

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

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

<sup>‡</sup>Corresponding author

---

Republication or reproduction of this report or its storage and/or dissemination by electronic means is permitted without the need for formal IUPAC permission on condition that an acknowledgment, with full reference to the source, along with use of the copyright symbol ©, the name IUPAC, and the year of publication, are prominently visible. Publication of a translation into another language is subject to the additional condition of prior approval from the relevant IUPAC National Adhering Organization.

# Postgenomic chemistry

## (IUPAC Technical Report)

*Abstract:* Numerous areas of chemistry can benefit from the ongoing genomic revolution. Here, we discuss and highlight trends in chemistry in the postgenomic era. The areas of interest include combinatorial approaches in organic chemistry; design and analysis of proteins containing unnatural amino acids; trace element-containing proteins; design and characterization of new enzyme types; applications of postgenomic chemistry in drug design; identification of lipid networks and global characterization of lipid molecular species; development of recombinant and self-proliferating polymers; and applications in food chemistry and bioanalytical chemistry based on new nanoanalytical systems and novel recognition elements.

*Keywords:* Postgenomic; combinatorial; bioanalytical; Division III; Division VII.

### 1. INTRODUCTION

Systematic investigation of DNA structures and the decoding of genomes of various microorganisms, plants, and humans established the basis for a quantitative leap in modern natural science. The decoding of genomes has a number of significant consequences which stimulated notable changes in the development of many fields, including:

- The possibility of accumulating structural information about large polymer molecules, generated according to the principle “want to know everything”, was demonstrated for the first time. This information includes the primary structure of all enzymes, all structural proteins, and all protein regulators in specific groups of organisms. That is, the demonstration of an up-to-the-minute scientific advance was reached by attempts to understand the structures and functions of all synthesized proteins (proteomics), lipids (lipidomics), and other biomolecules.
- The accumulation of a large amount of structural information resulted in the need for development of adequate methods of storage, treatment, and analysis of the data. The main outcome was the emergence of bioinformatics, a new life science-based informational technology, directed toward the development and application of computational methods of analysis and interpretation of genetic texts.
- Projects directed toward genome analysis at the molecular level demonstrated the potential for integrated analytical methods. This development has resulted in the industry of chemical structural analysis. Automated analytical technologies have provided extraordinary sensitivity and capacity of analytical methods and procedures, which allowed the implementation of this grandiose aim.

Genomic studies and research directed toward the analysis of all proteins in biosystems influence the methodology of modern chemistry, since this is an area of interest and is a part of IUPAC's mission and agenda. It is important to discuss the direction in which the various fields of chemistry are headed, informed by our knowledge of genome structure and the application of methodology and genomics. Obviously, the achievements of genomics and proteomics could considerably influence the outcomes of classical chemical science as well as development of new scientific directions and open new horizons in chemistry. To discuss these issues, a working group was formed under IUPAC's project “Postgenomic Chemistry”.

The main outcomes of the project were as following:

- More than 20 scientists from 11 countries (Belgium, Canada, Estonia, France, Germany, Italy, Poland, Russia, Sweden, UK, and USA) actively participated in the interdisciplinary project, which analyzed and discussed the most promising areas of bioorganic chemistry in light of information provided by recent advances in functional genomics.
- A workshop was held in Moscow (6–8 September 2003) that was an important step in the project, since it allowed experts to exchange views on chemistry in the postgenomic era and to discuss the implication of advances in genomics, proteomics, biomimetics, and biological and chemical informatics. Participants agreed that there is a need to create new educational programs and training courses for chemistry students in the area of chemical genomics.

The dissemination of these ideas was through lectures of experts involved in the discussion at various scientific forums. The goals and summary of discussion were presented at the XVII International Mendeleev Congress on General and Applied Chemistry (Kazan, Russia, 21–26 September 2003) [1].

This report is based on the opinions of experts working in various fields of biomolecular chemistry. The following discussion should be helpful in developing strategic directions for future developments of 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 more 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 have been overexpressed during a disease state have been identified. To understand the role of these proteins and eventually to regulate their activity, it is of the utmost importance to discover small molecules that will interact with them selectively. The ultimate goal is to have one low-molecular-weight compound to block or selectively activate each 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 high production methods, several thousands of analogs of pure small molecules of tremendous structural diversity. Therefore, supported and solution-phase synthetic methods and automated purification techniques are needed to complement the tools available to chemists and provide access to novel ligands with unique molecular scaffolds.

### **2.2 Artificial proteins**

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

Although artificial proteins can be prepared from scratch, the incorporation of manufactured 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 use of the multiwavelength anomalous dispersion (MAD) technique. Incorporation is done by biotechnology using selenomethionine. It is necessary to develop additional methods permitting incorporation of heavy atoms other than selenium. Manufactured amino acids can also serve to enhance the structural stability of protein biocatalysts, therefore prolonging their lifetime, and improving environmentally friendly industrial processes. Although elegant approaches have been developed for incorporating nonproteogenic amino acids through biotechnology or chemical ligation strategy, development of improved complementary techniques to these methods is necessary to expand the range of manufactured amino acids that can be incorporated and to augment the diversity of potential artificial proteins that can be created.

Describing all of the areas where organic chemistry will have an impact in the postgenomics era would require a complete issue of this journal. The above examples represent areas where the authors foresee direct and immediate impacts of organic chemistry in this field.

### **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 the molecular recognition phenomena is, therefore, an important element in the structural understanding of these binding events, which are prerequisites for controlled interference with biological functions mediated by respective protein–ligand interactions. 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 or not 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 generation of a large range of different molecules, including simple and more complex peptides, which may contain much more than 20 proteinogenic amino acids, peptidomimetics, and 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 and 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 interference with protein–ligand interactions is aimed at molecules which, owing to their specific molecular architecture, are capable of mimicking 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 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 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 the 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 many new possibilities for comparative analysis of the different biological taxonomical groups. Such analysis is directed toward deeper understanding of structural–functional and evolutionary relationships for biological objects at all levels. Genetic information encoded in the genome represents 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 along with 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 and the relatively newer 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 a structure–function relationship study, owing to the important role of posttranslational modification of proteins and their interactions with lipids, carbohydrates, and small endogenous bioregulators. As a result, structural research on glyco-, lipo- and nucleoproteins 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 the recognized structure–function relationships of 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 for new lead compounds for novel medicines can be performed in the following way:

- analysis of the 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;
- searching 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 the 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 21<sup>st</sup> century many problems of chemistry, biology, and medicine that were traditionally decoded by experimental methods will be initially modeled “in silico” by methods of bioinformatics, with subsequent experimental checking of computer estimates. Such an approach will create the basis for discovery of new, effective, and safer pharmaceuticals while significantly decreasing the time and costs involved in discovery, and will 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 its 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 the problems of chemical transformation of renewable raw materials such as carbohydrates, biomass components, and carbon dioxide. Various biocatalytic systems are expected to be used in the 21<sup>st</sup> century as the main instruments for the chemical modification of molecules.

The colossal volume of information disclosing molecular structural regularities and peculiarities of various enzymes is now collected owing to the recent development and wide application of physico-chemical methods (X-ray and NMR analysis, mass spectrometry, etc.). The up-to-date conception of protein structure and features correlation explains the catalytic mechanisms of various chemical reactions and the existence of enzymatic specificity from the fundamental basis of physics and chemistry. Currently, biocatalytic systems based on application of enzymes are widely used in various fields of science (organic and analytical chemistry, genetics, molecular biology, etc.), biotechnological processes, environment, and medicine.

The dilative range of reactions applying the enzymatic catalysts requires improvement of their selected characteristics such as pH or thermal stability, substrate specificity, enantioselectivity, etc. Genetic modifications of protein molecules and manipulations with genes encoding proteins (site-directed mutagenesis and directed evolution) are approaches commonly used to overcome the drawbacks of native enzymes. Further development of these methods may enable elaboration of families of new postgenomic biocatalysts with features unknown in nature. The recent prognosis is that the total world volume of enzyme production, taken in real-value terms, will become equal to the production of “classic” catalysts used in the chemical industry within the next decade.

Biocatalysis is fundamental for new resource technologies. The limitations of traditional sources used in the chemical industry as well as the limitations of hydrocarbon energy sources acutely raises questions about development of new sources of raw materials and energy. Enhancing interest in biocatalytic conversion of biomass into chemical raw materials for creation of new materials, polymers,

and energy sources (namely, hydrogen, methane, ethanol, and diesel oil) will characterize the next few decades. Significant progress is 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 is also 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, and polyethers. The availability of a large number of complete genomic sequences allows the identification of homologs of previously characterized enzymes. Phylogenetic analyses of these sequences may yield clusters of orthologous enzyme groups and reveal new enzyme families. Data obtained demonstrates that the practically unlimited variability of the primary structure of enzymes can result in a very confined number of types of catalytic centers providing effective transformation of molecules.

#### **4.2 Biocatalysis: New-generation enzymes**

Use of bioinformatics for investigating protein genetics enables identification of known enzymes in new microbial sources. The accomplishment of comparative analysis of genomes encoding the same proteins originated from mesophylic and thermophilic microorganisms allows the genetic improvement of enzymatic thermostability. This approach to obtaining novel enzymes with enhanced thermal stability and high activity creates the basis for development of novel biocatalytic systems and high-temperature biotechnological processes.

Development of new preparation methods of gene-expressed proteins obtained from manufactured amino acids is capable of supporting the creation of new catalysts, thus considerably widening the possibilities of biocatalysis. The biosynthesis of enzymes with manufactured 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.). Chemical modification of active sites by manufactured 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 new DNA technologies and the accumulation of information about the structures of enzymatic active sites promotes the elaboration of catalytically active substances imitating enzymes with selected features. The genetic construction of catalysts is anticipated to be carried out with the application of a minimal number of amino acids that presumably are necessary for the formation of enzyme-like molecular structures. Advances of chemistry in this direction should guarantee the appearance of a new generation of reproducible artificial enzymes.

#### **4.3 Biocatalysis: The foundation for new resource technologies**

Symptoms of the increasing world crisis in fuel and energy preclude the search and development of new energy and raw material sources. In the next decade, biocatalytic conversion of renewable biomass as a process of providing the chemistry for new raw materials useful in the production of polymers, solvents, fuels, etc. is expected to become the basis for progress in chemical technology. Significant progress may be expected in the development of biosystems catalyzing the reduction of carbon dioxide to 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 potential energy sources for industry and transport as well as modern medicine. Biofuel cells are based on enzymes with the ability to catalyze reactions on electrodes. This phenomenon is called bioelectrocatalysis. 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. The chemical energy of fuel combustion is converted directly

into electricity with 80–90 % efficacy. Development of fuel cells requires engineering of the enzymes to improve their stability and 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 artificial amino acid analogs and modify their activity. The interest in proteins containing artificial amino acids is stimulated by at least two aspects:

- The potential possibility of incorporation of artificial amino acids into proteins infinitely widens the spectrum of novel proteins that can be synthesized. Development of effective methods to synthesize proteins and enzymes using synthetic amino acid analogs containing heteroatoms gives one the possibility to prepare proteins and enzymes with previously unknown properties.
- Novel classes of drugs can be proposed using artificial amino acid analogs incorporated into proteins. It is known that the possible incorporation of some amino acids into polymer chains depends on the specificity of aminoacyl-tRNA synthetases. Enzymes in this group that have 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 the present time, several approaches have been developed to overcome these limitations:

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

However, there is another approach that is more universal and is potentially applicable on a wider basis. It is based upon modification of the specificity of aminoacyl-tRNA synthetases. The possibility of changing enzyme specificity by methods of genetic engineering and site-directed mutagenesis is well known and presupposes the replacement of those amino acids composing the selected sorption 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 for the 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 methods of biochemistry, and subsequently a variety of biophysical techniques were used for their characterization. However, these meth-



ods are labor-intensive and require relatively large quantities of proteins. In addition, metal-containing proteins with low abundance or limited expression patterns escaped identification when using classical methods.

The availability of over 500 completely and partially sequenced genomic sequences, including those of humans, and the rapid development in methods of bioinformatics provides researchers 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 the 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 21<sup>st</sup> amino acid. Until last year when the 22<sup>nd</sup> amino acid pyrrolysine was discovered, the discovery of selenocysteine, encoded by the UGA codon, had been the only addition to the universal genetic code since its discovery. Selenocysteine is inserted into polypeptide chains during ribosome-based protein synthesis when an RNA structure, designated as the selenocysteine insertion sequence (SECIS) element, is present in the selenoprotein genes.

The principal method for the identification of selenoprotein genes included the search for 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 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 method 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 the application of additional methods assessing the protein reaction mechanism and biological function. Understanding 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 the 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 cotranslationally, 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 func-

tion. 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 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 signaling and sensory perception

Chemical signalling is one of the main mechanisms of metabolic regulation, communication between cells and organs of the body, and 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 correct functioning of living organisms and 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 not known. This is 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 a simple chemical, NO, been determined. The situation is even more underdeveloped in the 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 proteomics of sensory systems allow screening for many of the 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 the sweet/bitter taste chemical signals (chemical compounds) would allow 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 the identification of smell-driven attractions and repulsions, critically 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 increase the existing means of treating many problems in sensory pathology, and consequently would improve human well-being in the domain of sensory perception, highly 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. It will soon be possible to monitor and analyze positive and negative interactions among the variety of chemical compounds present in foods, in our genomes and proteomes. Hence, in regard to human or animal health, more precise definition of beneficial or detrimental chemical entities present in foods and fodders is already possible. Consequently, our ability and 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 (i.e., the elderly, young children, pregnant women, and critically ill). It will also be possible to link nutrient choices with the genome of a given individual. Many toxic elements or modifications will be identified and made innocuous thanks to the reworking of the physical chemistry of food formulation and preparation, thus avoiding negative influences of xenobiotics and other harmful chemicals present in our foods. As an example, information on the specificity of interactions of polyvalent cations with proteins composing the human proteome is very scarce and the assessment of the more complex interactions involving the proteome is even more limited (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 enter the food chain, and antidotes in the form of small chemical compounds should become common due to intensive research efforts 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, the quality of human, animal, or plant life.

Information on benefits and risks associated with the 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 GMOs, which would yield better starches and healthier (unsaturated) lipids that are easier to digest, less allergenic, and susceptible to protein modifications, and better food-stuffs in general. These complex and beneficial chemical interventions will certainly contribute to improving the nutritional value and health impact of foods in 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 still be important for years to come; however, it could be revised to “Better life through postgenomic chemistry”.

## 8. POSTGENOMIC MACROMOLECULAR CHEMISTRY

### 8.1 Recombinant polymers

Biomolecular macromolecular chemistry for materials production is oriented to the so-called intelligent smart materials. 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 methodology of recombinant polymer synthesis is based on the construction of an artificial polymer gene in microbial cells. Various polyamide polymers with predicted properties can be obtained using this approach. Recently, the possibility of natural silk synthesis by recombinant methods was demonstrated. The introduction of non-natural amino acids in the structure of recombinant polymers will give them unique properties not present in natural polymers.

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

Biomacromolecules such as polynucleotides, polysaccharides, and proteins are essential to the survival of an organism. Their synthesis generally involves *in vivo* enzyme-catalyzed 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 the synthesis of polymers. The advantage of enzyme-catalyzed synthesis is that such reactions proceed under environmentally friendly conditions.

Nature demonstrates the unique example of polymer molecule multiplying by template-directed synthesis. This is the polymerase catalyzing the replication of DNA molecules. For chemistry, this is the example of production of the polymer identical to a template. The challenges in this area are connected with the possible application of enzymes (polymerases, oxidoreductases, esterases) for template-directed synthesis of macromolecules and with construction of 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

The success of the Genome Project was based on the development and application of new analytical methods. It stimulated the significant development of methods based on polymerase chain reaction (PCR) and DNA amplification. The notable progress in biochip technologies enabled the development of miniature devices for the investigation of characteristics of single molecules and molecular formations gave new opportunities for the investigation of nucleic acids and proteins.

### 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, including 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 a 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 PCR techniques initiated the development of a number of atypical chemical processes, for example, the 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 is connected to 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 for 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. Such a scheme 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 already been demonstrated. This approach is compatible with the commonly used PCR method of amplification 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 considerable potential for chemists 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 the generation and application of genomic information suggests a model for development of chemical projects on a similar scale. The use of chemistry in the development of new biomimetic structures that incorporate artificial 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 biosynthetic methods to prepare proteins containing these amino acids; studies on altering aminoacyl-tRNA-synthetase specificity by methods of 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 using the principles of enzyme catalysis;
- self-multiplying polymers: chemical models of DNA-polymerases; and
- new methods of analytical chemistry that are based on micro- and nanochip technologies.

Genomic and proteomic studies can significantly influence chemical education. At present, instruction in molecular biology, genetic engineering, genomics, and proteomics in chemistry departments at most universities is unsatisfactory and should rise to the postgenomic challenge. The development of new courses such as “Chemical basis of genomic studies”, “Genes and genomes for chemists”, “Bio- and chemoinformatics” are essential to increase the attractiveness of chemistry as a field of study and to accelerate the development of postgenomic chemistry.

## 11. REFERENCE

1. S. D. Varfolomeyev, T. K. Aliev, E. N. Efremenko. *Pure Appl. Chem.* **76**, 1781–1798 (2004).