Pure Appl. Chem., Vol. 76, No. 11, pp. 1985–1999, 2004. DRAFT

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

ORGANIC AND BIOMOLECULAR CHEMISTRY DIVISION*
CHEMISTRY AND HUMAN HEALTH DIVISION**

POSTGENOMIC CHEMISTRY

(IUPAC Technical Report)

Prepared for publication by
SERGEY VARFOLOMEYEV[1,‡], ELENA EFREMENKO[1], IRINA BELETSKAYA[1], IVANO BERTINI[2], G. MICHAEL BLACKBURN[3], ALEXEY BOGDANOV[4], RAIMOND CUNIN[5], JUTTA EICHLER[6], IGOR GALAEV[7], VADIM GLADYSHEV[8], DAVID O'HAGAN[9], THOMAS HAERTLE[10], JAAK JARV[11], ARKADIY KARYAKIN[1], ILIA KUROCHKIN[1], MARIAN MIKOLAJCZYK[12], VLADIMIR POROIKOV[13], IVAN SAKHAROV[1], FRITZ SPENER[14], NORMAND VOYER[15], AND JAMES WILD[16]

[1]Chemical Enzymology Department, The M. V. Lomonosov Moscow State University, Lenin's Hills, 1/11, Moscow 119992, Russia; [2]Magnetic Resonance Center, University of Florence, Italy; [3]University of Sheffield, UK; [4]Belozersky Institute, Moscow State University, Russia; [5]Vrije University, Belgium; [6]German Research Center of Biotechnology, Germany; [7]Lund Universitiet Brussel, Sweden; [8]University of Nebraska, USA; [9]University of St. Andrews, UK; [10]Institute of National Research Agronomy, France; [11]University of Tartu, Estonia; [12]Center of Molecular and Macromolecular Studies, Polish Academy of Sciences, Poland; [13]Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Russia; [14]University of Munster, Germany; [15]Laval University, Canada; [16]Texas A&M University, USA

*Membership of the Organic and Biomolecular Chemistry Division during the final preparation of this report (2002–2003) was as follows:
***President*:** T. T. Tidwell (Canada); ***Vice President*:** D. StC. Black (Australia); ***Secretary*:** G. M. Blackburn (UK); ***Titular Members*:** Atta-ur-Raman (Pakistan); S. E. Braslavsky (Germany); M. T. W. Hearn (Australia); M. Isobe (Japan); V. T. Ivanov Russia); G. J. Koomen (Netherlands); G. Q. Lin (China); L. McElwee-White (USA); N. S. Nudelman (Argentina); C. L. Perrin (USA); ***Associate Members*:** A. D. McNaught (UK); G. Mehta (India); F. Nicotra (Italy); W. H. Powell (USA); M.-F. Ruasse (France); ***National Representatives*:** J. A. R. Rodrigues (Brazil); A. M. P. Koskinen (Finland); S. S. Kim (Korea); A. Rauter (Portugal); T. G. Watson (South Africa); ***Provisional Members*:** B. M. Abegaz (Botswana); E. Dagne (Ethiopia).

**Membership of the Chemistry and Human Health Division during the final preparation of this report (2002–2003) was as follows:
***President*:** A. Kallner (Sweden); ***Titular Members*:** M. S. Chorghade (USA); J. M. Christensen (Denmark); R. Cornelis (Belgium); P. W. Erhardt (USA); U. Forsum (Sweden); C. R. Ganellin (UK); B. Heinzow (Germany); M. N. Liebman (USA); T. J. Perun (USA); ***National Representatives*:** P.-E. Bost (France); G. Gaviraghi (Italy); M. Totani (Japan); N.-S. Park (Korea); N. E. Nifantiev (Russia).

‡Corresponding author

Postgenomic chemistry

(IUPAC Technical Report)

Abstract: The identification of the most promising areas of chemistry that use genomic information is discussed. Attention is paid mainly to perspectives of development of novel trends in chemistry in the postgenomic era. Among these are: combinatorial approaches to the development of organic chemistry; management of biosynthesis by selected ligands; development of artificial proteins containing unnatural amino acids and proteins including trace elements; creation of and research on new types of enzymes; use of achievements of postgenomic chemistry for drug design; identification of lipid networks and global characterization of lipid molecular species; development of recombinant and self-proliferating polymers; and approaches to the development of food chemistry and bioanalytical chemistry based on new nanoanalytical systems and novel recognition elements.

1. INTRODUCTION

The past decade has been characterized by the dynamic integration of chemistry and biology. Chemistry, as the science investigating the structure and properties of molecules, comprises the fundamental basis for the development of biochemistry and molecular biology. Chemical science actively investigates and uses the principles of molecular organization of processes typical for so-called "living" systems. Chemistry and chemical technology peculiarly use biological materials and catalysts for the creation of a steadily increasing variety of useful molecules. A qualitatively new state of integration of branches of modern science with chemistry is one result of genome studies. The creation of a harmonious knowledge system of the nature of genetic information, genetic materials, and structures of genomes influences the development of chemistry, providing it with evolutionary new representatives, methods, and molecular structures. In the framework of proteomics, the problem of identification of chemical structure and function of all proteins catalyzing the processes in biosystems seems to be within reach. The complete elucidation of the human genome and genomes of many other organisms is a remarkable scientific breakthrough that will have an impact on our lives for decades to come.

However, there are numerous actual limitations that need to be overcome before exploiting successfully the wealth of information from genomic studies. Although not at the forefront during recent genomic efforts, chemistry will undeniably 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 be the basis of novel, improved detection methods that will speed up and facilitate diagnostics. Organic chemistry will have an even greater impact on biological sciences. All this provides chemistry with new possibilities and sets new tasks and challenges.

Moreover, the functioning of a living organism is realized at higher (compared to the genome) levels of organization. Therefore, profound studies of proteometabolism and physiology of cells are under way. Another important feature is the multidisciplinarity feature of problems, which can be solved only by the concerted actions of scientific teams, including different specialists (chemists, biologists, clinicians, mathematicians, and others).

Questions at this level are a part of IUPAC's mission and agenda. It is necessary to discuss the development direction of different fields of chemistry, informed by our knowledge of genetic material and genome structure as well as by the application of methodology and methods suggested by genomics. Obviously, the achievements of genomics and proteomics could influence considerably the solution of

classical chemistry tasks as well as the formation of new scientific directions and new-perspective scientific projects. To discuss all these problems, a working group was formed in the framework of IUPAC's project "Postgenomic chemistry". This report is the result of its activities. The text is based on the opinions of specialists working in the various fields of biomolecular chemistry. The materials and opinions presented reflect the state of the problem and could be useful in the formation of strategic directions for the future development of up-to-date chemical science.

2. ORGANIC CHEMISTRY: SYNTHETIC CHEMISTRY AND COMBINATORIAL APPROACH

2.1 Chemical proteomics. Management of biosynthesis by selective ligands: Unique biomolecular tools

In living systems, every biomolecule has one or several effector molecules that regulate its activity in a precise manner. Disruption of the ligand/receptor balance leads almost unavoidably to disease conditions. Through genomics and proteomics research, thousands of unknown proteins that are either important for cell survival or overexpressed during a disease state have been identified. To understand the role of these proteins and eventually to regulate their activity, it is of utmost importance to discover small molecules that interact with them selectively. The ultimate goal here is to have one low-molecular-weight compound to block or activate highly selectively the activity of every protein of an organism. Such specific molecular ligands will be tremendously useful to elucidate the function of biological macromolecules and, hence, can be regarded as biomolecular tools. Likewise, they could be used to engineer novel affinity media to facilitate the purification of interesting proteins. This field of research can be coined "chemical proteomics", in parallel with chemical genomics where small molecular effectors are used to control the expression of genes.

In order to formulate proteome-wide libraries of selective low-molecular-weight effectors, it will be essential to develop novel methodologies to prepare in a high-throughput fashion several thousands of analogs of pure small molecules of tremendous structural diversity. Therefore, supported and solution-phase synthetic methods, as well as automated efficient purification techniques, are needed to complement the actual tools available to chemists and to give access to novel ligands with unique molecular scaffolds.

2.2 Artificial proteins

Structural and functional proteomics has already generated a large number of proteins having novel folds evolved to carry out specific biochemical reactions. These proteins will undoubtedly inspire chemists and biochemists to engineer artificial proteins to be used in unprecedented industrial applications. Hence, the preparation of "synthetic" proteins with improved properties is an area that is meant to expand rapidly in the postgenomics era.

Although artificial proteins can be prepared from scratch, the incorporation of unnatural amino acids into known proteins is useful for many applications. On the one hand, it could tremendously facilitate protein structural studies by NMR and X-ray crystallography. For example, incorporating heavy atoms, such as selenium, speeds up structure determination through the use of the multiwavelength anomalous dispersion (MAD) technique. Actually, incorporation is done by biotechnology using selenomethionine, but there is a need for other methods that permit incorporation of other heavy atoms. On the other hand, unnatural amino acids can also serve to enhance structural stability of protein biocatalysts, therefore prolonging their lifetime, and improving industrial processes that are environmentally friendly. Although elegant approaches have been developed for incorporating nonproteogenic amino acids through biotechnology or chemical ligation strategy, development of improved comple-

mentary techniques to these methods is necessary to expand the range of unnatural amino acids that can be incorporated and to augment the diversity of potential artificial proteins that can be created.

Obviously, describing all the areas where organic chemistry will have an impact in the post-genomics area would require a complete issue of this journal. The above examples represent areas where the authors foresee direct and immediate impacts on organic chemistry.

2.3 Synthetic chemistry and combinatorial approaches for molecular recognition and protein function analysis

The molecular bases of all essential biological processes are specific binding events, which are initiated by molecular recognition between proteins and their ligands. The systematic study of molecular recognition phenomena is, therefore, an important element in the structural understanding of these binding events, which are prerequisites for the controlled interference with biological functions mediated by the respective protein–ligand interaction. The strategies used to identify and characterize regions of proteins that are involved in their interactions with other molecules (i.e., the protein binding sites) are based on different methodological repertoires. Using recombinant strategies, mutants of the proteins can be generated in which selected amino acid residues are replaced by others. Biological testing of the mutants reveals whether the altered residues are important for the respective biological function. Combinatorial approaches capable of generating vast populations of different proteins, such as phage display libraries, enable the directed screening for a particular biological activity (e.g., interference with protein–ligand interactions). Structural analysis of the protein in complex with its ligand, on the other hand, provides structural information on the amino acid residues involved in ligand binding. Unlike these strategies, which are limited to the use of proteins as molecular entities, synthetic chemistry enables the generation of a large range of different molecules, including simple and more complex peptides, which may contain much more than the 20 proteinogenic amino acids, peptidomimetics, as well as non-peptide small molecules. Furthermore, modern synthesis instrumentation enables the automated and parallel solid-phase synthesis of up to hundreds of compounds. Consequently, the design and generation of synthetic molecules that can interfere with protein–ligand interactions represents a promising strategy for the exploration an understanding of protein structure and function. In addition to their basic significance, such synthetic proteinomimetics are also useful tools for a range of biomedical applications.

2.4 Synthetic mimicry of biologically active materials

A rational design approach for the interference with protein–ligand interactions is aimed at molecules which, due to their specific molecular architecture, are capable of mimicking the binding sites of natural proteins. As the highly complex structures of proteins have been optimized over millions of years of evolution, developing compounds that can mimic their structure and/or functions is clearly a challenging endeavor. This challenge was initially approached by using short synthetic peptides to elucidate linear binding sites of proteins. Such peptides, representing overlapping fragments of the protein and, as a whole, spanning the entire protein sequence, are tested individually for binding to the respective ligand in order to identify the region(s) of the protein responsible for recognizing the ligand. The epitopes identified in this manner can be characterized further and optimized regarding their binding affinity to the respective ligands by determining the contribution of individual amino acid residues to antibody binding by systematically replacing all positions of the sequence with a range of other amino acids, a process called chemical mutation.

Sequentially discontinuous protein binding sites are composed of parts of the protein that are remote in the amino acid sequence, but brought into special 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.

Drug candidates may evolve from continuous or scaffold peptides through transformation of the peptide structure into non-peptide, "drug-like" structures. Furthermore, combinatorial libraries of synthetic compounds having propensity to mimic defined protein binding sites will be valuable tools for the investigation and characterization of unknown binding specificity of protein–ligand interactions involving newly discovered proteins with potential biomedical relevance.

In conclusion, synthetic chemistry can be expected to play a pivotal role in the identification and functional characterization of gene products, as well as in the development of diagnostic and therapeutic strategies based on modulating protein function through controlled interference with the underlying protein–ligand interactions.

3. POSTGENOMIC CHEMISTRY FOR DRUG DESIGN AND BIO- AND CHEMOINFORMATICS

The elucidation of complete genomes in humans, some mammals, plants, and microorganisms has created quite new possibilities for comparative analysis of different biological taxonomical groups. Such analysis is directed toward the deeper understanding of structural–functional and evolutionary relationships for biological objects, at all levels of their organization.

It is obvious that genetic information encoded in the genome represents only the potential of a particular individual, but realization of this potential strongly depends on the interaction of the individual with the environment.

Developments of new physicochemical methods of analysis are necessary in this field and their application to the study of structure and function of biological objects. They include techniques already used in proteomics, such as 2D-gel electrophoresis and mass spectrometry as well as relatively new biosensor techniques and atomic force microscopy, which are useful in the analysis of intermolecular interactions.

The analysis of proteomes is not the last stage of the structure–function relationship study, because of the important role of posttranslational modification of proteins and their interactions with lipids, carbohydrates, and small endogenic bioregulators. As a result, structural research on glyco-, lipo-, and nucleo-proteins will be accompanied by studying how structural modification could influence biological function. Studies of intermolecular interactions and the formation and function of molecular complexes are also of particular importance.

On the basis of recognized structure–function relationships in biological objects, new effective and safer biologically active compounds can be discovered. Complete research "from genome to drug" can be executed by the methods of bio- and chemoinformatics.

A search of new lead compounds for novel medicines can be performed in the following way:

- analysis of human genome in normal and pathological states or analysis of genomes of pathogenic microorganisms;
- finding of genes encoding the proteins—potential targets for new pharmaceuticals;
- analysis of amino acid sequences of macromolecular targets, determination of their function by experimental and/or theoretical methods;
- experimental elucidation or computer modeling of spatial structure of macromolecular targets and their active sites;
- search for new potential ligands in databases of compounds available for screening or their design de novo, analysis of protein–ligand interactions, estimation of the "binding energy" (scoring function); and

- optimization of the structure of lead compounds to provide the appropriate pharmacodynamics and pharmacokinetics parameters necessary for high efficiency and safety of drug candidates. "From genome to drugs" studies require the integration of existing and development of new computer programs and databases. Such integration of informational and software resources can be realized via the Internet.

It can be forecast with confidence that in the 21st century many problems of chemistry, biology, and medicine that were traditionally decoded by experimental methods will be initially modeled "in silico" by the methods of bioinformatics, with subsequent experimental checking of the computer estimates. Such an approach will create the basis for discovery of new effective and safer pharmaceuticals, significantly decrease the time and cost involved in discovery, and reduce the risk of obtaining negative results.

4. POSTGENOMIC BIOCATALYSIS

4.1 "Mining"of genome information: Identification and comparative analysis of enzymes in completely sequenced genomes

The recent decade can be characterized by considerable progress in the field of enzyme investigation. Enzymes are the most widespread and available catalysts obtained from renewable raw materials. Biocatalysis is the basis of important chemical processes. In past centuries, organic chemistry was mainly oriented toward the transformation of hydrocarbons and their different chemical derivatives, whereas the chemistry of the new century connects with the problems of chemical transformation of renewable raw materials such as carbohydrates, biomass components, and carbon dioxide. Biocatalysts are the instruments for the chemical modification of molecules of the 21st century.

At present, enzymes are the most studied and comprehensible macromolecular catalysts. Detailed structural information at the atomic level is already available for a large number of enzymes through application of X-ray and NMR structure analysis. The existing understanding explains the observed effects of acceleration of chemical reactions and the nature of specificity on the basis of fundamental physicochemical laws that are elementary for modern chemistry.

During the last 50 years, enzymes and biocatalytic systems have found a wide scope of application, including fine organic synthesis and multiple forms of chemical analysis. Enzymes provide the basis of many processes in the food industry, in industrial production, and in utilization of detergents. They possess key positions in environmental biotechnology and medicine as well as in the formulation of chemical and biological safety systems.

Many "weaknesses" of enzymatic catalysis have been overcome in the past few decades. Thus, methods of effective stabilization of enzymes were elaborated, new conditions for their exploitation over long time periods were suggested, and new methods of use of enzymes in media with organic solvents and at increased temperatures were proposed. Methods of genetic engineering as well as site-directed mutagenesis played a significant role in the analysis of enzyme action mechanisms and in the creation of enzyme action fundamentals. Potentially, all of these points make enzymes most effective biocatalysts. The most recent prognosis shows that the total world volume of enzyme production taken in real-value terms will be equal to the production of "classic" catalysts used in the chemical industry within the next decade.

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

Biocatalysis is fundamental for new resource technologies. The limitations of traditional sources used for chemical industry as well as limitations of hydrocarbon energy sources, already apparent, raise acutely questions about the development of new sources of raw materials and energy. The next decades

will be characterized by enhancing interest in the biocatalytic conversion of biomass as initial chemical raw materials for the creation of new materials, polymers, and energy sources, namely, hydrogen, methane, ethanol, and diesel oil. Significant progress should be expected in the creation of biosystems catalyzing the reduction of carbon dioxide to base carbon sources for the chemical industry (organic acids, alcohols, and monomers). Considerable progress also is expected in the use of enzymes as catalysts for electrochemical reactions for the purpose of electrocatalysis in fuel elements and biospecific electrosynthesis.

The role of enzymes is growing in the development of new preparation methods of polymer materials, including new polymers based on peptides, polysaccharides, polyethers, and others.

The availability of a large number of complete genomic sequences allows identification of homologs of previously characterized enzymes. Phylogenetic analyses of these sequences may yield clusters of orthologous enzyme groups and reveal new enzyme families. The data obtained demonstrate that practically unlimited variability of the primary structure of enzymes can result in a very confined number of types of catalytic centers providing the effective transformation of molecules.

4.2 Biocatalysis: New-generation enzymes

Modern methods of bioinformatics allow the identification of enzymes in different genomes as well as the performance of comparative analysis of protein content and molecular characteristics. The problem of enzyme thermostability could be solved at the genetic level by the comparison of mesophilic and thermophilic microorganisms. Genomic access to thermostable enzymes and proteins could underpin the development of unique technologies of high-temperature biotechnological processes.

The development of new preparation methods of gene-expressed proteins obtained from unnatural amino acids is capable of supporting the creation of new catalysts, thus considerably widening the possibilities of biocatalysis. The biosynthesis of enzymes with unnatural amino acids included in the polymer chain could result in the creation of new types of active sites to solve the problem of carrying out enzymatic reactions under extreme conditions (extreme values of pH, temperature, salt content, etc.). The chemical modification of active sites by unnatural amino acids should result in the appearance of enzyme families with altered catalytic efficiencies and mechanisms and characterized by transformed specificity and enantioselectivity.

The development of DNA technologies introduces the task of biocatalyst design de novo. Based on a continuation of methods already traditionally used, such as DNA-shuffling and directed evolution, it is possible to anticipate the appearance of methods that allow one to obtain enzymes having a new structural basis and newly selected properties. In this view, it is interesting to minimize the protein structure and to create proteins from the limiting number of essential amino acids. From one point of view, the approaches developed should guarantee a deep understanding of the physiological basis of biological phenomena and from another point of view should lead to new biocatalyst families of high practical significance.

4.3 Biocatalysis: The foundation for new resource technologies

The limitations of traditional sources used in the chemical industry, as well as limitations of hydrocarbon energy sources, 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. 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). Considerable progress is also expected in the application of enzymes as catalysts in electrochemical reactions. Biofuel cells are considered as potential energy sources for industry and transport as well as for modern medicine. Biofuel cells are based on the

enzymes that are able to catalyze reaction on electrodes. This phenomenon is called bioelectrocatalysis. The direct electron exchange between the enzyme active site and the electrode is possible. Enzyme electrodes are able to oxidize a variety of substrates and to reduce oxygen directly to water. Chemical energy of fuel combustion is converted directly into electricity with 80–90 % efficacy. The development of fuel cells requires engineering of the enzymes to improve their stability as well as to provide efficient electron transport between the electrode and the enzyme active sites. The role of enzymes is growing in the development of new preparation methods of polymer materials, including new polymers on the base of peptides, polysaccharides, polyethers, etc.

5. PROTEINS WITH UNNATURAL AND RARE AMINO ACIDS

5.1 Gene-expressed proteins containing unnatural amino acids

All proteins are known to be made up of 20 amino acids and some of their biotic derivatives. It is an intriguing question whether the proteins and enzymes could harbor unnatural amino acid analogs and modify their activity. The interest in proteins containing unnatural amino acids is stimulated by at least two aspects:

- The potential possibility of incorporation of unnatural amino acids into proteins infinitely widens the spectrum of novel proteins to be synthesized. The development of effective methods to synthesize proteins and enzymes using synthetic amino acid analogs containing heteroatoms gives one the chance to obtain proteins and enzymes with previously unknown properties.
- Novel classes of drugs can be proposed using unnatural amino acid analogs incorporated into proteins. It is known that possible incorporation of some amino acids in polymer chains depends on the specificity of aminoacyl-tRNA synthetases. Enzymes of this group having high specificity to both amino acid and tRNA provide the correspondence between codon and amino acid. The result is a limitation on the incorporation of amino acids having nonclassic structures into a polymer chain.

At present, several approaches are developed to overcome these limitations:

- Aminoacyl-tRNA synthetases as well as most enzymes have no absolute specificity and tolerate the incorporation of analogs having a structure close to natural amino acids into the compound with t-RNA. The incorporation of several fluorine-containing amino acids into the proteins is possible and can be done as follows:
- Aminoacyl-tRNA with analogs of amino acids could be obtained chemically (precharged RNA) and incorporated into proteins in systems of cell-free protein synthesis;
- Considerable experimental development is linked with the methodology based on the conversion of a codon corresponding to a certain amino acid into the “blank, nonsense codon” (nonsense codon suppression, amber codon). This is followed by chemical amino acylation of amber codon-suppressing aminoacyl-tRNA synthetase by the unnatural amino acid and expression of the gene of the required protein in vitro by the transcription-translation biosynthesis machinery. This method is useful enough, and it allows one to obtain 10–50 mg/ml of modified protein.

However, there is another approach that is more universal and potentially more widely applicable. It based on modification of the specificity of aminoacyl-tRNA synthetases. The possibility of changing enzyme specificity by the methods of genetic engineering and site-directed mutagenesis is well known, presupposing the replacement of amino acids composing the selected sorbtion site of the enzyme active center.

Advances in the field of site-directed mutagenesis of aminoacyl-tRNA synthetases should result in the creation of a universal method of preparation of proteins composed of unnatural amino acids.

5.2 Proteins and enzymes containing trace elements

Previously, proteins that contained trace elements were identified by biochemical methods, and subsequently a variety of biophysical techniques were used for their characterization. However, these methods are labor-intensive and require relatively high quantities of proteins. In addition, metal-containing proteins with low abundance or limited expression patterns escaped identification using classical methods.

The availability of over 500 completely sequenced genomic sequences, including those of humans, as well as dramatic development in the methods of bioinformatics provides researchers for the first time with the ability to identify full sets of micronutrient-associated proteins in organisms. In addition, methods have been developed that allow functional characterization and assessment of biomedical potential of various trace element-containing proteins.

Several studies were recently published that described identification of full sets of trace element-containing proteins that are encoded in a genome of interest. Two independent bioinformatics methods were used to identify all, or almost all, genes encoding selenocysteine-containing proteins in human, mouse, and *Drosophila* genomes, providing a first view on selenoproteomes in these organisms. Selenium-containing proteins contain selenium in the form of selenocysteine, which has been called the 21st amino acid. Until last year when the 22nd amino acid pyrrolysine was discovered, the discovery of selenocysteine, encoded by the UGA codon, has been the only addition to the universal genetic code since this code was discovered in 1960s. Selenocysteine is inserted into polypeptide chains during ribosome-based protein synthesis when an RNA structure, designated a selenocysteine insertion sequence (SECIS) element, is present in the selenoprotein genes.

The principal method for identification of selenoprotein genes included the search for SECIS elements. In addition, to identify mammalian selenoproteins, human and mouse (or human and rat) genomes were searched in parallel to identity pairs of orthologous SECIS elements. It was found that the human genome has 25 selenoprotein genes and the mouse genome 24, whereas the fruit fly has only 3 such genes.

A second method for selenoprotein identification included the searches for selenoprotein/cysteine-containing protein pairs of homologs. This method provided independent verification of the number of selenoproteins in various genomes. Subsequent characterization of new selenoproteins revealed examples of proteins with expression patterns limited to embryos or testes, as well as proteins with novel subcellular distributions. This example illustrates the power of bioinformatics in linking genomics and chemistry. Trace elements are typically present at critical sites within protein structures. Information on the presence and location of a trace element in a protein may allow application of additional methods that assess protein reaction mechanism and biological function. Understanding of the identities and functions of trace element-containing proteins will also provide new tools for nonspecific incorporation of these elements into protein. Several proteins were already designed to coordinate metals, which modified functional repertoires of these proteins. Moreover, understanding of selenocysteine insertion may reveal ways for targeted insertion of this residue into protein. This will provide improvements over nonspecific insertion of selenium-containing residues, such as selenomethionine, since the number of selenium atoms in a protein and their specific locations can be precisely controlled.

Although biologically relevant trace elements, other than selenium, are not known to be inserted into protein co-translationally, bioinformatics methods for their identification will likely be developed. For example, it may be possible to search for dicysteine-containing patterns of amino acids present within the context of secondary structure patterns that coordinate zinc, iron, and copper. Identification and characterization of full sets of metal-containing proteins in humans and other organisms should set up new challenges and highlight the need for genomic and postgenomic chemistry of biological trace elements.

6. LIPIDOMICS

From the chemical point of view, the youthful era of genomics and postgenomics focuses on efficient analyses of macromolecules (i.e., DNAs, RNAs, and proteins). A successful marriage of chemistry with high technology enabled an enormous collection of data. The knowledge thus gained on genomes, transcriptomes, and proteomes provided new leads for the design of experiments to understand gene function. This is not the case in the domain of low-molecular-mass molecules. These molecules are not only building blocks for macromolecules such as those described above, but also substrates, intermediates, and products in cellular and tissue metabolism, and—even more important—regulators of metabolism and cellular actions. In contrast to all other classes of biomolecules, lipids are not defined in accordance with common structural features. A multitude of structures is known to date, whose biosynthetic origin is mainly acetogenic or isoprenoid. The main classes are the "neutral" lipids, such as long-chain acylglycerols, fatty acids, and their oxygenated derivatives, the group of "complex" lipids such as phospho-, sphingo-, and glycolipids, and the plethora of steroids and derivatives. The properties of hydrophobicity and amphiphilicity result in an enormous power for self-organization that is mainly entropy-driven, linking up with carbohydrates or complexing with proteins enhances their information potential. All these properties make lipids indispensable for cellular life, be it for membrane formation, for serving as energy sources, or for regulator and signaling functions which include gene regulation by lipids.

If we call the entire spectrum of lipids in a biological system the "lipidome", then mapping this spectrum can be called "lipidomics", which now will furnish new leads to provide insight into the function of a single lipid molecular species. The next step would be to analyze and study lipids in the context of carbohydrates, proteins, and even nucleic acids, and finally in the context of cellular and tissue biology and physiology. This will create leads for the synthesis of xenobiotic lipids whose intended action, in turn, can be studied by subjecting them to the lipidomic cycle of increasing complexity just outlined.

The tasks for lipidomics in postgenomic chemistry can be defined by the following:

- *New analytical approaches for mapping the lipidome: Global characterization of lipid molecular species by HPLC/MS.* There is a need to improve efficiency of this method on the one hand, and to develop further methods for simultaneous analysis of different lipid classes by chip-type technology on the other hand.
- *Identifying the lipid network, including lipid mediators, for metabolic and gene regulation and its integration with non-lipid signaling*. The challenge here is the integration of lipidomics with proteomics to study and characterize homeostatic and aberrant cellular states.
- *Design and application of xenobiotic lipids for interaction with nucleic acids and nucleic acid/protein complexes*. Lipidomics allow one to investigate how lipids interact with nucleic acids in a cell-based approach. The perspectives are answers with respect to regulation at nucleic acid levels as well as leads for intervention with drugs.

7. FOOD CHEMISTRY

7.1 Chemistry of tastes and smells: Postgenomic analysis of chemical signalization and sensory perception

Chemical signalling is one of the main mechanisms of metabolic regulation, of communication between cells and organs of the body, and of sensory perception of the surrounding world by living organisms. It is also fundamental for the functioning of various nervous systems in living animal species.

Despite the utmost importance of chemical signals for the correct functioning of the living organism and for communication between cells and organisms, the identity, specificity, and scope of activities of the majority of compounds that play crucial roles in chemical signalling are either poorly known or simply unknown. This is already the case for several primary and secondary neurotransmitters

such as acetylcholine, serotonin, adrenaline, and others. Only recently has the role in chemical communication in the body of so simple a chemical as NO been determined. The situation is even more underdeveloped in the case of sensory perception of tastes and smells. Most of these chemical signals are transduced by systems of G-protein coupled receptors (GPCRs)—heptahelical transmembrane proteins homologous to rhodopsin. The importance of these events can be also perceived from the fact that the family of GPCR genes is one of the most numerous in the human genome.

Recent advances in genomics and in proteomics of the sensory systems allow screening for many still unknown sensory ligands that convey attractive and repulsive stimuli to the living organism. More detailed and creative research of the agonist and antagonist of sweet/bitter taste chemical signals (chemical compounds) would allow much better control of the attractions and repulsions induced by certain foods and the treatment of many alimentary pathologies. Postgenomic study of the structure/function effects of well-constructed libraries of volatile compounds would also allow identification of smell-driven attractions and repulsions, so important in many crucial elements of animal and human behaviors, with well-selected or -designed odorants. More complete knowledge of this area of neurochemistry would certainly increase the existing means of treating many problems in sensory pathology, and would consequently improve human well-being in domains of sensory perception, which is so important for physical and psychological health.

7.2 Postgenomic food chemistry and toxicology

Improved methods of genomics and proteomics should provide access to emerging elements of food chemistry. Soon it will be possible to monitor and analyze positive and negative interactions among the variety of chemical compounds present in foods and in our genomes and proteomes. Hence, with regard for human or animal health, more precise definition of beneficial or detrimental chemical entities present in foods and fodders is already possible. Consequently, our ability or possibilities to modify the composition and processing of different foods will be increased and made more rational and target-oriented. Different alimentary compositions will be advisable for different population groups (elders, children, pregnant women, and sick people). It will also be possible to link the nutrient choices with the genome of a given individual. Many toxic elements or modifications will be identified and made innocuous thanks to reworking of the physical chemistry of food formulation and preparation, thus avoiding the negative influences of xenobiotics and other harmful chemicals present in our foods. For example, information on the specificity of interactions of polyvalent cations with proteins composing the human proteome is still very scarce, whereas the assessment of more complex interactions involving the proteome is even more challenging (e.g., rational prevention of toxicity of recently used—and widely spread in nature—alloys, provoking the so-called "Gulf syndrome").

Many military, industrial, and other pollutants are entering the food chain, and antidotes in the form of small chemical compounds should become common due to the concerted research in combinatorial chemistry, applied genomics, and proteomics. Hence, the combined approaches of postgenomic biology and chemistry will stimulate research in food chemistry and toxicology to improve the quality and safety of our foods and, as a result, of the quality of human, animal, or plant life.

Information on the benefits and risks associated with consumption of certain dietary chemical compounds (medicines, foods) may allow reprogramming of some of the key plant and animal species and result in the production of the GMOs, which would yield better starches and healthier (unsaturated) lipids that are easier to digest, less allergenic and susceptible to protein modifications, and better foodstuffs in general. These complex and beneficial chemical interventions will certainly contribute to improving the nutritional value and health impact of the foods of the third millennium. Postgenomic food chemistry may trigger one of the most important revolutions ever in human nutrition.

In this sense, the slogan "Better life through chemistry" is and can be still important for years to come; however, it could be revised to "Better life through postgenomic chemistry".

8. POSTGENOMIC MACROMOLECULAR CHEMISTRY

8.1 Recombinant polymers

Materials with excellent performance and novel functions are the basis for future technology. 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.

Nature's fine-tuned control of macromolecular structure far surpasses that which can be achieved in chemical polymerization processes. Many biologically derived materials spontaneously organize into noncovalently bound, complex structures. The challenge is to identify specific technological and commercial opportunities that demand highly engineered molecular materials. Once the performance requirements for such applications have been defined, the power of the biomolecular approach can be brought to bear on the problems of materials design, synthesis, and fabrication.

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

The production of artificial proteins by recombinant methods already has been demonstrated in several laboratories. Some groups have succeeded in the development of artificial proteins with smart properties. The behavior of molecular machines based on similar artificial proteins produced by chemical synthesis was extensively studied. Electro-active polymers already 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 practically any desirable amino acid sequence and the recent breakthrough in industrial microbial cell fermentation and downstream processing of proteins allows production of recombinant polymers. Precise control of the composition and structure of recombinant polymers endows them with unique properties. The introduction of unnatural amino acids in the structure of recombinant polymers will give them unique properties not present 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 through utilization of the conformational transitions of polymers rather then movement of solid machinery.

8.2 Template-directed synthesis of polymers: Self-proliferating polymers

Biomacromolecules such as polynucleotides, polysaccharides, and proteins are essential to organism survival. Their synthesis generally involves in vivo enzyme-catalyzed chain growth polymerization of monomers. In many cases, natural catalysts carry out polymer syntheses that are impossible to accomplish using conventional chemistry. Because of this, many chemical laboratories now use enzymes in synthesis of organic compounds. The advantage of enzyme-catalyzed synthesis is that such reactions proceed under environmentally friendly conditions.

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

Template-guided synthesis in the presence of enzymes has been demonstrated.

It is possible to expect the development of this approach in several directions:

- creation and investigation of chemical models of DNA polymerase; investigation of conditions for DNA proliferation in systems in vitro with the application of chemical analogs of DNA co-polymers;
- widening of possible applications of enzymes (polymerases, oxidoreductases, esterases) for template-directed matrix synthesis (polyaniline, polyethers, etc.);
- investigation of possible catalytic synthesis of polymers with proliferation of polymer matrix; and
- chemical models of polymerase chain reaction.

Investigations in this field should result in the creation of self-proliferating polymers. New systems that record information in chemical language different from classical biological language could be suggested on the basis of a number of such polymers, and could lead to the construction of new forms of life.

9. BIOANALYTICAL CHEMISTRY

A number of new analytical technologies that qualitatively changed the face of modern bioanalysis have been developed during the realization of the Genome project and its further applications. The methods of DNA amplification catalyzed by DNA-polymerases have been widely disseminated. Considerable progress was achieved in the creation of biochip analytical devices and biosensors. The development of proteomics provides an application of mass spectrometric methods for the analysis of proteins and peptides.

Advances in genomic and nano/materials/info technologies are being combined to provide analytical devices and systems with potential global effects on individual and public health, safety, economic, social and political systems; and business and commerce.

9.1 Protein and small-molecule microarrays

DNA microarrays have revolutionized the way nucleic acids are characterized. One can easily foresee that protein and small-molecule microarrays will have a major impact in many areas, such as protein structure–function and protein–protein interaction studies, as well as in lead compound identification. Toward such practical microarrays, novel chemistries that ensure selective, reliable, and efficient linkages of synthetic and biological molecules onto solid supports 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.

Thus, DNA analysis machines and chip-based systems will likely accelerate the proliferation of genetic analysis capabilities, improve drug searches, and make biological sensors practicable. Such new bioanalytical systems (some genetically engineered) may also aid in detecting biological warfare threats, improving food and water quality testing, continuous health monitoring, and medical laboratory analyses. Such capabilities could fundamentally change the way health services are organized by greatly improving disease diagnosis, understanding predispositions, and improving monitoring capabilities.

9.2 New nanoanalytical systems

Recent techniques such as functional brain imaging and knock-out animals are revolutionizing our endeavors to understand human and animal intelligence and capabilities. These efforts should make significant inroads into improving our understanding of phenomena such as false memories, attention, recognition, and information processing, with implications for better understanding of humans and designing and interfacing artificial systems such as autonomous robots and information systems.

Neuromorphic engineering (which bases its architecture and design principles on those of biological nervous systems) has already produced novel control algorithms, vision chips, head–eye systems, and biomimetic autonomous robots.

9.3 Single-molecule registration

General principles of DNA polymerase reactions and polymerase chain reaction (PCR) techniques initiated development of a number of atypical chemical processes, for example, production of conductive polymer films using oxidoreductase and polyelectrolytes (polyaniline, polypyrrole, polythiophene, etc. films). Such conductive films and polymers have wide potential application in sensors, biosensors, rechargeable batteries, molecular electronic devices, and electrochromic displays.

Another bioanalytical application connects with the detection of single molecules. A single-enzyme molecule can generate polymers and the formation of nanoparticles, which can act as markers of immuno- and nucleic acid-based reactions. Scanning probe microscopy can be used as a counter of such nanoparticles. This approach promises to be extremely specific for the registration of single-molecule interactions.

9.4 New recognition elements

Genetically modified organisms might be engineered to produce biopolymers (plastics) for engineering and analytical applications. Manufacturing of new recognition elements and other materials with DNA might represent the ultimate biomimetic-manufacturing scheme. It consists of "functionalizing small inorganic building blocks with DNA and then using the molecular recognition processes associated with DNA to guide the assembly of those particles or building blocks into extended structures". Using this approach, a highly selective and sensitive DNA-based chemical assay method using gold nanoparticles with attached DNA sequences has been demonstrated already. This approach is compatible with the commonly used PCR amplification method of the amount of the target substance. The DNA-based self-assembly mentioned above might be achieved by attaching nonlinking DNA strands to metal nanoparticles and adding a linking agent to form a DNA lattice. This can be turned into a biosensor or a nanolithography technique for biomolecules.

10. CONCLUSION

Recent outstanding achievements in genomics and proteomics illustrate the considerable potential within the scientific community to investigate the structure and functions of biomacromolecules and biosystems. The bulk of information currently available on protein-coding nucleotide sequences in genomes of different species is growing exponentially. Bioinformatics and chemoinformatics methods raise the challenge of identification of all molecular structures in biological systems. In turn, these developments may lead to new fields of chemistry. Development of new large-scale and high-throughput projects oriented toward generation and application of genomic information suggests a model for development of chemical projects on a similar scale. The use of chemistry in the creation of new biomimetic structures that incorporate unnatural amino acids or other compounds is of particular importance. The following directions are avenues for potential advances in postgenomic chemistry:

- combinatorial chemistry and automated chemical synthesis;
- synthesis of new classes of unnatural amino acids and development of new biosynthesis methods for preparation of proteins containing unnatural amino acids; studies on altering aminoacyl-tRNA-synthetase specificity by methods of site-directed mutagenesis and molecular evolution;
- chemical management of biosystems at the molecular level;
- new approaches for classification of enzymes on the basis of structures of enzymatic active sites;

- creation of new polymer catalysts on the basis of enzyme action principles;
- self-multiplying polymers: chemical models of DNA-polymerases; and
- new methods of analytical chemistry based on micro- and nanochip technologies.

Genomic and proteomic studies can significantly influence education in chemistry. At present, instruction in molecular biology, genetic engineering, genomics, and proteomics in chemistry departments at most universities is not satisfactory and should rise to the postgenomic challenge. The development of new courses such as "Chemical basis of genomic studies", "Genes and genomes for chemists", and "Bio- and chemoinformatics" are essential to raise the attractiveness of chemistry as a field of study and to accelerate the development of postgenomic chemistry.