to identify the protein region(s) responsible for the ligand recognition. The epitopes identified by this way were characterized and optimized regarding their binding affinity to the respective ligands.

Synthetic chemistry should play a significant role in the identification and functional characterization of gene products, as well as in the investigation of proteins' functions through controlled interference with the protein–ligand interactions.

## **BIOCATALYSIS: ENZYMES OF NEW GENERATION**

The methods of bioinformatics allow identification of comprehensive sets of genes encoding enzymes in various organisms as well as comparative and functional analysis of these proteins. For example, the problem of enzyme thermostability could be solved at the genetic level by comparing genomes of mesophilic and thermophilic microorganisms. Genomic access to thermostable enzymes and proteins may result in development of unique technologies of high-temperature biotechnological processes [17].

Enzymes are the most widespread catalysts that can be obtained from renewable raw materials. Biocatalysis is the basis of important chemical processes. During the past two centuries, organic chemistry was substantially oriented toward the transformation of hydrocarbons and their different chemical derivatives, whereas the chemistry of the 21<sup>st</sup> century will be dealing with the problems of chemical transformation of renewable raw materials such as carbohydrates, biomass components, and carbon dioxide. Biocatalysts are the best tools for the chemical modification of molecules in the 21<sup>st</sup> century.

The detailed structural information at the atomic level is already known for a large variety of enzymes from X-ray crystallographic and NMR spectrometric methods. The existing concepts explain the observed effects of acceleration of chemical reactions and the nature of enzyme specificity and are based on fundamental physicochemical laws [18].

During the last 50 years, enzymes and biocatalytic systems were realized in the wide range of applications, including fine organic synthesis and multiple forms of chemical analysis. Enzymes provided the basis of many processes in food industry, industrial production, and utilization of detergents. They served as key components in environmental biotechnology and medicine and improved chemical and biological safety of various industrial processes.

Many “weaknesses” of enzymatic catalysis were overcome during the past decades. Thus, the methods of effective stabilization of enzymes were developed and the new conditions for their exploitation were discovered, and new methods of the use of enzymes in the media with organic solvents

at increased temperatures were described. Methods of genetic engineering and site-directed mutagenesis played a significant role in elucidating reaction mechanisms of various enzymes and providing molecular basis for their biotechnological applications. Potentially all these features make enzymes the most effective biocatalysts. The up-to-date prognosis suggests that within one decade, total enzyme production in the world will reach the level of or even exceed the production of "classic" catalysts used in the chemical industry within the next decade.

The creation and construction of enzymes with new properties is based on the application of genetic information and modern achievements of genomics. It is possible to expect the appearance of new scientific directions of enzyme investigations and the development of new biocatalytical technologies.

The use of new methods of preparation of gene-expressed proteins obtained from unnatural amino acids might be capable of supporting the development of principally new catalysts, considerably widening the possibilities of biocatalysis. Biosynthesis of enzymes with unnatural amino acids included in the polymer chain could result in the development of new types of active sites, solving the problem of the use of enzymatic reactions under extreme conditions (extreme pH, temperature, salt content, etc.). The chemical modification of active sites by unnatural amino acids should result in the application of enzyme families with altered catalytic efficiencies and reaction mechanisms, and those that are characterized by transformed specificity and enantioselectivity.

In addition to methods traditionally used in protein design, such as DNA-shuffling, and directed evolution, it is anticipated that new methods allowing generation of enzymes with predefined properties will emerge. In this regard, the new methods allowing construction of proteins from the limited number of essential amino acids are quite attractive. These approaches should result in deeper understanding of physiological functions of proteins and lead to the appearance of new biocatalyst families of high applied significance.

## **BIOCATALYSIS: NEW RESOURCE TECHNOLOGIES**

The limitations of traditional sources used in the chemical industry as well as limitations of hydrocarbon sources of energy, illustrate the need for development of new raw and energy sources. Growing interest in the biocatalytic conversion of biomass as the initial chemical raw materials for production of new materials, polymers, and energy sources (i.e., hydrogen, methane, ethanol, and diesel oil) may be envisioned as characteristic of the next decades [19]. Significant progress may be expected in the development of biosystems catalyzing the reduction of carbon dioxide to the basic carbon sources for the chemical industry (organic acids, alcohols, and monomers) [20].

## **ENZYME CATALYTIC SITES AND MECHANISM OF CATALYSIS FROM SEQUENCE DATA**

Computer methods are widely employed in modern chemistry and molecular biology for various purposes [21,22]. During the last decade, two approaches, bioinformatics and molecular modeling, have become especially popular and intensively developing areas. The methodology of bioinformatics is based on the informational analysis of nucleotide sequences and proteins [23–30]. Within a framework of adequate physical considerations, molecular modeling allows to characterize protein structures and their changes induced by some treatments or local structural alterations. In combination with methods of molecular dynamics, molecular modeling represents a powerful approach that provides understanding of numerous physical and chemical aspects of protein molecules. Graphic capacities of modern work stations, relatively low cost and potency of modern processors, storage capacity of accumulated information, availability of a large number of resolved spatial protein structures, and nucleic acids and their complexes are the background for rapid worldwide distribution of computer modeling methods among scientists. These methods of computer modeling are widely used for determination of structures

of biological molecules (X-ray analysis and multimer nuclear magnetic resonance), theoretical studies of their interactions, studies of the interaction between membrane and membrane proteins, structural analysis of site-directed mutagenesis and prediction of new mutation sites, and protein structure prediction (method of homologous modeling).

Well-known software allow the construction of various biological molecules, edit their structure (by substituting some groups, changing conformation of molecular groups) followed by subsequent structural relaxation by means of methods of molecular mechanics and dynamics. These programs also allow the calculation of various characteristics of the constructed molecules (e.g., electrostatic potential, solvent accessible surface, etc.), study of interactions between various molecules, alignment of protein sequences, creation of homology-based structures, and the investigation of ligand docking to active sites of enzymes.

### **Multiple alignment of amino acid sequences allows recognition of the enzyme catalytic site**

Amino acid sequence determines structure and properties of each protein. Now, good evidence exists that in almost endless variability of proteins some structural elements are rather conservative, and these elements mainly determine the function of the protein molecule. This is especially demonstrative in the case of catalytic proteins. For example, in the case of hydrolases, which represent about one-third of all enzymes (about 1100 of 3700 enzymes) listed in the enzyme classification, only four main types of sites forming the catalytic structure are known [31]. Consideration of active site structure of enzymes requires subdivision of the active site into two structural constituents [31–33]:

- substrate-binding subsite, which is responsible for binding, fixation, and certain orientation of substrate(s); it determines enzyme specificity;
- catalytic subsite, which is responsible for chemical transformation of substrate molecule; this site usually employs general acid-base catalysis.

It is possible that within one large enzyme superfamily the substrate-binding subsite responsible for the enzyme specificity exists as quite variable protein structure corresponding to variations in the substrate structures. However, catalytic sites should represent rather conservative structural elements due to a limited number of catalytic site types. To test this hypothesis, we have employed the bioinformatic approach based on the comparison of amino acid sequences of proteins constituting one large family. We analyzed results of sequence alignment of a few large enzyme families from HSSP database (<<http://www.sander.embl-heidelberg.de/hssp/>>) [33]. These enzyme families were selected by the following criteria:

- the number of analyzed enzymes should exceed 100; this provides reasonable statistical significance of the results;
- this analysis requires selection of enzymes with known active site structures and well documented catalytic mechanism.

One of the quantitative criteria of position conservatism for each residue in the protein sequence is the statistical criterion in the form of Shannon entropy. In information theory, Shannon entropy is one of the most important functions. The informational entropy (Shannon entropy) is a very convenient function for comparison of related proteins with distinct amino acid sequences. The alignment procedure vs. some reference protein represents sequence positioning one over another one, followed by fixation of homologous sites, and recognition and eliminations of inserts. Such comparison of a large number of protein sequences allows the calculation of probability of localization of some amino acid residue in a certain position. This probability is determined as relative frequency of the amino acid  $j$  in

a given position  $i$ . The entropy function for all 20 amino acids in each position in the protein sequence can be calculated using the following formula:

$$H_j = -\sum_i p_i^j / \log_2 p_i^j /$$

This function tends to zero in the case of events with high ( $p_i^j \rightarrow 1$ ) and low ( $p_i^j \rightarrow 0$ ) probability. So calculation of Shannon entropy may recognize positions in the protein sequence which are common (absolutely conservative) for a given  $j$  amino acid in the whole protein family. This is a position in which probability of appearance of this amino acid is close to unity, whereas in all other cases probability tends to zero. High values of Shannon entropy are typical for positions in protein sequences characterized by high variability of amino acid residues, whereas low values are indicative for positions of conservative residues. Within  $p_i^j \rightarrow 1$  (absolute conservatism)  $H_j \rightarrow 0$ .

Hydrolases were used as the object of research. Four mechanisms of water activation in reactions catalyzed by these enzymes are known [34]. They involve:

- aspartic (glutamic) acid carboxyl group, water activation by nucleophilic mechanism;
- histidine imidazole group, water activation by nucleophilic mechanism;
- complex with zinc or cobalt ions, water activation by electrophilic mechanism;
- complex with magnesium or manganese ions, water activation by electrophilic mechanism.

The following enzyme families representing these four mechanisms of water activation were chosen: pepsin family (activation by carboxyl group), chymotrypsin and subtilisin family (activation by imidazole group), alkaline phosphatase family (activation by complex with zinc or cobalt ions), pyrophosphatase family (activation by complex with magnesium or manganese ion). Figure 3 shows alignment profiles for all representatives of these four hydrolase families.

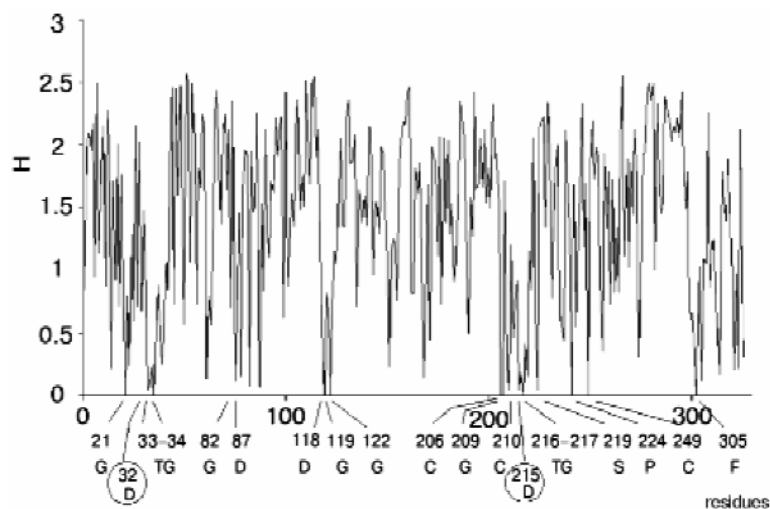


Fig. 3 Shannon entropy for multiple alignments of proteins of aspartyl proteases vs. pepsin.

Analysis of highly conservative amino acids (for which  $H_j \sim 0$ ) revealed that during alignment, amino acid residues forming catalytically active subsites always represent conservative elements of the protein sequence. The catalytic site of acidic pepsin type proteases includes carboxyl groups of Asp31 and Asp215. These residues are recognized during alignment of amino acid sequences of pepsin family as conservative positions characterized by minimal value of Shannon entropy (Fig. 3).



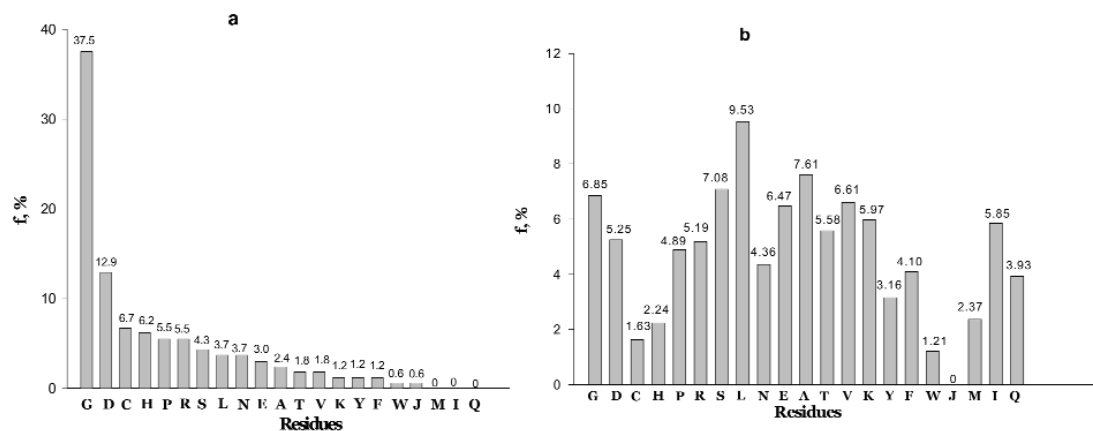
The catalytic site of the chymotrypsin family includes Ser195, His57, and Asp102, whereas the catalytic site of the alkaline phosphatase family includes Asp51, Asp369, His370, Asp327, His412, His331, and Ser102. Catalysis by inorganic phosphatase involves Glu20, Asp65, Asp70, and Asp102 [35].

All these residues are recognized as conservative ones during multiple alignment of amino acid sequences of corresponding protein families. Thus, the bioinformatic approach allows recognition of side chains of amino acid residues forming a catalytic site of the enzymes and responsible for nucleophilic/electrophilic substrate conversion.

The comparison of enzyme amino acid sequences also revealed that Gly and Asp are the most frequently recognized as absolutely conservative residues [31,36]. The finding that Gly is the most conservative residue was rather unexpected. Asp takes the second position in this list, and the sum of Gly and Asp represents about 50 % of all conservative residues recognized.

Amino acid residues were ranked by their conservatism in these four protein families. For each amino acid, we determined its frequency as the conservative element ( $H_j \sim 0$ ), normalization on total number of conservative positions for all amino acid residues in these families. Figure 4a shows rating of amino acid conservatism. Figure 4b shows frequency of amino acid residues in proteins determined using Swiss-Prot database. Total frequency of conservative residues in enzymes completely differs from total frequency of amino acid residues in proteins, where Leu is the most frequent residues, whereas Gly takes only the fourth place in this list. (Frequency of Gly as the most conservative residue is 37 %.)

Gly, Asp, Cys, Pro, and His are the most frequent conservative residues in enzymes. They represent about 70 % of all conservative positions in these enzymes, whereas Met and Ile represent the most variable elements in the amino acid sequences. Thus, the most conservative residues can be separated into two principally different groups: (1) residues involved in substrate activation and acting as acids and bases (Asp and His); (2) residues forming active site architecture (Gly, Cys, Pro).



**Fig. 4** Frequency of amino acid residues as conservative elements during multiple alignment of amino acid sequences in enzymes (a) [31], and frequency of amino acids in proteins of Swiss-Prot database (b).

### Aspartic acid and histidine in the enzymatic catalytic cycle

The bioinformatic approach used in this study demonstrates a crucial role of aspartic acid and histidine residues in the functioning of the active site. Let us consider this role in more detail based on the analysis of mechanism of hydrolase action. As mentioned above, hydrolases represent the largest class of all known enzymes and their molecular mechanisms of catalysis have been well characterized. For most hydrolases, functional groups of amino acid residues constituting catalytic sites have also been identified and the interaction between these groups during catalytic cycle has been elucidated.

Based on the active site structure and mechanism of action all hydrolases can be arbitrary subdivided into the four main types:

- hydrolases containing aspartic or glutamic acid residues in the active site (lysozyme-pepsin type);
- hydrolases using imidazole group for water activation (the type of pancreatic ribonuclease, chymotrypsin, subtilisin, papain, lipase);
- hydrolases using complexes with  $Zn^{2+}$  or  $Co^{2+}$  for activation of water and substrate (type of alkaline phosphatase, carboxypeptidase A, and organophosphate hydrolase);
- hydrolases using  $Mg^{2+}$  and  $Mn^{2+}$  for activation of water and substrate (inorganic pyrophosphatase type).

Analysis of the catalytic mechanisms revealed that in most types of catalytic sites aspartic acid and histidine play principally important role [31].

### Role of Gly in formation and conformational flexibility of active structure

It is clear that conservative Gly residues do not play an important role in the activation of water molecules during the catalytic cycle because Gly does not have any substituent at the *alpha*-carbon atom, and therefore it lacks pronounced chemical function.

Nevertheless, Gly residues are essential for protein structure [37–42]. This was demonstrated by mutation experiments where conservative Gly residues were substituted for any other amino acid. As a rule such substitutions resulted either in complete loss or significant decrease in catalytic activity (see [40] as an example).

Apparently, conservative Gly residues are principally important for the following functions.

1. Being a unique amino acid with the most energetically favorable rotation along C–N and C–C bonds of polypeptide chain ( $\varphi$  and  $\psi$  Ramachandran angles) glycine may be a “junction point”, providing change in the direction of polypeptide chain during “assembly” of amino acid residues into functionally competent active site. The essential role of Gly for structural function of enzyme-active sites was demonstrated by molecular modeling method [35]. Thus, the presence of conservative glycines may explain the structural paradox of enzymatic catalysis, when completely identical active sites are formed from completely distinct protein sequences. They just share several common features such as the existence of conservative glycines and factors stabilizing the assembled structure. In this connection, we should mention that cysteine residues involved in disulfide bond formation (one of the most common stabilizing factors) take the third place in the rating of conservatism.
2. Conservative glycine residues may also function as a “hinge” responsible for known conformational flexibility of the active site. In many cases, conservative glycines are located near catalytically active residues. For example, the following conservative motifs were found in hydrolases from various families: Asp215XGly217 (pepsin), Asp170XXGly173 (thermolysin), Asp32XGly, His63Gly64, Gly119XSer221 (subtilisin), Gly17XSer177 (trypsin), His76Gly77, Ser153XGly155, Gly75XAsp177 (lipases). In these enzymes, Asp, Ser, and His residues are from the active site structures.

For some enzymes, values of *phi* and *iota* angles for amino acid residues of the catalytic site are out of the energetically “relaxed limits”. This was found for amino acid residues of *alpha*-chymotrypsin (His57, Asp102, Ser195) by using the Ramachandran map of these residues. The active site of these enzymes is conformationally tensed (values of  $\varphi$  and  $\psi$  angles are in an energetically unfavorable region).

In enzymatic catalysis, the conversion of initial substrate into reaction product(s) involves a series of intermediates possessing distinct structures. Conservative glycines of the active site may function as “relaxing” elements adapting active site conformation for the catalytic conversion of the next intermediate.

Cysteine and proline residues take top positions in the rating of amino acid conservatism. They are also essential for active site formation. Proline is a unique amino acid that unfolds the polypeptide chain.

Cysteine residues provide “required” active site structure by forming a disulfide bridge, which promotes fixation of catalytic residues (often located in different positions of a polypeptide chain) in the active site structure. For many enzymes, disulfide bond formation is the final step terminating active site assembly.

## ENZYMES VS. CLASSICAL CATALYSTS

A wide diversity of enzymes, great potentials of enzyme modification by genetic engineering methods, and a deep insight into enzyme-related mechanisms potentially provide their diverse applications [43,44]. Enzymes are widely applied in fine organic synthesis, medicine, and analysis as active components of washing powders. Enzymes are most widely applied in the food industry. However, enzymes can also be fairly competitive in solving some classical chemical tasks. As example, consider the problem of creation of catalysts for fuel cells. In view of progressing deficit of hydrocarbons, as a basis of modern power engineering [45], much attention is nowadays attracted to fuel cells, i.e., the electrochemical energy converters where fuel oxidation is spatially separated from reduction of oxidizer, usually air oxygen, resulting in the electrochemical potential generated on two electrodes. Fuel cells afford a direct conversion of oxidation energy to the electric form without the intermediate emission of heat. Due to this fact, the efficiency of fuel cells is fairly high, approaching 100 %.

The electrocatalytic problem of fuel oxidation and oxygen reduction is quite involved. On the one hand, the electrocatalyst should accelerate the electron transfer from the electrode matrix to oxygen. On the other hand, it should take an electron out of the fuel molecule without intermediate formation of free-radical particles in both anodic and cathodic reactions. At present, hydrogen–oxygen fuel cells are most elaborated [46,47]. Supposedly, the fuel cells are an alternative to the internal combustion engines for motorcars. The only appropriate catalyst in catalytic and corrosive parameters for hydrogen–oxygen cells is metallic platinum [48]. However, platinum, as electrocatalyst for fuel cells, has some principal drawbacks: first of all, the high price and sensitivity and, hence, the quantity of CO and H<sub>2</sub>S in the fuel. CO actually totally and irreversibly inhibits platinum at 100 ppm concentration of the former in the fuel hydrogen. A principal limitation of platinum as catalysts in fuel cells for motorcars is imposed by highly limited resources for the world production of platinum. To-date, the world production is about 180 t platinum per year. To produce 60 million motorcars per year (the level of motorcar production in 2000), about 5800 t platinum are required according to the most modest estimates.

Biological catalysts are totally renewable and industrially producible in any necessary quantity. To create the enzyme-based fuel cells, we applied hydrogenases (hydrogen oxidation enzymes) and blue copper-containing oxidases of laccase type (oxygen reduction enzyme). The fundamental ground for creation of enzyme fuel cells is the application of biocatalysis, i.e., the enzyme-catalyzed acceleration of electrochemical reactions of electron transfer at the interface electronic conductor/ionic conductor upon direct “electric catalysis” between conductor and enzyme-active site.

The effect of direct electron transfer between conductor and enzyme-active site was disclosed in our laboratory and registered as the discovery (State registration No. 311, 1987). It should be noted that the USSR had the system of registration of priority research “discoveries”, which did not follow from direct theoretical investigations, were unexpected, and promised novel technological advances. It was shown that at the electrodes made of any electron-conducting material, hydrogenase chemisorption in air oxygen establishes a reversible equilibrium of hydrogen potential.

At the electrodes with sorbed laccase, the potential is established that approached the equilibrium potential of oxygen reduction. Hydrogen and oxygen enzyme electrodes underlay the construction of the fuel cell [49], affording us to obtain a record coefficient of conversion for the free energy of hydrogen oxidation by oxygen, 95.7 %. The latest information on enzymatic hydrogen cells was presented

elsewhere [50,51]. The detailed comparison of electrochemical behavior of platinum and hydrogenase in hydrogen oxidation reaction showed that the exchange currents for hydrogenase by 2–3 factors higher than for platinum (referred to the catalyst active site) [52]. Note that the hydrogenase electrodes are not inhibited by H<sub>2</sub>S and practically not inhibited by CO [53].

Thus, for hydrogen activation reaction in electrochemical mode, it turned out that the enzymes have notably higher parameters than the “classical” catalyst. The possibility to produce enzymes in any necessary quantities is the ground for real practical creation of fuel cells with unique technological parameters as energy converters.

## CHEMICALLY MODIFIED PROTEINS: NEW PROPERTIES, NEW CAPABILITIES

Structural and functional proteomics has already identified a large number of proteins with novel folds designed to carry out specific biochemical reactions. Some of these proteins have inspired chemists and biochemists to engineer artificial and semi-artificial proteins, which may be used in unprecedented industrial or pharmacological applications. Therefore, the preparation of “synthetic” proteins with improved properties can become a rapidly developing part of postgenomic chemistry.

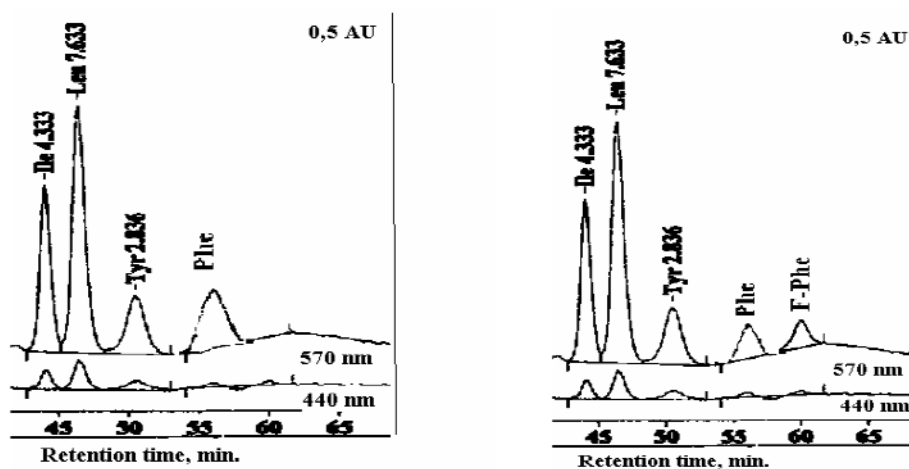
### Fluoro-containing proteins and organisms

The interest to the unnatural amino acid-containing proteins is stimulated by at least three aspects. These are: (1) A potential possibility of the unnatural amino acid incorporation into proteins infinitely widens the spectrum of the novel proteins to be synthesized. The elaboration of the effective methods to synthesize the proteins and enzymes containing element-organic amino acid analogs gives us hope to obtain the proteins and enzymes with the properties previously unknown; (2) Novel classes of anti-cancer and antiviral agents could be proposed using the amino acid analogs and their derivatives [54,55]; (3) The replacement of any amino acid significant part by their synthetic analogs is expected to lead to the formation of microorganisms with qualitatively novel properties. Amino-acyl tRNA synthetase is the key enzyme of amino acid recognition in the protein biosynthesis. Some analogs could not be incorporated into the proteins because of the high specificity of the above-mentioned synthetase. Nevertheless, there are several main ways of unnatural amino acids' incorporation in protein structure [56].

One way represents itself: cultivation amino acid auxotrophic *E. coli* strain with a cloned enzyme gene under the inducible promoter or in the presence of a specified protein expression inhibitor in a medium containing all natural amino acids [57]. After significant biomass amount is accumulated, the medium is replaced and an amino acid analog is introduced in place of its natural counterpart. After a short incubation period, the protein expression is induced by the addition of inducer. The second way supposes chemical charging of the suppressor amino-acyl tRNA with the amino acid analog and an in vitro introduction of the analog at the nonsense mutated positions of the enzyme gene under investigation in the cell-free translation system [58]. Another way to overcome the specificity of amino-acyl tRNA synthetase is mutagenesis of the enzyme [59]. This approach may widen the enzyme specificity and enables one to incorporate the substances from a long analog list.

Realizing discussed approaches, the Tyr-auxotrophic strain *E. coli* B-2935 was transformed with plasmid pTES-His-OPH [60] and inoculated in the medium containing F-Tyr. The biosynthesis of F-labeled organophosphate hydrolase was carried out. The mass-spectrometric analysis (MALDI-TOF) of purified protein testified to the replacement of 80 % Tyr-residues by their F-containing analogs (data not published).

It was shown that phenylalanine analog, namely, 4-fluorophenylalanine, can be introduced into the cellular protein by biosynthetic machinery. It was revealed that the relative content of FPA in the whole yeast protein from the cells grown up in the medium contained 1 mg/ml FPA. The amino acid analysis data are presented in Fig. 5 [56]. It was established that a significant part (35–40 %) of phenyl-



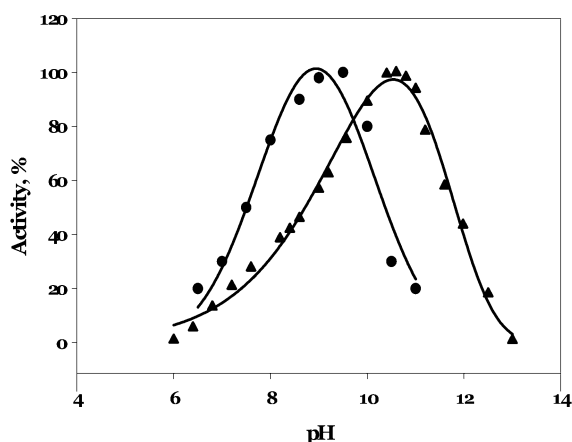
**Fig. 5** Results of amino acid analysis of the total yeast cell protein, obtained without (a) and with (b) substitution of natural phenylalanine by its F-containing analog.

alanine is replaced by FPA. This investigation illustrates a possibility of creating the living systems with a significant part of their natural amino acid replaced by its element-organic analog.

### Hybrid proteins

The use of recombinant proteins has increased significantly in recent years as a response to the rapidly developed fields of proteomics. Recombinant hybrids containing a polypeptide fusion partner, termed affinity tag, to facilitate the purification of the target polypeptides are widely used [61–66]. Many different proteins, domains, or peptides can be fused with the target protein. The production of recombinant proteins in a highly purified and well-characterized form has become a major task for the protein chemistry in the postgenomic era. Several different strategies have been developed to produce recombinant proteins on a large scale. One approach is to use a very small peptide tag that should not interfere with the fused protein properties [61,65]. The effect on tertiary structure and biological activity of fusion proteins with small tags depends on the location and on the amino acid composition of the tag [66]. Another approach is to use large peptides or proteins as the fusion partner since the large partners can increase the stability and solubility of the target proteins [62]. Some hybrid partners are specially used for the investigation of protein–protein interactions or changing properties of a target protein. For example, the catalytic properties of organophosphate hydrolase containing oligohistidine sequence considerably differed from the native enzyme. The optimum pH for hybrid enzyme shifted to the basic area and became equal to 10.5 instead of 9.0 that was typical for the native enzyme (Fig. 6). The thermostability of the developed enzyme increased, and temperature optimum of action shifted from 48 °C up to 55 °C [60,67]. In addition to that, the effective constants for hydrolysis of many substances catalyzed by the hybrid enzyme enhanced 10 times compared to the native form.

The wide application of green fluorescent protein as fusion partner opened new capabilities in the field of proteomics [68,69]. The importance of affinity-tag technology will increase because of its necessity for the peptide/protein chip design, high-throughput purification, peptide/protein libraries, drug delivery systems, and strategies of large-scale production.



**Fig. 6** pH-dependence of enzyme activity of native (circles) and oligohistidine-containing (triangles) organophosphate hydrolase.

## POSTGENOMIC MACROMOLECULAR CHEMISTRY

Future technology will be based on materials with excellent performance and novel functions. There is a large demand for so-called intelligent smart materials: self-recovery, self-adjustment or control, self-diagnosis, stand-by capability for detecting nonlinear onset, ability to be externally tuned, etc. [70].

Recombinant DNA technology can be utilized to produce polymers with specific properties and behavioral characteristics. The production of polymers by recombinant methods affords significant advantages over traditional polymer synthesis technologies, including control of microstructure, stereochemistry, and biocompatibility.

The production of artificial proteins by recombinant methods has already been demonstrated in several laboratories. Some groups have succeeded in the development of artificial proteins with smart properties. The electroactive polymers have been used to make artificial muscles. Hybrid intelligent hydrogels were developed via self-assembly of synthetic polymers and recombinant polypeptides.

Modern gene technology allows expression of any desirable amino acid sequence, and the recent breakthrough in industrial microbial cell fermentation and downstream processing of proteins allows production of large quantities of recombinant polymers. Precise control of the composition and structure of recombinant polymers renders them with unique properties. The introduction of unnatural amino acids in the structure of recombinant polymers will provide them with further unique properties not observed in natural polymers.

Recombinant polymers will combine both high mechanical and chemical stability as construction materials with the ability to react in a strong and predictable way to environmental changes and, hence, perform as active materials. In fact, these polymers will pave a way to soft machines where chemical energy is converted to mechanical energy with high efficiency utilizing the conformational transitions of polymers rather than movement of solid machinery.

The development and production of recombinant polymers is also a response to the demand for “green” alternatives to petroleum-based chemistry and waste-generating stoichiometric reagents. The recombinant polymers are produced by the environmentally friendly process of fermentation of microorganisms. The resources needed for the production of recombinant polymers are renewable, and by-products generated (e.g., microbial biomass) could be easily utilized and converted. Moreover, the recombinant polymers are biodegradable and pose minimal environmental threat as compared to traditional synthetic polymers.

One could easily foresee that recombinant polymers along with other natural polymers like carbohydrates, polyesters, and different composite materials based on these polymers, would to a large extent replace many traditional synthetic polymers.

Some of natural polymers are synthesized using template-guided synthesis. A classical example of this type of the synthesis is polymerase-catalyzed replication of DNA molecules where novel chain is produced by one of two chains of primary DNA. In this case, it allows production of biomolecules identical to the template.

One could envision the development of this approach into several directions:

- development and investigation of chemical models of DNA polymerase; fine-tuning of conditions for the DNA proliferation in the systems *in vitro* with the use of chemical analogs of DNA copolymers;
- widening of possible application of enzymes (polymerases, oxidoreductases, esterases) for the template-directed matrix synthesis (polyaniline, polyethers, etc.); and
- investigation of catalytic synthesis of polymers that employ proliferation of polymer matrix; chemical models of polymerase chain reaction.

Investigations in this field should result in the creation of self-proliferating polymers. The new systems recording the information in chemical language different from the classical biological one could be suggested on the basis of a number of such polymers.

## NEW BIOANALYTICAL CHEMISTRY

A number of new analytical technologies that qualitatively changed the face of modern bioanalysis have been developed in various genomic projects and their further applications. In addition to DNA sequencing methods, considerable progress was achieved in the development of biochip analytical devices and biosensors. The development of proteomics illustrates the application of mass-spectrometry method for the analysis of proteins and peptides.

Advances in genomics, in combination with the progress in nano/materials/info technologies, are being integrated to enable analytical devices and systems with potential global effects on individual and public health, safety, economic, social and political systems.

DNA microarrays have revolutionized the way of nucleic acid characterization. One can easily foresee that protein and small-molecule microarrays will have the same impact in many areas, such as protein structure–function and protein–protein interaction studies, as well as in lead compound identification [71]. Toward such practical microarrays, novel chemistries that ensure selective, reliable, and efficient linkages of synthetic and biological molecules onto solid support are desperately needed. It is reasonable to believe that new reagents and chemoselective reactions operating in water will significantly improve the preparation of useful, stable microarrays that will find applications in many areas, especially in the drug development process.

General principles of DNA polymerase reactions and PCR technique initiated development of a number of atypical chemical processes. It was the first example of single-molecule registration.

A single enzyme molecule can generate polymerization and formation of nanoparticles that can play the roles of markers of immunological interactions and nucleic acid-based reactions. Scanning probe microscopy can be used as a counter of such nanoparticles. This approach promises to be extremely specific at registration of single molecular interactions.

## CONCLUSION

The review concerns the new trends in chemistry development induced by advances in genetic engineering and genomics. The paper is based on the research works performed at the Chemistry Enzymology Department (Chemistry Faculty, The Lomonosov Moscow State University) as well as on

the results of IUPAC Project “Post genomic chemistry” (2001-005-1-300). A few examples were used to demonstrate one of the fundamental and philosophic results of the genomic project, i.e., a possible identification of all molecular participants of biological processes, including the low-molecular-weight substances and biopolymers. Protein identification resulted in the new field “chemical proteomics” aiming at creation of regulators—“chemical keys” (inhibitors, activators, suppressors) for all proteins known to exist in nature. The approaches to creation of a new generation of biocatalysts by the methods of genetic engineering and protein design are analyzed. Application of the methods of bioinformatics to investigation of the nature and functions of enzymes provides the solution of fundamental problems—the identification of catalytic site structure and the mechanism of catalysis from the data on the amino acid sequence in protein polymer chain. The possibilities of biocatalyst use in creation of new resource technologies (fuel production, CO<sub>2</sub> reduction) are demonstrated. Applications of enzymes are discussed for solving classical chemical problems, in particular, upon creation of electrocatalysts effecting the ionization of molecules at the electron conductor/ion conductor interface. The enzymes in the reactions of this type are shown to be by 2–3 factors more efficient compared to metallic platinum, their “classical” catalyst. Potential approaches are shown to create the hydrogen–oxygen fuel cell with 95.7 % conversion efficiency for free energy of fuel oxidation. Possibilities for creation of gene-expressed proteins on principally different chemical ground—from unnatural amino acids—are discussed. This approach is illustrated by the experimental data with application of proteins containing fluorinated aromatic amino acids. A possible approach is shown to design the organisms containing proteins with some amino acids up to 40 % substituted by their fluorinated analogs. Possible approaches are considered to develop polymers using methods of recombinant DNA technology and to design template-directed synthesis of polymers and self-proliferated polymers. One of the results of the genomic project is the creation of some new analytic technologies underlying the development of the new bioanalytic chemistry. The advances in elaboration of biosensors, biochips, the protein and small-molecule microarrays and the systems for recording single molecules are discussed.

## ACKNOWLEDGMENTS

The authors would like to thank following scientists for their help in the preparation of the text and fruitful participation in the IUPAC project “Post genomic chemistry” (2001-005-1-300): Jutta Eichler (German Research Center of Biotechnology, Germany); Igor Galaev (Lund University, Sweden); Vadim Gladyshev (University of Nebraska, USA); Thomas Haertle (Institute of National Research Agronomy, France); Arkadiy Karyakin (Moscow State University, Russia); Ilia Kurochkin (Moscow State University, Russia); Marian Mikolajczyk (Center of Molecular and Macromolecular Studies Pol. Acad. Sci, Poland); Vladimir Poroikov (Institute of Biomedical Chemistry Russ. Acad. Med. Sci., Russia); Fritz Spener (University of Munster, Germany); Normand Voyer (Laval University, Canada); and James Wild (Texas A&M University, USA).

## REFERENCES

1. G. F. Cambs Jr., L. C. Clark, B. W. Tumbull. *Biofactors* **14**, 153–156 (2001).
2. D. L. Hatfield and V. N. Gladyshev. *Mol. Cell. Biol.* **22**, 3565–3569 (2002).
3. G. V. Kryukov, S. Castellano, S. V. Novoselov, A. V. Lobanov, O. Zehtab, R. E. Guigo, V. N. Gladyshev. *Science* **300**, 1439–1443 (2003).
4. G. V. Kryukov, V. M. Kryukov, V. N. Gladyshev. *J. Biol. Chem.* **274**, 33888 (1999).
5. E. N. Efremenko and V. S. Sergeeva. *Izv. Akad. Nauk, Ser. Khim.* 1743–1749 (2001).
6. V. S. Sergeeva, E. N. Efremenko, G. M. Kazankov, A. K. Gladilin, S. D. Varfolomeev. *Biotechnol. Tech.* **13**, 479–485 (1999).
7. V. S. Sergeeva, E. N. Efremenko, G. M. Kazankov, S. D. Varfolomeev. *J. Mol. Cat. Part B: Enzymatic* **10**, 571–579 (2000).



8. E. I. Rainina, E. N. Efremenko, S. D. Varfolomeyev, A. L. Simonian, J. R. Wild. *Biosens. Bioelectron.* **11**, 991–1000 (1996).
9. A. L. Simonian, E. N. Efremenko, J. R. Wild. *Anal. Chem. Acta* **444**, 179–186 (2001).
10. E. N. Efremenko, V. I. Lozinsky, V. S. Sergeeva, F. M. Plieva, T. A. Makhlis, G. M. Kazankkov, A. K. Gladilin, S. D. Varfolomeyev. *J. Biochem. Biophys. Methods* **51**, 195–201 (2002).
11. I. G. Danilova, A. D. Ryabov, S. D. Varfolomeev. *J. Mol. Catal., Part A: Chemical* **118**, 161–166 (1997).
12. R. S. Gold, M. E. Wales, J. K. Grimsly, J. R. Wild. In *Enzymes in Action: Green Solutions for Chemical Problems*, NATO Science Series: 1. Disarmament Technologies, B. Zwanenburg, M. Mikolajczyk, P. Kielbasinski (Eds.), pp. 263–286, Kluwer, Dordrecht (2000).
13. K. Sproulea, P. Morrilla, J. C. Pearson, S. J. Burton, K. R. Hejnæs, H. Valorec, S. Ludvigsen, C. R. Lowea. *J. Chromatogr. B: Biomed. Sci. Appl.* **740**, 17–33 (2000).
14. M. Hans and S. Wilfried. *J. Med. Chem.* **42**, 4506–4523 (1999).
15. L. V. Hijfte, G. Marciniak, N. Froloff. *J. Chromatogr. B: Biomed. Sci. Appl.* **725**, 3–15 (1999).
16. R. A. Houghten, C. Pinilla, J. R. Appel, S. E. Blondelle, C. T. Dooley, J. Eichler, A. Nefzi, J. M. Ostresh. *J. Med. Chem.* **42**, 3743–3778 (1999).
17. *Materials of Intern. Workshop High Temperature Biotechnology*, Petropavlovsk-Kamchatsky, Russia, 16–22 May 1995, Logos, Moscow (1996).
18. M. Garcia-Viloca, J. Gao, M. Kaplus, D. G. Truhlar. *Science* **303**, 186–195 (2004).
19. S. Varfolomeyev and S. Kalyuzhny. In *Enzymes in Action. Green Solutions for Chemical Problems*, NATO Science Series, B. Zwanenburg, M. Mikolajczyk, P. Kilbasinski (Eds.), Kluwer, Dordrecht (2000).
20. E. I. Rainina, M. A. Pusheva, A. M. Ryabokon, N. P. Bolotina, V. I. Lozinsky, S. D. Varfolomeyev. *Biotechnol. Appl. Biochem.* **19**, 321–329 (1994).
21. M. S. Gelfand and A. A. Mironov. *Mol. Biol. (Moscow)* **33**, 969–984 (1999).
22. M. Pellegrini. *Curr. Opin. Chem. Biol.* **5**, 46–50 (2001).
23. S. F. Altshul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lippman. *Nucleic Acids Res.* **25**, 3389–3402 (1999).
24. A. Bairoch and R. Apweiler. *Nucleic Acids Res.* **28**, 45–48 (2000).
25. A. Bateman, E. Birney, R. Durbin, S. R. Eddy, K. L. Howe, E. L. Sonnhammer. *Nucleic Acids Res.* **28**, 263–266 (2000).
26. K. Hofmann, P. Bucher, L. Falquet, A. Bairoch. *Nucleic Acids Res.* **27**, 215–219 (1999).
27. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. H. Bhat, H. Weissing, I. N. Shindyalov, P. E. Bourne. *Nucleic Acids Res.* **28**, 235–242 (2000).
28. A. Bairoch and R. Apweiler. *Nucleic Acids Res.* **28**, 45–48 (2000).
29. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Kapp, D. L. Wheeler. *Nucleic Acids Res.* **28**, 15–18 (2000).
30. G. Stoesser, W. Backer, A. van den Brock, E. Camon, M. Gareia-Pastor, C. Kanz, T. Kulikova, V. Lombard, R. Lopes, H. Parleison, N. Redaschi, P. Sterk, P. Stoehr, M. A. Tulu. *Nucleic Acids Res.* **29**, 17–21 (2001).
31. S. D. Varfolomeev and K. G. Gurevich. *Izv. Akad. Nauk, Ser. Khim.* 1629–1637 (2001).
32. D. J. Voadlo, G. J. Davies, R. Laine, S. G. Withers. *Nature* **412**, 835–838 (2001).
33. C. Sander and R. Schneider. *Proteins* **9**, 56–68 (1991).
34. S. D. Varfolomeev and A. E. Pozhitkov. *Vestnik Mosk. Univer., Ser. 2, Khimiya* **41**, 147–156 (2000).
35. S. D. Varfolomeyev, I. V. Uporov, E. V. Fedorov. *Biochemistry (Moscow)* **67**, 1099–1108 (2002).
36. S. D. Varfolomeev, K. G. Gurevich, V. V. Poroikov, B. N. Sobolev, A. E. Fomenko. *Dokl. RAN* **379**, 548–550 (2001).
37. N. G. Esipova and V. G. Tumanyan. *Mol. Biol. (Moscow)* **6**, 840–849 (1972).
38. M. Sasai. *Proc. Natl. Acad. Sci. USA* **92**, 8438–8445 (1995).

39. I. Friedberg, T. Kaplan, H. Margalit. *ISMB* **8**, 162–170 (2000).
40. Z. Q. Cheng and B. A. McFadden. *Protein Eng.* **11**, 457–465 (1998).
41. A. C. Martin and J. M. Thornton. *J. Mol. Biol.* **263**, 800–810 (1996).
42. A. V. Efimov. *FEBS Lett.* **355**, 213–217 (1999).
43. J. R. Wild, S. D. Varfolomeyev, A. Scozzafava (Eds.). *Perspectives in Bioremediation. Technologies for Environmental Improvement*, NATO ASI Series, Kluwer, Dordrecht (1997).
44. B. Zwanenburg, M. Mikolajczyk, P. Kilbasinski (Eds.). *Enzymes in Action. Green Solutions for Chemical Problems*, NATO Science Series, Kluwer, Dordrecht (2000).
45. R. W. Bentley. *Energy Policy* **30**, 189–205 (2002).
46. J. O. M. Bockris. *Int. J. Hydrogen Energy* **24**, 1–15 (1999).
47. L. Barreto, A. Makihiraa, K. Raihia. *Int. J. Hydrogen Energy* **28**, 267–284 (2003).
48. R. J. Farrauto. *Sci. Technol. Catal.* 21–29 (2003).
49. A. I. Yaropolov, A. A. Karyakin, S. D. Varfolomeyev, I. V. Berezin. *Bioelectrochem. Bioenerg.* **12**, 267–276 (1984).
50. S. D. Varfolomeyev, A. I. Yaropolov, A. A. Karyakin. *J. Biotechnol.* **27**, 331–344 (1993).
51. S. V. Morozov, E. E. Karyakina, N. A. Zorin, S. D. Varfolomeyev, S. Cosnier, A. A. Karyakin. *Bioelectrochemistry* **55**, 169–171 (2002).
52. S. V. Morozov, E. E. Karyakina, O. A. Zadorny, N. A. Zorin, S. D. Varfolomeyev, A. A. Karyakin. *Electrochemistry (Russ.)* **38**, 113–114 (2002).
53. A. A. Karyakin, S. V. Morozov, E. E. Karyakina, S. D. Varfolomeyev, N. A. Zorin, S. Cosnier. *Electrochem. Comm.* **4**, 417–420 (2002).
54. Y. Kiso. *Biopolymers* **40**, 235–239 (1996).
55. R. Larsson, S. Dhar, H. Ehrsson, P. Nygren, R. Lewensohn. *Br. J. Cancer* **78**, 328 (1998).
56. A. E. Pojtkov, E. N. Efremenko, S. D. Varfolomeyev. *J. Mol. Catal. B: Enzymatic* **10**, 47–55 (2000).
57. J. F. Parsons, G. Xiao, G. L. Gilliland, R. N. Armstrong. *Biochemistry* **37**, 6286–6293 (1998).
58. J. A. Ellman, D. Mendel, P. G. Schultz. *Science* **255**, 197–202 (1992).
59. D. R. Lju, T. J. Magliery, M. Pastrnak, P. G. Schultz. *Proc. Natl. Acad. Sci. USA* **94**, 10092 (1997).
60. E. N. Efremenko, Yu. A. Viotchitceva, T. Aliev, S. D. Varfolomeyev. Application for Russian Federation Patent, # 2003136646.
61. M.-Q. Xu and H. Paulus. *Methods Enzymol.* **236**, 376–418 (2000).
62. K. Terpe. *Appl. Microbiol. Biotechnol.* **60**, 523–533 (2003).
63. P. Sweda, P. R. Pladzyk, R. Kotlowski, J. Kur. *Protein Expr. Purif.* **22**, 467–471 (2001).
64. D. Sachdev and J. M. Chirgwin. *Methods Enzymol.* **326**, 312–321 (2000).
65. C. M. Halliwell, G. Morgan, C. P. Ou, A. E. Cass. *Anal. Biochem.* **295**, 257–261 (2001).
66. M. H. Bucher, A. G. Evdokimov, D. S. Waugh. *Biol. Cryst.* **58**, 392–397 (2002).
67. B. DiSoidi, J. K. Grimsly, K. Lai, J. R. Wild. *Biochemistry* **38**, 2866–2872 (1999).
68. D. Rochu, N. Beaufet, F. Renault, N. Viguie, P. Masson. *Biochim. Biophys. Acta* **1594**, 207–218 (2002).
69. W. F. Patton. *J. Chromatogr., B* **771**, 3–31 (2002).
70. V. I. Lozinskii, I. A. Simenel, E. A. Kurskaya, V. K. Kulakova, V. Ya. Grinberg, A. S. Dubovik, I. Yu. Galaev, B. Mattiasson, A. R. Khokhlov. *Dokl. Chem. (Engl. Transl.)* **375**, 273–276 (2000).
71. S. D. Varfolomeyev, I. N. Kurochkin, A. I. Yaropolov. *Biosens. Bioelectron.* **11**, 863–871 (1996).