

Letter from the Editor

Last months of the past year were marked by the growing interest of various government officials and parliamentarians, as well as the media, to the problems of the development of the biological information market and biotechnology. Even a simple review of dates and facts shows that our country's leaders decided to take a serious approach to the revival of Russia's technological potential in this area. Presidential Advisory Board for the Questions of Modernization and Technological Development held its meeting in Pokrov, where President D.A. Medvedev inspected the technological grounds for new medical compounds production. Prime Minister V.V. Putin visited the pharmaceutical company in Zelenograd, where medicines based on the recombinant proteins technology are produced. On October 15, 2009, debates in the State Duma of the Russian Federation were held, the topic of which was "a perfectibility of the legislative support for the biotechnological industries." Industrial Committee in the State Duma has initiated that meeting. Another session with the participation of the Ministry of Science and Education and the Ministry of Industry and Trade of the Russian Federation was held to discuss the science and technology issues, and finally, a Presidential message to the Federal Assembly of Russia contained a significant section on biopharmaceuticals.

It is a common knowledge that before Perestroika the Soviet Union held one of the leading positions in the world in the field of biotechnological industries. A lot was achieved due to fast development of both academic and practical "bioscience," which managed to bounce back rapidly during "post-Lysenko" development. Unfortunately, that period of positive growth in biology, like in many other areas of scientific and technological progress, was abruptly stopped by the revolutionary developments of the early 90s. Nowadays a lot of questions are being sharply posed about the revival of 'life sciences' in combination with new biotechnologies and the biopharmaceutical industry.

The concerns are connected both with the national security and the participation of our country in the international division of labor in the XXI century. Discussions are mainly focused

on accelerating the development of biological sciences and biotechnologies, as well as on methods of increasing research effectiveness to cope with international standards (publications in leading scientific journals, high citation index, intensive patenting). The resources for the development of the biotechnology – be it governmental corporations' or private companies' funds – are still undefined. Perhaps it is possible to use both sources in reasonable proportions, but the mechanisms and details of the investing process are vague, and its possible pace is unclear. Another great concern is the expertise of the projects. After political barriers have been taken down, Russian scientists had no formal restrictions to migration to the West, and the turmoil of the 90s led to mass emigration of scientists from Russia, thus hurting the biotechnological sphere severely. We have to wonder now if we could aim towards "re-immigration," in the way China had focused its efforts on the return of its former citizens to raise their industry?

Along with the global challenges facing our leaders, scientists and businessmen, there are some other problems that need immediate attention. Many of them are tied to the customs control of international transportation of biological compounds, including those that are necessary for the media cultivation for cell lines and animals. Nowadays the process of their delivery to the Russian Federation is rather complicated, and in certain cases (for example, chemicals kept at low temperatures), is practically impossible. Also, we have to note that laboratory instruments, materials, and technological equipment imported into the Russian Federation are substantially more expensive for the local consumers than for their colleagues in European countries or the U.S.. This all makes the growth of a biological industry in Russia significantly more difficult than in the West.

All the above said led the editorial board of this journal to dedicate its Forum section to the problems of Bio-farm. We selected several articles from authors in the science world, in business, as well as from the media: we sought to show the opinion spectra of various experts on the problems connected to the development of the biological industry in Russia.

Save 10% on Subscription for 2010

Details at www.actanaturae.ru

RESEARCH ARTICLES

Stacking approaches are further improved by implementing new algorithms of the evolutionary search and new scoring functions (we look to estimate the free energy of ligand binding). Some of them may include other components of molecular recognition force fields (2D or empirical terms, e.g. hydrophobic bonds described by their geometrical parameters) [4]. In this work we studied stacking interactions, which usually are not properly taken into account in widely used scoring functions.

THE PARAMETERS OF STACKING INTERACTIONS

Of all the various types of interactions in biomolecular complexes (such as hydrogen bonds, salt bridges, etc.) the stacking of aromatic substances deserves special attention. Most drugs include aromatic fragments in their chemical structure, and stacking often plays a crucial role in their recognition by protein chains. We have recently shown that an explicit account of stacking in scoring functions increases the efficiency of ADP docking [5]. The aromatic interactions were identified by the mutual rotation of two planes defined by geometrical parameters: the height h and displacement d of one ring relative to the other, and the angle between their planes (Fig. 1).

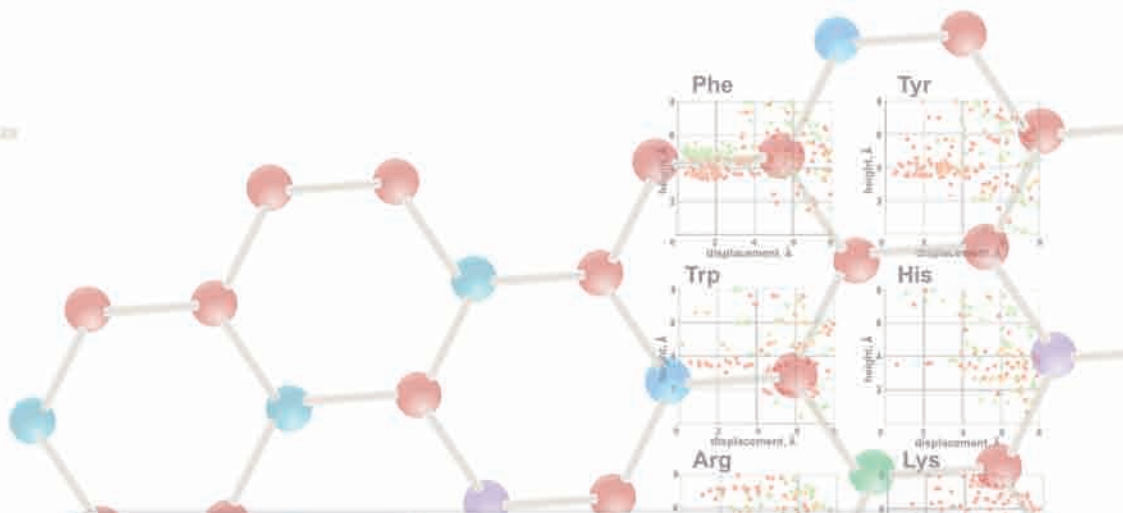
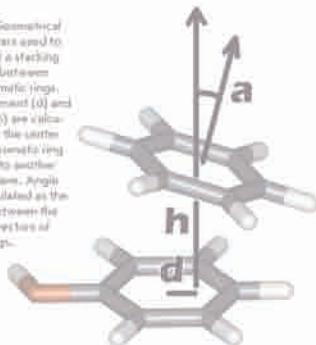
However, the range of these parameters, which corresponds to the presence or absence of a stacking contact, is not too very well defined and usually takes an arbitrary [6, 7]. Defining it more accurately would assist in describing more efficient scoring functions and should improve the prediction quality of the spatial structures of protein-ligand complexes by molecular docking methods. With this aim in view, we performed an analysis of the spatial structures of protein-ligand complexes determined experimentally with a same resolution which ligands contained aromatic or quaternary amine substituents.

The well-known example of stacking interactions is the parallel packing of guanine and pyrimidine nucleobases in DNA [8, 9]. Some aromatic compounds tend to orient perpendicular to each other (T-shaped stacking), as has been shown for amino acids in proteins [1, 10] and for model systems of carbon atom rings (benzene and naphthalene) [11–14]. Besides, such compounds participate in cation- π interactions, where a positively charged group interacts with the negatively charged orbital of aromatic nucleobases [15–17].

Taking all that into account, we analyzed the distribution of geometrical parameters h , d , and θ . The contacts of amino acids with guanine residues of ligands with the assistance with cluster of receptor amino-acids Phe, Tyr, Trp, and His, as well as with the positively charged quaternary group of Arg and amino group of Lys. The results obtained for guanine are presented in Fig. 2.

It can be seen that two distinct orientations are typical for Phe: parallel and perpendicular to the guanine plane (Fig. 2, orange and green, respectively). The displacement d lies in the same range 2–5 Å for both types of contacts. Meanwhile, they clearly differ in the value of height h , which is ≈ 3.3 Å for parallel d and ≈ 3.1 Å for perpendicular orientation. Similar distributions were obtained for Tyr, Trp, and His, though the data are noisier in them (not shown). The T-shaped contact is not as typical for Tyr, Trp, and His as it is for Phe.

Fig. 1. Geometrical parameters used to describe a stacking contact between two aromatic rings. Displacement (d) and height (h) are calculated for the centers of one aromatic ring relative to another ring's plane. Angle θ is calculated as the angle between the normal vectors of both rings.



APRIL/JUNE 2009, No. 1

ActaNaturae

SYNTHETIC ANTIBODIES
FOR CLINICAL USE

REGULATING TELOMERASE IN ONCOGENESIS
R. 81

THE STRUCTURE OF THE MITOCHONDRIAL GENOME AS AN ACTIVATOR OF OPISTHORCHIASIS
R. 89

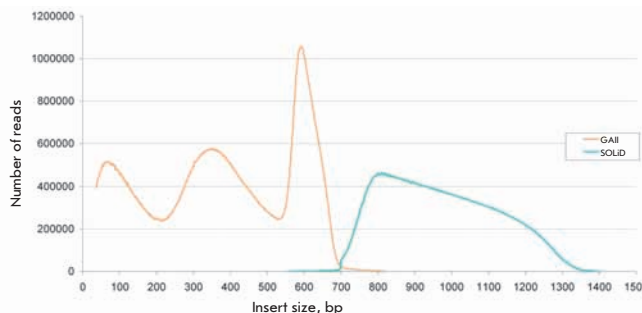
STACKING INTERACTIONS IN COMPLEXES OF FIBERS WITH ADENINE-AND GUANINE CONVENING MOLECULES

HIGHLIGHTS

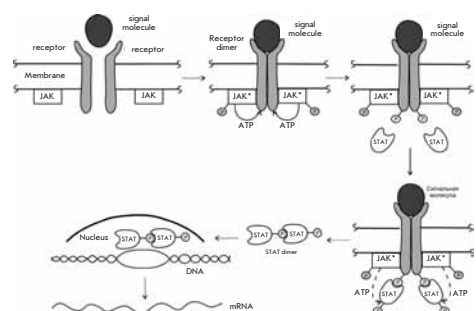
K.G. Skryabin, E.B. Prokhortchouk, A.M. Mazur, E.S. Boulygina, S.V. Tsygankova, A.V. Nedoluzhko, S.M. Rastorguev, V.B. Matveev, N.N. Chekanov, D.A. Goranskaya, A.B. Teslyuk, N.M. Gruzdeva, V.E. Velikhov, D.G. Zaridze, M.V. Kovalchuk

Combining Two Technologies for Full Genome Sequencing of Human

The study is the first Russian effort towards whole human genome sequencing. In this study, the whole genome sequencing of a human (Russian man) was performed using two technologies currently presented on the market - Sequencing by Oligonucleotide Ligation and Detection (SOLiDTM) and sequencing technologies of molecular clusters using fluorescently labeled precursors. Optimal algorithm of using the latest methods of sequencing was established for analysis of individual human genomes



Graph of distances distribution between pairs of reads for the human reference genome ver. hg18 libraries



A schematic representation of signal transduction via JAK-associated membrane receptors

D.G. Knorre, N.V. Kudryashova, T.S. Godovikova

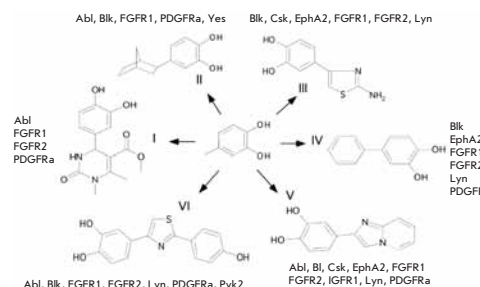
Chemical and Functional Aspects of Posttranslational Modification of Proteins

This paper reviews the chemical and functional aspects of the posttranslational modifications of proteins, which are achieved by the addition of various groups to the side chain of the amino acid residue backbone of proteins. Much attention is paid to the role of posttranslational modification of proteins in the regulation of biochemical processes in live organisms

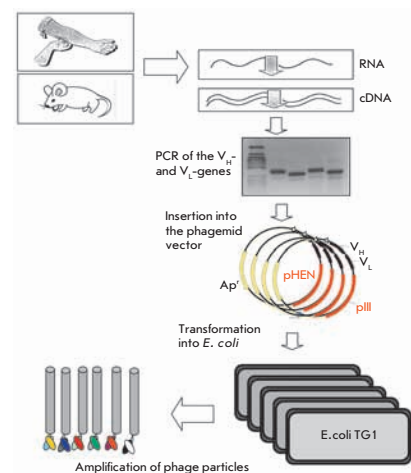
T.V. Rakitina, O.V. Yudkina, E.V. Smirnova, A.V. Lipkin

Protein Tyrosine Kinase Panel As a Tool for Anticancer Drug Design

This work is an example of a combination of inhibitor experimental search with the computational analysis of the potential mechanism of the inhibitors' action, which allowed to propose the 2-hydroxyphenol group as a scaffold for the design of new tyrosine kinase inhibitors.



The small molecule tyrosine kinase inhibitors that were found by screening of the chemical collection



N.V. Tikunova, V.V. Morozova

Phage Display on the Base of Filamentous Bacteriophages: Application for Recombinant Antibodies Selection

The display of peptides and proteins on the surface of filamentous bacteriophage is a powerful methodology for selection of peptides and protein domains, including antibodies. An advantage of this methodology is the direct physical link between the phenotype and the genotype, as an analyzed polypeptide and its encoding DNA fragment exist in one phage particle. This review is an introduction to phage display methodology. It presents recombinant antibodies display in more details

A schematic representation of the construction of an antibody fragment phage library

Founders

Russian Federation Agency
for Science and Innovation,
Lomonosov Moscow State University,
Park Media Ltd

Editorial Council

Chairman: A.I. Grigoriev
Editors-in-Chief: A.G. Gabibov, S.N. Kochetkov

V.V. Vlassov, P.G. Georgiev, M.P. Kirpichnikov,
A.A. Makarov, A.I. Miroshnikov, V.A. Tkachuk,
M.V. Ugryumov

Editorial Board

Managing Editor: V.D. Knorre
Publisher: A.I. Gordeyev

K.V. Anokhin (Moscow, Russia)
I. Bezprozvanny (Dallas, Texas, USA)
I.P. Bilenkina (Moscow, Russia)
M. Blackburn (Sheffield, England)
S.M. Deyev (Moscow, Russia)
V.M. Govorun (Moscow, Russia)
O.A. Dontsova (Moscow, Russia)
K. Drauz (Hanau-Wolfgang, Germany)
A. Friboulet (Paris, France)
M. Issagouliants (Stockholm, Sweden)
A.L. Konov (Moscow, Russia)
M. Lukic (Abu Dhabi, United Arab Emirates)
P. Masson (La Tronche, France)
K. Nierhaus (Berlin, Germany)
V.O. Popov (Moscow, Russia)
I.A. Tikhonovich (Moscow, Russia)
A. Tramontano (Davis, California, USA)
V.K. Svedas (Moscow, Russia)
J.-R. Wu (Shanghai, China)
N.K. Yankovsky (Moscow, Russia)
M. Zouali (Paris, France)

Project Head: R.R. Petrov

Editor: E.V. Dorogova

Science Editor: V.V. Sychev

Strategic Development Director: E.L. Pustovalova

Designer: K.K. Oparin

Photo Editor: I.A. Solovey

Art and Layout: K. Shnaider

Copy Chief: Daniel M. Medjo

Address: 119991 Moscow, Russia, Leninskiye Gory, Nauchny
Park MGU, vlad.1, stroeniye 75G.
Phone/Fax: +7 (495) 930 80 06
E-mail: knorre@d@gmail.com, rpetrov@strf.ru, vsychev@strf.ru

Reprinting is by permission only.

© ACTA NATURAE, 2009

Номер подписан в печать 18 декабря 2009 г.
Тираж 200 экз. Цена свободная.
Отпечатано в типографии «МЕДИА-ГРАНД»

CONTENTS

Letter from the Editor. 1

FORUM

A.I. Gordeev, Fund "Open Economics"
The Pharmaceutical Industry in Russia: Reality and Prospects 6

A.L. Konov, A.A. Leonov
Biopharma: How Can the "Valley of Death" between R&D and Innovation Be Overcome? 10

Recipe for Russian Insulin. 16

REVIEWS

N.V. Tikunova, V.V. Morozova
Phage Display on the Base of Filamentous Bacteriophages: Application for Recombinant Antibodies Selection 20

D.G. Knorre, N.V. Kudryashova, T.S. Godovikova
Chemical and Functional Aspects of Posttranslational Modification of Proteins 29

V.P. Puzyrev, M.B. Freidin
Genetic View on the Phenomenon of Combined Diseases in Man. 52

A.P. Grigorenko, S.A. Borinskaya, N.K. Yankovsky, E.I. Rogaev
Achievements and Peculiarities in Studies of Ancient DNA and DNA from Complicated Forensic Specimens 58

V.S. Baranov
Genome Paths A Way to Personalized and Predictive Medicine 70

RESEARCH ARTICLES

N.M. Novozhilova, N.V. Bovin
D-Arabinose Methabolism: Characterization of Bifunctional Arabinokinase / Pyrophosphorylase of *Leishmania major* 81

T.V. Rakitina, O.V. Yudkina, E.V. Smirnova, A.V. Lipkin
Protein Tyrosine Kinase Panel As a Tool for Anticancer Drug Design. 84

I.G. Shabalin, K.M. Polyakov, V.I. Tishkov, V.O. Popov
Atomic Resolution Crystal Structure of NAD⁺-Dependent Formate Dehydrogenase from Bacterium *Moraxella sp. C-1*. 89

I.V. Shapovalova, W.B.L. Alkema, O.V. Jamskova, E. de Vries, D.T. Guranda, D.B. Janssen, V.K. Švedas
Mutation of Residue β F71 of *Escherichia coli* Penicillin Acylase Results in Enhanced Enantioselectivity and Improved Catalytic Properties 94

C.G. Ordzhonikidze, L.K. Ramaiyya, E.M. Egorova, A.V. Rubanovich
Genotoxic Effects of Silver Nanoparticles on Mice *in Vivo* 99

K.G. Skryabin, E.B. Prokhortchouk, A.M. Mazur, E.S. Boulygina, S.V. Tsygankova, A.V. Nedoluzhko, S.M. Rastorguev, V.B. Matveev, N.N. Chekanov, D.A. Goranskaya, A.B. Teslyuk, N.M. Gruzdeva, V.E. Velikhov, D.G. Zaridze, M.V. Kovalchuk
Combining Two Technologies for Full Genome Sequencing of Human 102

N.I. Grineva*, T.V. Akhlynina, L.P. Gerasimova, T.E. Manakova, N.G. Sarycheva, D.A. Schmarov, A.M. Tumofeev, N.M. Nydenova, L.Yu. Kolosova, T.I. Kolosheynova, L.G. Kovaleva, S.V. Kuznetsov, A.V. Vorontsova, A.G. Turkina
Cell Regulation of Proliferation and Differentiation *ex vivo* for Cells Containing Ph Chromosome in Chronic Myeloid Leukemia 108

E.V. Gromenko, P.V. Spirin, E.A. Kubareva, E.A. Romanova, V.S. Prassolov, O.V. Shpanchenko, O.A. Dontsova
Deamination of 5-Methylcytosine Residues in Mammalian Cells 121

E.S. Piruzian, V.V. Sobolev, R.M. Abdeev, A.D. Zolotarev, A.A. Nikolaev, M.K. Sarkisova, M.E. Sautin, A.A. Ishkin, An.L. Piruzyan, S.A. Ilyina, I.M. Korsunskaya, O.Y. Rahimova, S.A. Bruskin
Study of Molecular Mechanisms Involved in the Pathogenesis of Immune-Mediated Inflammatory Diseases, using Psoriasis As a Model 125

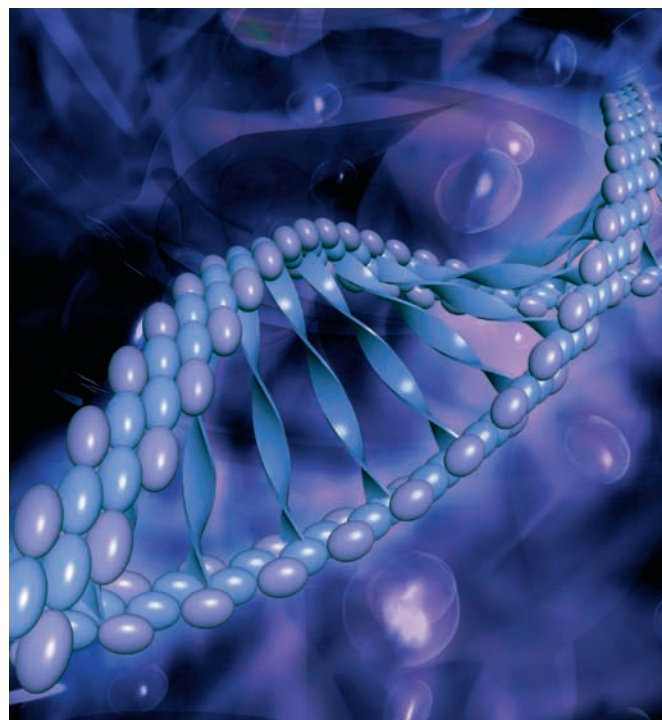


IMAGE ON THE COVER PAGE
 Bigstockphoto.com/STEELCOAST CREATIVE/Sean Acierno

The Pharmaceutical Industry in Russia: Reality and Prospects

A.I. Gordeev, Fund «Open Economics»

This section “Forum” is about the pharmaceutical industry in Russia. We were encouraged to debate this topic after the unveiling of the Strategy of Development of the Pharmaceutical Industry in the Russian Federation developed by the Ministry of Industry and Trade of the Russian Federation. The majority of our experts, who are authorities in the federal government, business, academia, and industrial science, believe that Russia needs a fully developed pharmaceutical industry. What are the main arguments for an intensive development of a Russian pharmaceutical industry? In our opinion, there are four major reasons.

1. NATIONAL SECURITY

According to the majority of our experts, the provision of national security is one of the most important arguments underpinning the necessity to develop a domestic pharmaceutical industry. Most importantly, it will serve to provide the country with pharmaceutical drugs in case of an emergency. It is, indeed, a very important point; we need to remember, however, that our pharmaceutical industry can provide simple drugs for our people no matter the situation. According to statistical data and inquiries of leading clinics, provided by STRF.ru, the share of Russian-made drugs on the market is about 70 %; we need to mention, however, that Russian companies have a tendency to use foreign-made raw materials instead of domestic ones.

It is true that we produce hardly any cutting-edge or very innovative drugs, but that is not a crucial factor of national security in case of a natural or military disaster. However, national security has not only a military and political dimension, but also an economi-

cal one. Import-substitution, especially when there is instability in foreign currency markets, is an obvious priority of the state’s social policy, because it is directly related to basic constitutional values.

2. TECHNOLOGICAL DEVELOPMENT AND THE ECONOMY

The pharmaceutical industry is believed to be one of the most high-tech-intensive industries. That means that we need at least to put serious attention on its development, as part of the implementation of the state’s effort to re-calibrate the Russian economy towards more high-tech industries. The economical parameters of the Russian pharmaceutical market, such as its capacity, dynamics, stability of demand, could be a good reason for the state to focus more attention. It is remarkable that the number of people employed in the industry is not very high (50-70 thousands). The strength of any hi-tech industry is in the stimulation of scientific research, as a result of a competitive development. The innovation that

drives progress in pharmacy are the achievements in biological and medical sciences. Progress in the pharmaceutical industry helps develop chemistry, physics, mathematics etc., and that will stimulate demand for the universities to train more highly qualified specialists. It is clear that a strong pharmaceutical industry drives the demand for results from the academic and industrial R&D research sector, as well as it stimulates growth in economic employment.

Pharmaceutical companies receive orders from the medicine branch to produce certain drugs, acquire information on new research and investigations, and, on the other hand, these companies can stimulate such investigations and propose new ideas to scientists. So, these companies influence both fundamental and applied research. In the entire cycle of innovation, from R&D to mass production, there is demand for very educated, highly trained people.

3. IMPROVING PEOPLE’S QUALITY OF LIFE

At the level of the state, the pharmaceutical industry is a mediator between the state and people with a couple of important social functions required to lengthen and improve the quality of life of the Russian population. It is obvious that we cannot realise those functions if we lack the technologies and the creative energy that would push us to do more and more research! Without the development and use of technologies, without possession of a certain amount of them, without the infrastructure for permanent innovation in this area, it is impossible to achieve a serious improvement in the quality of our health care sector. Even



Photo: I.A. Solovey

A.I. Gordeev

when we have new knowledge, new research in Russia, the pervasive nature of foreign innovations hobbles the development of an infrastructure for the production of new drugs and the development of new treatment methods. The absence of a mediator---a fully developed pharmaceutical industry---will result in low-quality of medical services and lack of demand for certain areas of science. There is no doubt that we can copy and use foreign research; however, we have no legal access to the best of it, and what we can buy is usually outdated and not the best. In other words, if we stop buying imported products, we will have second-tier medicine in Russia.

4. ACCESS TO THE INTERNATIONAL PHARMACEUTICAL AND BIOTECHNOLOGY MARKET

During the last several years Russia has lost practically all its positions on the international market of pharmaceuticals and biotechnology. A return into this market could be an important step in the effort to move from a “row material economy” to the fundamentally different participation in international trade.

KEY PROBLEMS FOR DISCUSSION:

What exactly is the Strategy “Pharma-2020”? In short, we believe it is an

industry strategy with the purpose of supporting Russian pharmaceutical companies (by “Russian pharmaceutical companies” we mean any enterprise that provides a full cycle of drug production on the Russian territory). “Pharma-2020” was created with full account of the prospects of a growing domestic pharmacology market, and with the assumption that the state will invest resources into its development. Those behind the Strategy claimed that their primary goal was to create conditions for a “transition to an innovation-based model of development” of the Russian pharmaceutical industry. Yet they only schematically indicated both the mechanisms and instruments that were to ensure this transition. This is not surprising, because this was a strategy of industrial development in conditions where it is practically impossible to ensure coordination between the bureaucracies of different departments. Practically, the Strategy left untouched “border” questions, such as the provision of medical services to the people, other programs for the development of science and technology, as well as the new legislative rules that will appear as a result of collaboration between science, business, and manufacture. What are the key points we should focus on and which are crucial to potential success in the implementation of the Strategy?

1. LACK OF A CLEAR, FUNCTIONING MECHANISM UNDERPINNING INVESTMENT IN INNOVATION

Innovation (we plan to focus a lot of attention on innovation in this issue) is key in the creation of a full-fledged pharmaceutical industry. Do we have money for this? The world pharmaceutical industry is 2nd in the level of investment in research and development. This was mentioned in the analytic review «R&D Scoreboard 2006» of the Ministry of Trade and Industry of the United Kingdom, based on an analysis of 800 British and 1,250 companies worldwide that are the most advanced in terms of investment in R&D. Pharmaceutical companies are sandwiched between producers of high-tech gears (1st place) and car producers (3rd place); software companies come in only 5th.

2. LACK OF A HIGHLY ELABORATED REGULATORY SYSTEM THAT SHOULD STIMULATE THE DEVELOPMENT OF THE INDUSTRY

We do not consider here the rules and regulations for registration, market entry, and the marketing of pharmaceutical companies, because they are issues that refer to the relationship between the Ministry of Industry and Trade and the Ministry of Health and Social Development. We want to focus on the problems of regulation of early steps in the innovation chain in pharmacy. If we consider the situation “in general,” it would appear that there are no problems. There is a significant increase in funding for academic science, and the development of specialized industrial and state-controlled special-purpose programmes (2 programmes of the biological branch of the RAS, the Federal Special-Purpose Program of Rosnauka); there is also a new chapter in the Russian Fund of Fundamental Research; there are new state corporations and venture capital funds with state participation and the Russian Venture Company; there are the programmes “Start,” “Temp,” and “Pusk” at the Bortnik’s Fund; and there are some changes in the regulation of Intellectual Property. However, the situation evolves very slowly, especially in pharmacology. The state is not ready to pass on the results of state-funded R&D to business; it is not ready to support high-tech industrial business without dictating the rules at all stages. The state is trapped in its own logic of administrative reform; state management has become a set of competing programmes with an unbelievable amount of criteria and regulations. That leads to lack of agreement between different departments during the implementation of the policy; and, even more confusing, the legal documents for this policy are very uncertain and contradictory. If we fail to solve the problem now, all of our steps further down the road and innovation programmes in pharmacy-industry will be ineffective. Collaboration between the Ministry of Health and Social Development and other departments and players on the pharmaceutical market is still deficient. Even if we have some reasonable ideas on how to solve this problem, we lack the mechanisms and instruments needed for implementation. Innovative

development of the pharmaceutical industry should be based on a full innovation chain – from R&D to the distribution of ready drugs. That means that this strategy cannot be fully elaborated without a detailed elaboration of the mechanisms of collaboration between the Ministry of Health and Social Development and other departments and state institutions. Recently, very active and effective consultations between 2 key players in the development of a domestic pharmaceutical industry – the Ministry of Industry and Trade and Ministry of Health and Social Development – were launched. The Ministry of Industry and Trade is responsible for pharmacy as an industry; and the Ministry of Health and Social Development is a key player on the pharmaceutical market, and, particularly, a) it regulates legal aspects for the majority of parameters underpinning the process of production and marketing of drugs; b) it has departmental industrial organisations which are players on this market; c) and it is one of the major buyers on the pharmaceutical market. For a long time, the policy of the Ministry of Health and Social Development had not been helpful in the development of the Russian pharmaceutical industry. Even if the ministry sometimes proposed reasonable plans, not all of them were realised. Our bureaucrats have a fantastic ability to kill the solution to any problem and leave the situation as is convenient to them. In a modern competitive world, this approach does not work. After the replacement of the management of the Ministry of Health and Social Development, the policy of the department began to change, and some reasonable actions to improve the

domestic pharmaceutical market were proposed. We have reason to believe that the realization of those proposals will solve this problem.

3. ABSENCE OF NATIONAL PRIORITIES IN THE DEVELOPMENT OF MEDICINE AND PHARMACY

This is one of the key problems, without the solution of which development will enter a “blind alley.” Investment is considerable, and technological cycles are long in the pharmaceutical industry, and company executives want assurances that the new drugs they develop after much investment and effort will be on demand. Development of a set of priorities, in other words selection of the main areas of medicine in need of innovative drugs is therefore a very important step; without this step, a stable development of the pharmaceutical industry is almost impossible. Are those priorities fighting cancer, cardiovascular diseases, infections or something else? This question needs an answer now, with full account of the Russian reality and the opinion of scientists and clinicians. The set of priorities is not just a list of the most important diseases. It is also important, especially for high-tech bio pharmacy, to establish priorities in the area of development of multi-functional methods and universal platforms for new molecular substances and drugs designs, such as recombinant proteins or new methods of targeted drug delivery. Because of Additional Drug Provision and the system of state procurement, the Ministry of Health and Social Development has become one of the major buyers on the market. This department influences much of what goes on on this market; nothing

will change in Russian pharmacy without a well-thought policy by the Ministry of Health and Social Development as a major player and buyer. In this edition, we plan to look into all of the above-mentioned problems in details, and to offer a chance to the experts involved in modern drugs, technologies, and enterprises in biotechnology and pharmacy to voice their opinions.

UPDATE OF THE SITUATION IN THE PHARMACEUTICAL INDUSTRY; DO WE HAVE INNOVATIVE DEVELOPMENTS!

Some experts claim that we never had a fully developed pharmaceutical industry and related scientific research. They often offer for argument the fact that we produced generic drugs, and that we borrowed most of the technology from abroad and simply focused on manufacturing. Moreover, they claim that all of that manufacturing went on in the republics we lost after the collapse of the Soviet Union, or in the Warsaw Pact countries. In reality, that is not true. Firstly, we still have a lot of factories in Russia. Yes, they are weak now, but we can use them to develop a future manufacturing base. Secondly, such claims have more to do with the “chemical” aspect of pharmacy; i.e. organic synthesis. In biotechnology, as experts believe, we were more advanced. A lot of our research in this area was really competitive as compared with foreign research. One of the explanations for this state of affairs is the fact that this sphere of science is related to the State Military-Industrial Complex and the design of biological weapons and means for protection against them. We still have good teams, schools, and experienced people in biotechnology. Therefore, it is possible to develop bio-pharmacology in our country. In fact, the “chemical” and biotechnological bases of pharmacy are not contradictory, but complementary. They are, in fact, two ways towards the possible development of a Russian pharmaceutical industry.

In any case, we can definitely say that

a) a domestic high-tech pharmaceutical industry still exists in Russia

b) all of the innovations in this area are based on Russian scientists’ R&D (even business and science hardly criticize each other).

REFERENCE

“Pharma-2020” (The strategy of Development of the Pharmaceutical Industry in the Russian Federation Up to 2020) was developed and proposed by the Ministry of Industry and Trade in 2008. The aim of the strategy is to develop ways towards the realisation of the priorities set in the development of the Russian pharmaceutical industry, to be the basis for a public-private partnership in the different aspects of development of the pharmaceutical industry, to provide coordination between state institutions in the development of this industry, to define vectors of development and fine-tune legislative rules regulating pharmacy and to be the basic strategy for major state decisions regarding the development and implementation of special-purpose programmes and projects.

Here, we will mention few examples of those R&D:

1) The group of companies “Bio-process,” in collaboration with “Gos NIIGenetika,” is designing important bio generics as a part of an innovative project (the state is co-funding this project to the tune of about 1 billion rubles). As a part of this project, they are developing an interferon alpha-2b of world-class quality with the expectation to bring it to market. This substance is needed for the treatment of influenza, upper respiratory tract viral infections, viral hepatitis B and C, and some onco-diseases. They are also developing a substance of erythropoietin and some enzymes.

2) “Pharmstandart” PLC is actively marketing certain drugs that were developed at the MM Shemiakin and YA Ovchinnikov Institute of Bioorganic Chemistry of RAS; for example, “rastan” (the first domestic growth hormone), neipomax or philgrastim (it is a granulocyte colony-stimulating filter)

3) The company “Biocad” is developing its own research center. The main campus is in Liubuchany, Moscow Region, it is based on one of the state research institutes, and another campus is in Novosibirsk and is based on the Research Institute of Genetics and Molecular Biology. It is obvious that “Biocad” uses educated people who received good training from the state. But the private company, in fact, has saved the research institute and created a new system of management. Is it a real example of effective continuity; it appeared “despite,” not “thanks to” the conjecture.

At almost every pharmaceutical company we were told that Russian scientists are regularly called upon to do some research. Our best intellectual resources in biotechnology and organic chemistry are concentrated in research institutes of the Russian Academy of Sciences. However, collaboration between research institutes and pharmaceutical companies seldom leads to any success. When the first private pharmaceutical companies appeared in Russia, they tried to contact all the reasonable research teams and to identify interesting projects. The result was close to nothing, because they looked for “almost ready to market” products at the stage of marketing or, at the very least, at the stage of clinical trials. Russian scientists did not have such innovative products. Why? Because as we mentioned earlier, funding of research institutes remained very pure for a long time. Research grants were dolled out, but practically not enough funding for serious preclinical-stage research. There are very few precedents. Russian scientists also did not have modern equipment such a sequencers. They had a lot of ideas and theoretical research, a lot of small but high-quality experiments, but not as a system. However, it often appeared that interesting scientific results that make scientists proud were neither useful nor interesting as potential pharmaceutical products. And time was needed to move them to the stage of marketing. Because of the pure technological basis and lack of practical experience in high-tech manufacturing, development of a new

drug takes a lot of time. There is limited experience of successful collaboration between scientists and industry; however, a stable working relationship does not exist. Let’s repeat: the main idea in the Strategy is the active innovative development of the Russian pharmaceutical industry, i.e. capturing a the substantial share of the innovative drugs market, and developing production of common generic drugs. Supporting and developing the production of Russian generic drugs is a simple and technologically understandable task. It is a question of political will and quality of management. Production of generic drugs is not complicated; it is very understandable from the point of view of economic processes, technology, and pharmaceutical development. There will be, obviously, fierce competition with India and China, but we can fit into this economical model of existence and develop it. What’s more, there are examples when our generic-producing companies had good economical achievements. Yes, they obviously do not make a lot of profits, but 15% profitability in high-tech production is better than no profitability at all or factories with very old equipment and poorly qualified workers. Developing an innovative way for bio-pharmacy in Russia is possible, because there are a lot of scientific teams in our country that can compete with foreign teams. The main problem is that coordination between drugs manufacturers and scientific research teams is extremely weak, and without such coordination, some areas of scientific research are withering. ●

Biopharma: How Can the “Death Valley” between R&D and Innovation Be Overcome?

A.L. Konov, A.A. Leonov

“Today Russian biopharma has two obvious main problems,” believes Alexey Konov, investment director, and Andrey Leonov, investment manager, at Bioprocess Capital Partners Ltd. Russian satirical poet Igor Guberman’s line about a “crowd of researchers staring at life’s enigma” was written more than 25 years ago. If we consider the sad end to that quatrain (about “life that sends all of those researchers far, far away”) as an illustration of the process of putting new knowledge to practice, we must admit that the current situation is even worse. In fact, until the collapse of the USSR (in the end of the 1980s), the pharmaceutical and biopharmaceutical industries provided most of the drugs the nation required. Biopharmacy in the USSR functioned the way modern corporations do. Each branch had specialized Research Institutes in which production-orientated R&D was conducted. There were also communal usage centers that worked on certain scientific tasks of the Soviet biopharmaceuticals industry. Centralized R&D and innovations made the Soviet Union one of the world’s leaders in certain fields.

TWO KEY PROBLEMS

The two key problems today are (1) the lack of promising Russian R&D and (2) the absence of an internal market for that R&D

The first problem is the absence of innovations ready for registration and to go to market. The booming development of biopharmacy throughout the world in the 1980s–90s occurred as science and technology stagnated in Russia. The gap between us and the developed world in this field is very wide: we have no products that are ready for introduction nor do we even have the technology to produce them.

This is especially true for products produced by eukaryotic cells (recombinant proteins, blood coagulation factors, and therapeutic monoclonal antibodies). For this reason, the percentage of locally produced biotechnological products used in the Russian Federation (RF) for drug production is critically small (only 2%). For comparison, the share of locally produced hi-tech chemical substances used in our medical industry is signifi-

cantly higher; 15 % quantity-wise and 5% money-wise. This gap is filled by imported substances. Our main partners are China and India, and large Western companies provide the most expensive biotechnological preparations. At the same time, development of new industrial strains and technologies happens very quickly abroad: the technologies there are much more developed than the domestic technologies that have been in use in Russia for 15 years. How can we design a portfolio of innovative projects? Obviously, there are two possibilities:

- we can attempt a transition from R&D to industry ourselves,
- we can try importing the good western innovations that have appeared on the market and are ready for registration and industrial production

The second problem is the absence of modern producers ready to accept the most advanced innovations. The standard scheme in the innovation process starts with research, followed by testing, and then introduction into the

market. Even if we try to modify this process and adapt it to Russian realities, it is very unlikely we would succeed in the absence of an internal market of venture activity. Even if we create mechanisms for the incubation of these projects, we will face the problem of finding a buyer for them.

We lack any significant market in modern biotechnology in Russia; i.e., there are no big players ready to step in as the main consumers of the new technologies that might appear as the result of investments. The development of the biotechnological industry in Russia today lags behind that of most leading countries; Russia’s share in the global production of biotechnological products is less than 0.3%, and we are almost absent in biopharmacy.

This means that we need to solve those two problems in combination: we need a portfolio of projects, and we need to build a system for accepting the products.

In this paper we will not address the important problem of importing innova-



Photo: Beprocess Capital Partners, Ltd

A.L. Konov

tions into Russia, we also will not touch on the extremely important problem of creating “acceptors” for those innovations.

We want to concentrate on the possibility of creating a domestic product that is ready to go to market. We will try to show below that none of the existing institutions of development and venture capital is fully adapted for work at the early (incubation) stages, when the product needs to be developed practically from the ground up.

DOMESTIC DEVELOPMENT: HOW TO OVERCOME TWO “DEATH VALLEYS”

As said in the strategy of the Pharma-2020 project, “In the current situation, the Russian pharmaceutical industry can not survive because it is squeezed between Western transnational corporations, which dictate the rules in the areas of technology and intellectual property, and producers from India and China, who put a constant downward pressure on price.” Neither new budding Russian pharmaceutical companies nor the created venture funds are yet ready to finance the long-term and risky development of innovative pharmaceutical products.

There are two “Death Valleys.” The first is the transition from ideas and successful primary experiments to a working model. The second is the move from the stage of a brand-new business to that of a rapidly growing company. In standard practice around the world, venture capital takes care of the first valley and “cultivating funds” and “angels” (investors) cover the second valley. The angels in developed markets are usually what are called the 3 Fs: fools, family, and friends. It is not an easy task for Russia to find a way to cover both “valleys,” but we need to find this solution!

Let’s consider the standard western algorithms of commercialization of research in pharma/biopharma and try to spot the typical Russian “white spots” in this process (see the scheme in Fig. 1)

State laboratories perform research, which sometimes yields very promising results regarding the design of new drugs and technologies. However, there is a so-called “regulation process” between the drug prototype and drug on the market or industrial technology; this “regulation process” is very expensive, lengthy, and risky.

In Fig. 2, the scheme is shown: the risks of rejection and spending are extremely high in the early stages; however, spending is lower in the trial stage.

The imperfection in the law and the underdevelopment of institutions that specialize in providing seed money lead to a trough (the grey zone in Fig.1): most potentially promising projects go into a financial trough, and only a few go directly from the green zone to the yellow zone, where the mechanisms of venture capital financing start to work (because of this the grey zone is not very significant there). (See text for a more detailed description – *Ed.*)

The active substance (molecular) produced at the R&D stage should be tested on animals before trials on humans can begin. At the preclinical testing stage, the toxicity of the new molecular substance and its pharmacokinetic and pharmacodynamic parameters should be assessed and the effect should be modelled. Regulatory agencies analyze the information about a new drug from the preclinical tests and decide whether or not trials should proceed on humans.

Clinical trials of drugs before official permission for medical usage is received usually proceed in three stages traditionally called “phases of clinical trials.”

At the first phase of clinical trials (phase I), clinico-pharmacological and biomedical trials are provided to a small group of (usually 36) healthy volunteers. At this stage, researchers investigate the side effects of a dose of the drug and its pharmacokinetic parameters and pharmacodynamic effects. This phase is important, because knowledge about the side effects and safety of the drug is necessary before deciding whether to continue investigations or end research.

The initial dose, regularity, and method of administering a certain drug are all usually established in preclinical tests on laboratory animals. However, because of the difference in human and animal pharmacokinetics and pharmacodynamics, correcting dosages could be required.

If the drug is safe and has no side effects, investigators start the second phase of clinical tests (phase II). This phase requires more volunteers, usually from 100–150 patients, but with diseases or conditions that the active ingredient in this drug is designed to cure, diagnose, or prevent.

The aim of phase II is to prove the clinical effectiveness of this drug for a certain group of patients (estimating the short-term safety and determining the therapeutic dose and scheme of dosage). Phase II trials are the most important step in deciding whether or not to continue the drug’s development.

If the drug proves effective and safe in the second phase, the investigation continues into phase III. Clinical trials in the third phase are closely monitored investigations designed to assess the safety and effectiveness of the drugs in conditions that are very close to the real conditions of medical treatment.

The aim is to determine the long-term ratio of safety to effectiveness for the medical forms of the active component. Usually, these investigations are related to existing standard therapy (or placebo for a new class of drugs). Innovative products can be registered after this phase of clinical trials. The number of patients in this last phase of research could amount to a thousand people,

depending on the drug and condition. Therefore, each new step is more expensive than the previous one for the drug's developer.

It is obvious that state institutions are incapable of bringing a new medical drug to market. Promising products designed in Western countries, mostly in universities and laboratories, are proposed for sale. Pharmaceutical companies buy many products and technologies that are still simply ideas, through licences or direct purchases. More often, development at the early stages happens through a “partnership,” with the possibility for the large company to own the product if it is successful.

If a team of researchers from a state institution see that the product or technology created has serious commercial potential and believe they can start their own business, then they do so. The conditions in this situation are clear: all the rights to the product or technology belong to the company, and the state institution employing the team of researchers takes a stake in it. After that, there is a business plan, the aim of which is to attract external financing and continue research with the intention of bringing the product to market and the promise of a huge profit. After that, the founders of the company try to attract all potential venture capital funds. To secure investment, they need to persuade, convince, and lay out their dream. Most startups end at this stage, but some are lucky. If so, they receive some seed money in exchange for shares. They spend money on investigations, and if they have promising results, the startup can attract more investment. The results are usually not promising, investment dries up, and the startup dies. The reality is that very few survive. The companies that are lucky create a product that brings a big profit in comparison with the money invested, and shareholders benefit.

In Russia, we need direct licensing from state institutions to the pharmaceutical companies, and, even more importantly, we need to create a way for small startups with ideas to succeed. Let's consider the major problems which, in our opinion, could be stalling the effective development of small innovative pharmaceutical companies, particularly biopharmaceutical companies.

Organisational Problems: How Should the Company Be Founded and How Can Laboratory Investigations Be Done Legally and Effectively?

There is no tradition of small companies being founded by scientists in Russia. They simply do not understand what needs to be done, which papers are required, etc. We do not have specialised industrial parks or incubators which are ready to provide not only legal and organizational support, but also laboratories and logistical infrastructure. This is important, because in this case we are talking about a company involved in research. That means that, apart from the usual problems of a small company, this company will need permission to work with chemical reagents, biological objects, radioactivity, etc. Today this problem is usually solved by renting property at an institute where research has been done before. Very often, companies formally rent just a couple of square meters, but in practice they use all of the equipment there. This is possible, however, only when the director or dean allows it. Otherwise, the scientists work illegally or semi-legally.

How Should the Intellectual Rights to the Project Be Transferred to the New Company?

If a patent has not been registered yet, in the condition of “know how,” the big question is who should be registered as the owner? If the owner is a state institution and the scientist is only the developer of the idea, the most he can ask for after commercialisation is to receive royalties. However, not a single state institution today is capable of bringing the product to market because of lack of funds, authority, and motivation. You're thinking maybe it is possible to register the idea to your name or to a company that has yet to be established? That's illegal, because all of the research was conducted during work hours, on equipment at work, and with the state's financial support. So, the key is to solve the question of intellectual property, with the rights to the results of the investigation going to the researchers and with the legal possibility of separating from the mother organization and founding a small enterprise (inside or outside the technology park).

Today, this is becoming possible because of very important amendments to the law.



A.A. Leonov

According to Federal Law 217, budget and education establishments, as well as scientific establishments, universities, and colleges that are founded by the state's science academies, “Have the right, without the agreement of the proprietor and with the notification of the federal body of the executive power which provides the functions of elaboration of state policies and normative legal regulation in the sphere of scientific and scientific-technical activity, to found (together with other persons) business companies whose activity practically apply (introduce) the results of intellectual activity.”

Regarding the production of “custom” work in state establishments with the aim of introducing it into industry, we think that this scheme has proven to be inefficient. Nothing can be done without the direct material interest of developers, not only in the project report, but also upon bringing the product to market and growing the company.

HOW TO SECURE FINANCING!

Today, practically the only form of state support for a startup in Russia is the program of the Fund for Assisting Small Enterprises in the Scientific-Technical Area.

Over the last few years, the fund has created real working mechanisms for the commercialisation of scientific investigations. Financial support from the “Start,” “Start-up,” “Universities,” and “Youth Business” programs

support projects at the early stages of R&D.

For several years, the fund provided up to 20% of the program's funds in the area of biotechnology and medicine. Remarkably, one of the fund's tasks for the nearest future is to collaborate with venture capital funds (as well as other structures) in order to provide financial support to projects throughout the whole innovation cycle.

If the company is established in Moscow, the Moscow government provides support. The government of Moscow founded a not-for-profit agency for developing innovative entrepreneurship, which has created a couple of interesting programmes, in particular:

- subsidies for small innovative enterprises for the production of prototypes

or small batches (this is a very important step for demonstrating the "proof of principle," which is exactly what should be shown to venture capital investors when requesting financial support),

- subsidies for patenting the results of innovative activity both in the Russian Federation and abroad.

Similar local programs have begun in other regions as well.

Finally, we should mention the Rosnauka programs for development at the middle stages. The contribution of those programs in supporting some of the leading Russian laboratories of the Russian Academy of Sciences (RAS) is hard to overestimate; however, we should remember that these programs are mainly aimed at financing applied

research. This may be why the effectiveness of these programs is far from what is desirable. Due to their status, mentality, and motivation, academic laboratories do not adapt very well to the design and introduction of their products into the market if they have to meet deadlines.

HOW AND WHEN TO CONDUCT TRIALS!

The company has been founded, R&D completed, and the trial (preclinical and clinical) stage begins. How and where can these trials be conducted according to Western standards with the possibility of selling this product in Western markets in the future? We are sure that today no small innovation company can support the whole cycle of trials by itself. A good plan for venture capital fi-

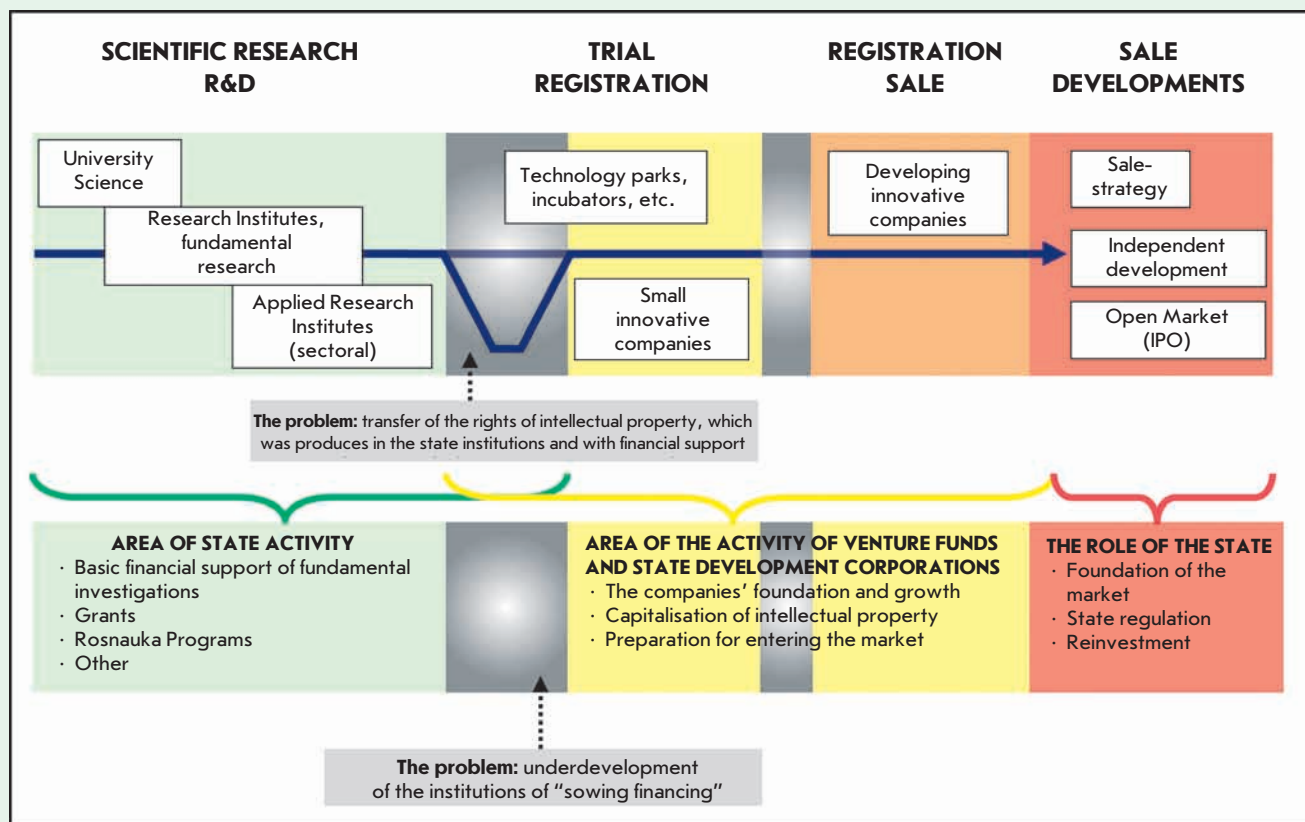


Fig. 1. Standard scheme in the innovation process. The colors illustrate the maturation stages of projects from the initial (green) to the mature stage (red). In the area of pharmacy and, in particular, biopharmacy, development goes from the R&D stage through the trials and registration to appearance on the market (blue line in the figure). If the development appeared in a state institution (university or research institute), the first problem that appears on the way to future commercialization is how to transfer the rights to the intellectual property from the state institution to a private company. The long-awaited and recently passed amendments to the law (see "Federal Law N 17" FL dated August 3, 2009) should solve this problem; however, we need to wait and see how this new law works in practice. The second problem, in our view, is more serious: it is financing companies when they are in their early stages

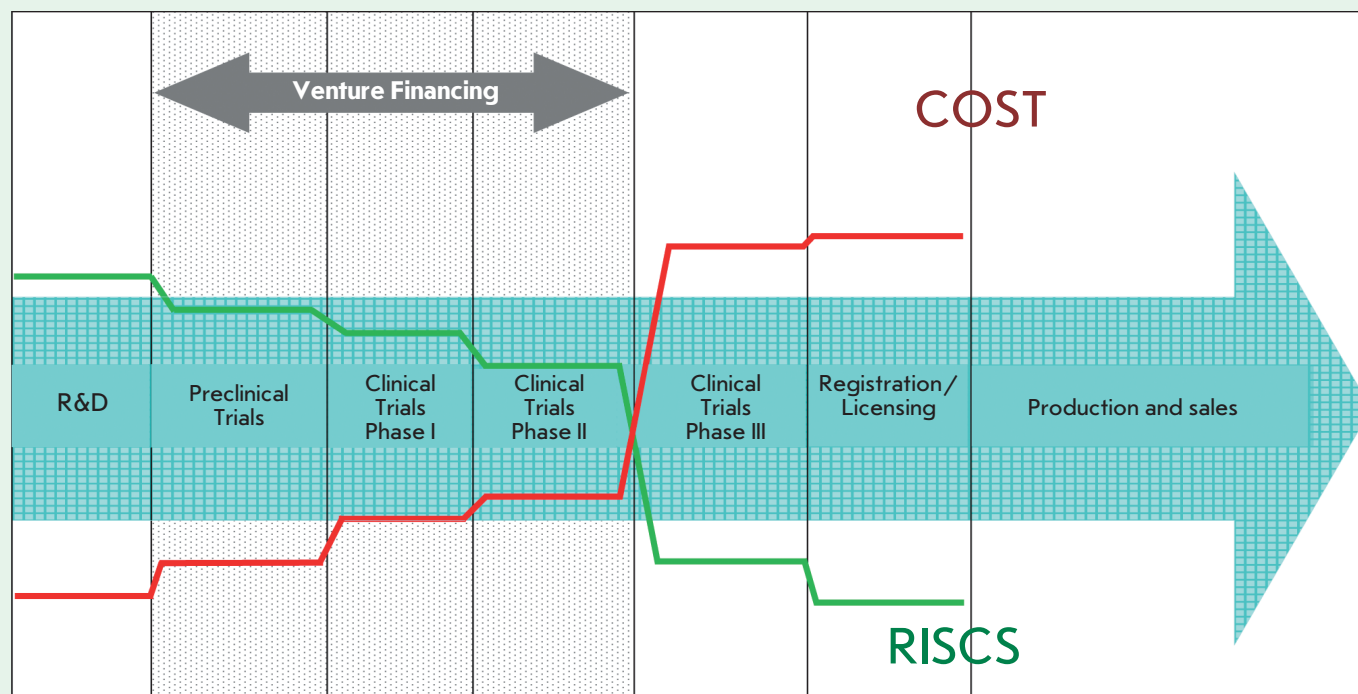


Fig. 2. Production of a new drug: from development to the market.

Venture financing is very commonly tapped after a successful completion of R&D. Funds usually dry up for the project before the clinical phase III, and often even earlier, after the second and even first clinical trials. (See description of stages in the text – Ed.)

ancing of the proposal, professionals in the preclinical and clinical trials, and secure partners (who are specialists in these types of research contract organizations) are required for this stage.

Project management and marketing are two parts of a successful pharmaceutical company; they should be fully covered by the company's own recourses.

Other activities are potential candidates for outsourcing. A broad variety of contract service organizations (CSOs) could work as subcontractors, i.e., Contract Research Organizations (CROs), which provide contracted services in preclinical and clinical trials, or Contract Manufacturing Organizations (CMOs), which optimize the process of production and production of active pharmaceutical substances and finished medicines for the trials, and, afterwards, for the market. Even today, small innovation companies around the world spend up to half of their entire R&D budget on CSO services.

HOW AND WHEN THE LIFE OF START-UP ENDS

The company has completed a whole set of trials, and the results of the first

phase are good. What now? Attract a lot more money for phase II and phase III trials? Search for a partner? Sell the business?

It is hard to provide a simple answer. This depends on the type of drug and the situation in the company and on the market at the moment. Let's make a couple of general remarks.

If in the 1970s–1980s the largest companies in the world introduced 5–7 new molecules every year, today they introduce a maximum of 2–3 fully original molecules in the course of several years. To cover the expenses for both the leading development, as well as for the spinoff developments, the new molecular must bring in hundreds of millions of dollars every year. Over the last 20 years, the largest medical producers in the world had other options, mainly because of their biotech “hits,” which were mainly developed by small companies. Today, the patents covering most of the bio-blockbusters that have been on the market for 15–20 years are running out; the industry is desperate for new “hits.”

The difficulty with compounds reintroduced into the market, the appear-

ance of new knowledge about how they act, and new statistics about their side effects have led to a shift in major expenses from R&D to trials. The EMEA and FDA, the leading international regulating agencies, demand more and more complicated trials of new drugs. As a result, today expenses for the introduction of a new molecular to the market have increased 3–5 times and can amount to 300–900 million dollars, which creates an insuperable barrier for small companies when they try to repeat the successes that venture legends Genentech and Amgen achieved 25 years ago.

However, small innovation companies can still attract venture capital at the early stages and develop new products up to the stage of the first trials (the end of preclinical trials or phase I of the clinical trial); after that, they can establish long-term cooperation with a “big player” as an investor in the next stage or simply sell the product to them.

The beginning of a recession in the global economy and the dearth of liquidity affecting the aforementioned big players and potential strategic

investors have brought forth the tendency to finance new developments in more advanced stages by attracting the next rounds of venture capital financing. For example, in September–October 2008, five companies secured the largest amount of late-stage venture capital financing; three of them were biotech companies. The company Proteolix, for instance, attracted 79 million dollars from venture capital investors to support the phase II clinical trial of a biopharmaceutical drug against autoimmune diseases and cancer; MacroGenics attracted 25 million dollars for the phases II/III of the clinical trial of an original monoclonal antibodies to fight diabetes; Link Medicine received 40 million in venture capital financing for the late stage of development of a cure for autoimmune diseases.

Therefore, in the next 3–4 years, the situation on the market will be very favorable to small venture capital funds and developers of certain products: the large companies that weather the hard times will continue to actively buy new companies to boost their product port-

folios. On the other hand, after raising funds in additional rounds of venture capital investment, there is now the unique opportunity to either grow the company to the stage when the product can be taken to market or sell the company to a strategic investor at a significantly advanced stage and for a much better price.

Regarding IPOs, we should note that, before the crisis, some biotechnological giants successfully went public. However, this is uncommon, because selling to a strategic investor in the form of a corporation involved in the same field remains the most attractive option for a venture biotech company. Today the IPO option is practically closed, and it is hard to predict the situation that will prevail in the next 3–4 years.

CONCLUSIONS AND RECOMMENDATIONS

To set up a process that will ensure mass production of innovations in the industry of biopharmaceuticals, we need, on the one hand, to solve the organizational, legal, and regulatory problems we have briefly mentioned

in this paper. On the other hand, it is unlikely that that process would be efficient without the creation of an internal market for the results of the activities of small companies: we do not have companies of the size of Bayer or AstraZeneca, which are able to spend \$100 million or more on development. In this context, small innovation companies must sell their developments to the West. The state should focus attention on creating a system in which Russia will have a chance to keep its rights at least on the Russian market. How should this be done? That's a separate question. Let's only say that one approach could be to consolidate several small innovation businesses into a large "virtual" big company; this would not involve a consolidation of buildings and equipment (i.e., immovable assets), but a consolidation of the rights to the intellectual property under development. Such a "disintegrated" company, which could outsource other steps in the process but conduct R&D and marketing itself, would, in effect, be a bureau that is flexible and can quickly react to the demands of the market. ●

THE DESIGN OF INNOVATIVE DRUGS INSIDE INDUSTRIAL COMPANIES

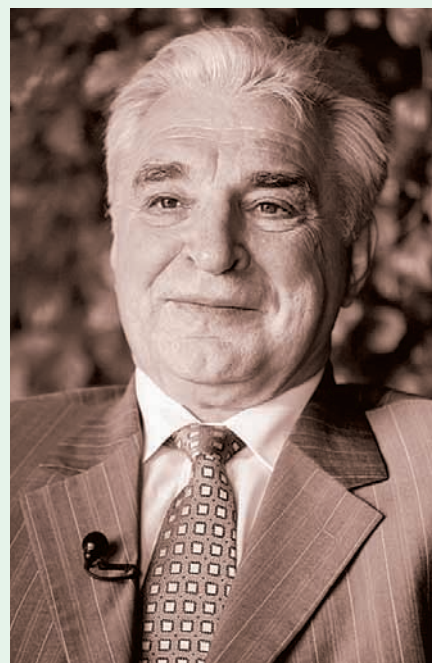
Figure 1 shows how a product designed by an industrial company is developed. Big transnational corporations have their own R&D departments with a wide range of research activities; however, efficiency here (the ratio between the amount spent and the quantity and quality of the newly developed drugs) is significantly lower than that of a small innovative company, which usually aims to create a certain drug or technology. Can we expect the development of an innovative product in Russian industry? That is unlikely! Existing companies, regardless of how they are run, are still, for the most part, not ready to invest in high-risk drug innovations. Usually Russian projects in the field of biopharmaceuticals follow the strategy "What's being done in the West? Let's replicate it quickly!" It is important to say that it is not a bad strategy. It works, and works well. One of the businesses we developed at "Bioprocess" (a producer of biogenerics such as interferon alpha, erythropoietin, and granulocyte-colonies stimulating factors) was built based on this model, and only after 5–6 years, when it began making a profit, was it possible to develop new products. China and India also built their industries based on this model, at least in the field of biopharmaceuticals. At the same time, industry in both countries had strong governmental support: direct funding, tax holidays, preferences in registration and purchases,

etc. Now the Indians and the Chinese have begun beating developed countries at their own game, i.e., in the area of innovative products. It is important for us in Russia not to miss out! However, the sizes of even our own, most successful companies do not allow them to invest tens of millions of dollars into the development and registration of really innovative drugs without state support. Until recently, there were two main mechanisms for securing state co-funding for innovative pharmaceutical development: through the Federal Agency for Science and Innovation (Rosnauka) and through the program of the "Fund for the Support of Small Enterprises in the Sphere of Science and Technology." The fund's programs are meant for small companies; they are not suitable for large industrial ones. Rosnauka's programs, unfortunately, are not suitable for industry: the maximum funding period is three years, and the company must fulfill the requirements for "program indicators," which means production should generate a certain amount of money as a return on investment in that period of time. If development is at the stage of the first clinical trials, three years is not enough time not only for production and sale, but also for simply registering the product on the market.

The Ministry of Industry and Trade (Minpromtorg) has similar programs: the time limit for the return of investment has increased to four years, and the beginning of industrial production is the only indicator. This is a bit more realistic for businesses, but it is still far from what is desirable.

Recipe for Russian Insulin

The share of locally produced insulin on the Russian market estimated at more than seven billion roubles in 2008 is, in reality, microscopic: national industry covers only 2 % of the market (see Table 1). This is unreasonably small, especially because a) Russian producers have enough capacity to produce insulin for the entire country, with similar or even higher quality as compared with foreign products, b) hundreds of thousands of insulin-dependent patients in some particular political or economical circumstances could suffer without this life-saving treatment. The World Health Organization recommends that any country with a population of more than 50 millions have its own manufacturing base for insulin. Anatoly Miroshnikov is an assistant director of the M.M. Shemiakin and Y.A. Ovchinnikov' Institute of Bioorganic Chemistry of the RAS, which produces insulin with the trade name Insuran. He talks to *Acta Naturae* on why Russia needs its own factories for the production of genetically engineered drugs.



MEDICAL HISTORY

Anatoly Ivanovich, the share of Insuran on the insulin market is 1,04 %. Is it our limit?

– No. We can provide all the insulin the Russian Federation needs. But let me first recall a bit of history. In 1978, at the dawn of the era of biotechnology in the world pharmaceutical industry, our research institute also began successfully working in this direction. By 1990, only 3 years after the start of production of genetically engineered products in the U.S. and Denmark, we had also created an insulin-producing strain and started semi-industrial research at a biochemical complex in Stepnogorsk, Kazakhstan. At the time, our lag compared with the West was minimal. After 1991, everything went “dead.” Science and the chemical industry in Russia were on the verge of collapse, and western producers divided to our pharmaceutical market and, particularly, the market of insulin. In

2000 did Yurii Luzhkov, the mayor of Moscow, issue an instruction setting up experimental production of genetically engineered insulin? At IBC RAS, we received a 120 million rouble loan and began supplying short and long-form insulin to Moscow clinics 3 years later. According to Luzhkov's plan, we should have ramped up production to 300,000 phials a year. However, the city refused to buy such an amount of insulin: they argued that Moscow will slowly move away from phials insulin. Unfortunately, nobody gave us money for a cartridges production line. Now we are producing 95,000 phials a year and providing insulin to 15 % of Moscow patients suffering from diabetes.

Have you had any reclamation on Insuran?

– Not one in the five years of manufacturing insulin. We produce it according the GMP standards, and each batch of 100 phials undergoes quality control. Chief endocrinologist of Mos-

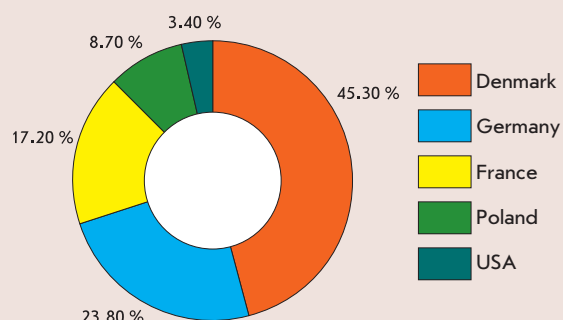
Anatoly Ivanovich Miroshnikov is full member of Russian Academy of Sciences, deputy director of the M.M. Shemiakin and Y.A. Ovchinnikov Institute of Bioorganic Chemistry of the RAS, president of the presidium of the Pushino Science Centre, vice-president of the Russian public organisation “Y.A. Ovchinnikov's Society of Biotechnologists of Russia.”

cow Michael Borisovich Amzyferov can testify to the quality of our insulin, and clinicians also agree that there is no difference between our insulin and Western-made insulin (Elly Lilly, Novo Nordisk). With the same quality, our “Insuran®” is cheaper. However, it is not easy to increase production, because the system for insulin supply to major clinics and pharmacies is already established and nobody likes new players. Long-established connections and, possibly, traditions of corruption play

"Our medicine is not only cheaper but also better. In the production of Insurane, we do not use toxin bromocyan, but Elli Lilly still uses toxin, Novo Nordisk uses a cheap strain of yeast, which is less effective; now they are also switching to bacteria. However, all of this does not mean that a "green street" will be open for us. I am sure that we will have a lot of difficulties with the marketing of our product. Western companies will fight for a market like the Russian one. And they will try to brain-wash doctors that the quality of our insulin is worse. That means that principles of market competition will apply."

From the interview of Anatoly Miroshnikov, 2000, "Meditsinskiy vestnik" v.4

Top 5 importers of the origin of insulin and its analogues, 1st quarter of 2008.



Centre for Marketing Research "Pharmexpert"

a role... As a result, 90% of the Russian insulin market is in the hands of American and Danish companies. But we are not giving up: in the last 4 years, I have been struggling to build a factory in Pushino. The problem was almost solved: on an order of the government of the Moscow Region and the initiative of the RAS and IBC RAS, the joint stock company Bioran was created. The company should anchor the project of industrial design of medical genetically engineered drugs; we even had a plot and communications for this project. But, unfortunately, the financial crisis hit, and Bioran has struggled to secure a loan. If the state had provided a loan guarantee at that time, we would have had money. The factory should have started functioning in 4.5 years, and we would have repaid the loan in 8 years.

What volume of purchase did you have in mind?

– A full-cycle factory should have produced 400 kg of substance. That should have been enough for Russia, the Commonwealth of Independent States, and even for export to Arab countries and South East Asia.

Do I understand you correctly that your technology is not inferior to foreign technologies?

– I can say that our technology is even better!

INSTRUCTIONS FOR USE.

– Does it mean that this question is economical and political?

Absolutely! And we try to explain it to our executive branch. In a letter to Igor Sechin, deputy prime minister, I indicated that, if we want to rebuild our pharmaceutical industry, we need to build at least 2 factories: one must be in chemistry and pharmacy and one for genetically engineered drugs production (there are about 40 factories

of the type in the world). I believe that the only location for a factory for the production of genetically engineered drugs is Pushino. Pushino has everything needed for the development of a bio-pharmaceutical industry: 7 specialized research institutes in physical and chemical biology, two universities, and experimental research facilities with in-

Table 1

Company-producer	2008 y., roubles.	2008 y., roubles %
Novo Nordisk	3 402 384 126	47.39
Sanofi-Aventis	1 720 706 535	23.97
Elli Lilly	1 760 622 611	24.52
MARVEL LIFESCIENCE S PVT.Ltd	129 616 186	1.81
MM Shemiakin and YA Ovchinnikov'snstitute of Bio-organic Chemistry of RAS	74 424 660	1.04
BIOTON S.A.	25 986 679	0.36
Pharmstandart Llc	26 651 492	0.37
TONGHUA DONGBAO PHARMACEUTICAL CO.Ltd	2 695 064	0.04
others	27 136 365	0.38
M.J.BIOPHARM	9 081 318	0.13
Nazionalnye Biotehnologii Plc	40 667	0.00
"Brynzalov-A" Ltd	30 979	0.00
Total	7 179 376 682	

REFERENCE:

The Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry already has ready-for-introduction developments of human albumin, some blood factors and enzymes based on which it is possible to develop anti-viral and anti-tumour drugs. Also, scientists from the institute have developed a new analogue of insulin, which has a quicker and long-term effect. Already on the market are the first Russian-produced insulin and growth hormone.

ternational certificates. It is possible to produce insulin, interferon, blood factors, and other drugs there. There was a time when our chemical-pharmaceutics factories ranked 5th in the World behind the US, Germany, Japan, and France. Today we import around 85% of our drugs and we produce around 12% of the drugs we need from foreign raw materials. At the same time, the majority of clinical trials of western drugs take place in Russia. Yes, in the 1990s, when, for example, oncological centers gained access to western drugs in this way, it was possibly a reasonable policy. But why they use us only as a testing ground? It is not obvious that all clinical trials will be successful. And they conduct experiments on people.

– **So... What do we have to do if we want to wean ourselves off imported insulin?**

Firstly, our government should decide about building at least two fac-

tories, and state corporations such as Rosnano or Rostekhnologii should be responsible for the realisation of this plan. It is nonsense to believe that private capital will invest money into factory-building projects. However, it will be possible to sell this factory to a private Russian company after completion. Bioran can build this factory, and I deeply believe that Bioran should give 50% plus 1 share to the state or a state corporation.

Second, I believe that we need to talk not only about the development of the pharmaceutical industry, but also and even more about the development of dependent import-substitution. If RAMS will unveil the list of the 2,000 most wanted drugs for the cure of onco-, cardio-, and pulmonary diseases and others to scientists, the scientists will quickly find a way to avoid patenting problems! However, the next step – experimental production – will be re-

quired. One of the very few biotechnological manufacturing site in Russia is at our institute.

CONTERINDICATION TO IMPORT

Today almost 70% of the drugs produced in Russia are non-innovative drugs. Should we also produce raw materials or it would be easier to buy them? What do you think?

– To produce! Definitely to produce! At least, the control of production should be on the National level, – it is the security of the country. It is possible that Russia needs a big disaster to understand that we need our own pharmaceutical industry with a full cycle of production of at least the most important drugs. Imagine that one day the import of insulin from abroad is stopped – that means that the lives of the hundreds of thousands of diabetics will be in danger! It seems that that will never happen. But we already have lessons of that kind. We started to develop the 7-th blood factor, which is involved in blood clotting, only after the Beslan tragedy. Director of the Science Centre of Haematology of RAMS Andrei Ivanovich Vorobiev told me that we lost a lot of lives because of the absence of this drug. Russia imports it from Denmark, it is very expensive, and we buy it in small quantities; in case of an emergency, we simply do not have enough of it. The situation with insulin, as you see, is similar... ●


INNOVATIVE RUSSIA

discussion club

We form a dialogue between students, scholars, teachers, businessmen, managers, innovators, investors, designers, art historians, architects, photographers – all socially active people.

Find more
at WWW.STRF.RU

We invite anyone with something to say and
the desire to share their ideas



Tel.: +7 (495) 930-87-07, 930 8850
E-mail: seminar@strf.ru

Phage Display on the Base of Filamentous Bacteriophages: Application for Recombinant Antibodies Selection

N.V. Tikunova*, V.V. Morozova

Institute of Chemical Biology and Fundamental Medicine, Siberian Branch,
Russian Academy of Science

*E-mail: tikunova@niboch.nsc.ru

ABSTRACT The display of peptides and proteins on the surface of filamentous bacteriophage is a powerful methodology for selection of peptides and protein domains, including antibodies. An advantage of this methodology is the direct physical link between the phenotype and the genotype, as an analyzed polypeptide and its encoding DNA fragment exist in one phage particle. Development of phage display antibody libraries provides repertoires of phage particles exposing antibody fragments of great diversity. The biopanning procedure facilitates selection of antibodies with high affinity and specificity for almost any target. This review is an introduction to phage display methodology. It presents recombinant antibodies display in more details, construction of phage libraries of antibody fragments and different strategies for the biopanning procedure.

Keywords: phage display, filamentous bacteriophages, phagemids, phage display libraries of antibody fragments, biopanning, single chain antibody fragments, Fab-fragments.

INTRODUCTION

In the mid-eighties, a novel molecular-biological methodology which revolutionized the engineering of peptides and proteins was developed. This approach is known as phage display. It is based on the experiments of George Smith performed in the mid-80s [1]. Initially, Smith demonstrated that an exogenous protein can be expressed on the surface of the filamentous M13 phage. This was achieved by inserting the gene that encoded a part of the EcoRI endonuclease into the ORF of the phage's minor capsid protein pIII. Using polyclonal antibodies specific to the EcoRI endonuclease, Smith demonstrated the ability of phages carrying the chimeric EcoRI-pIII protein to specifically bind the appropriate antibodies. Furthermore, it was shown that phages with this insertion could be selected from a mixture containing wild-type phages by affine enrichment using polyclonal antibodies against the EcoRI endonuclease.

These experiments led to two important conclusions: first, using DNA-recombination methods, it is possible to create phage populations of different representativity ($10^6 - 10^{11}$ variants), wherein each individual phage displays a random peptide on its surface. Such populations were named "combinatorial phage libraries." Second, physical link between the analyzed polypeptide and the gene encoding it in the same phage particle provides the opportunity for easy selection of the needed variants and their identification.

G. Smith termed the result of expression of exogenous oligo- and polypeptides on the surface of viable filamentous phages "phage display." Furthermore, a method of affinity enrichment named "biopanning" was developed. According

to this method, phages bearing inserted sequences with affinity to specific ligands can be selected from a phage library. The term "biopanning" was suggested in 1988 [2].

The small number of pIII molecules in the phage particle (5 copies) limits the use of phage displays in selection of synthetic immunogens. Still, attempts to obtain phages exposing exogenous peptides as portions of the pVIII protein, which is present in 3,000 copies in each virion, were unsuccessful. Only the studies performed by Russian researchers managed to map a site on the N-terminus of pVIII that was exposed on the surface and was immunogenic but did not lead to significant disturbance of the filamentous phages' morphogenesis [3, 4].

In the 1990s, phage display was used in order to expose the antigen binding fragments of immunoglobulins on the surface of the fd phage [5]. This led to a novel combinatorial approach in the development of recombinant antibodies, which was an alternative to the traditional hybridoma technology. According to this approach, the phage system allows to replace all the stages after immunization of animals and spleen removal by simple manipulations with DNA and bacteria. In addition, it reduces the time needed to obtain stable antibody-producing clones from months to weeks. It also reduces the cost of the whole procedure.

Years of using phage display have led to several important areas of application:

Phage Display of Peptides

- The study of receptors and mapping of antibody binding sites
- The creation of immunogens and nanovaccines
- Mapping of substrate binding sites for proteases and kinases

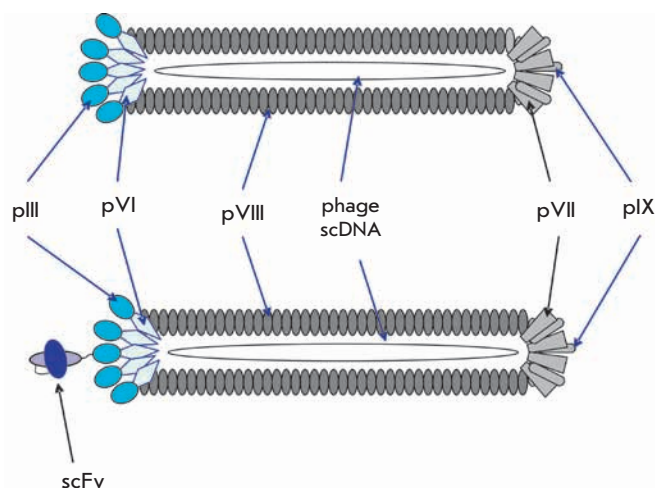


Fig. 1. A schematic representation of a filamentous phage: (a) wild-type filamentous phage; (b) filamentous phage-based phage antibody

Phage Display of Proteins and Protein domains

- Selection of antibodies with specific properties
- Study of protein-ligand interactions
- Screening of expressed cDNA fragments
- Directed evolution of proteins

This review describes the main principles and methods applied in phage display technology based on filamentous phages. Special attention is paid to the display of recombinant antibodies.

MORPHOLOGY AND LIFE CYCLE OF FILAMENTOUS BACTERIOPHAGES

Phage display methodology and its success are defined by the features of filamentous phages. Currently, several filamentous bacteriophages are known to infect gram-negative bacteria. The best characterized are the M13, f1, and fd phages, which infect *Escherichia coli* strains that carry an F-conjugative plasmid. The genomes of these phages have been sequenced and are 98 % homologous [6, 7]. Based on this homology and also on the dependence of infection on the presence of an F-plasmid, these phages are all termed Ff-phages.

An Ff-phage genome is a single-stranded covalently closed DNA, 6407(8) nucleotides in length, which encodes 11 genes. These genes are grouped in the genome according to their functions: the first group (genes II, V, X) encodes proteins needed for the replication of the phage DNA; the second group (genes III, VI, VII, VIII, IX) encodes surface-envelope proteins; and the third group (genes I, IV, XI) encodes proteins necessary for virion assembly. In addition, the phage DNA carries an intergenic region which contains an ori (origin of replication) site for synthesizing (+) and (-) DNA chains, as well as a site called the “packaging signal,” which initiates virion assembly.

Ff-phage DNA is enclosed in a flexible cylinder comprised of approximately 2,700 molecules of pVIII (Fig. 1). One end of the Ff-phage carries 5 copies of the minor surface proteins pIII and pVI; the other, pVII and pIX.

Infection of *E. coli* cells by the Ff-phage starts with a specific interaction between pIII and the top of the F-pilus, which is a protein tube made up of pilin subunits. Retraction of the pilus, caused by de-polymerization of pilin subunits, pulls the bacteriophage towards the cell [8]. After the phage DNA penetrates into the cytoplasm, it is turned into a replicative plasmid-like molecule (RF molecule) by the *E. coli* replication enzymes. This molecule then acts as a template for transcription and translation of phage proteins.

Phage protein production is enhanced as the number of RF-molecules increases, and after reaching a certain concentration of the pV protein, the newly synthesized ssDNA forms a DNA-pV complex, which is used in bacteriophage assembly. Assembly takes place in an area of tight contact between the cell wall and the cellular membrane and continues until the end of the phage DNA is freed and the phage leaves the cell. Assembly of the virions does not lead to cell lysis, and the infected cells continue to divide, although slower than uninfected cells [9].

FILAMENTOUS PHAGE-BASED VECTOR SYSTEMS

A number of novel vector molecules have been constructed on the base of filamentous phage DNA. These so-called “phagemids” combine the characteristics of plasmids and phages [10]. Phagemids contain a replication origin and the packaging signal of the Ff-phage, as well as a replication origin for the chosen plasmid, gene III, a polylinker, and an antibiotic-resistance gene [11].

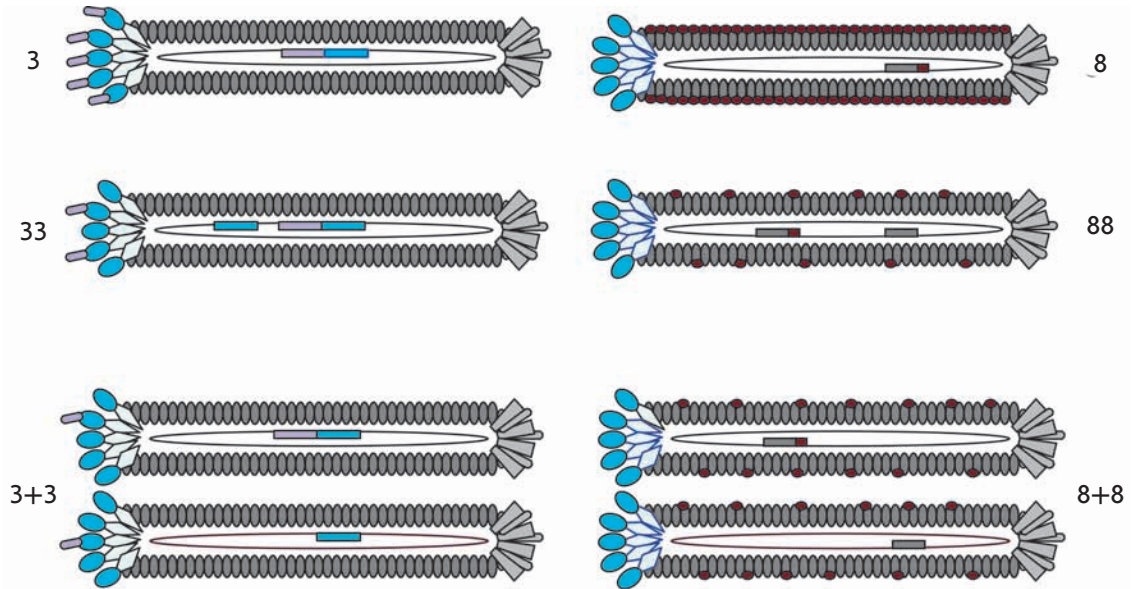
Several phage display systems were developed on the base of five capsid proteins [12], but systems based on the minor envelope protein pIII or the major envelope protein pVIII are mainly used (Fig. 2). These vector systems are termed system 3 and 8, respectively [10].

Depending on the size of the exogenous insert, the envelope protein of the phage may lose its normal virion assembly or cell infection, both of which lead to decreased viability of the phage. In order to restore infectivity of phages with inserts, specialized vector systems were developed using introduction of an additional gene encoding wild-type pIII or pVIII protein. The vector systems, termed 3+3 or 8+8, carry both the wild-type phage gene and the recombinant gene. The vector systems referred to as 3+3 and 8+8 have the recombinant gene on a phagemid, while the additional wild-type gene is introduced into the *E. coli* cells via a “helper” phage. In both cases, the replicated phages carry both normal and hybrid proteins and can replicate, despite the presence of exogenous inserts.

The 8+8 vector system exhibits multivalence: both wild-type and recombinant pVIII proteins are exposed on the surface of the phage particle, but since pVIII is a major envelope protein, only several hundred exogenous fragments can be exposed. Such a high valence is useful when low-affinity ligands need to be selected.

On the other hand, the main characteristic of the 3+3 vector system is its virtual monovalence: both recombinant and wild-type pIII proteins are exposed on the surface of the phage particle, and the number of recombinant protein copies varies from 0 to 5 for each single virion. Notably, only 10 % of the phages carry even one copy of the chimeric protein, and the percentage of phages carrying 2 or more molecules

Fig. 2. Types of vector systems for peptide and protein display based on filamentous phages. Lilac and red represent the exogenous inserts into the genes encoding the pIII and pVIII proteins, respectively



of recombinant pIII is considerably smaller. About 90 % of the phages carry no chimeric protein [13]. This low valence leads to limited avidity, which in turn allows the selection of high-affinity molecules, and the 3+3 system is mostly used for selection of antibody fragments.

MAIN STEPS OF A PHAGE DISPLAY PROTOCOL

In order to select for target antibodies with the desired characteristics, two important steps are required: first, an adequate phage antibody library should be used; and second, the right biopanning strategy should be chosen.

A combinatorial phage library of antibodies is a phage population in which each individual phage exposes a unique antigen-binding antibody domain on its surface as a part of a chimeric pIII protein. This antibody domain is most often a single chain-variable fragment (scFv) or Fab-fragment (Fig. 3). The structural and functional traits of these antibody fragments are reviewed in [14]. The creation of such a library involves the cloning of PCR-amplified DNA fragments encoding Fab or scFv from different sources into the ORF of the pIII protein (Fig. 4). This yields a repertoire of phagemids, and each phagemid includes a DNA fragment encoding an individual antigen binding domain. Transfection of *E.coli* cells with this phagemid repertoire provides a library of phages, and each phage exposes an individual combination of variable heavy- and light-chain domains on its surface. DNA ligation and bacterial transformation are the key steps, since they determine the size of the library [15].

Such libraries are characterized by the affinity level of the obtained antibodies, as well as the size of the library and its functional size. The affinity of the selected antibodies is primarily determined by the size of the library that is limited by the efficiency of *E.coli* transformation. The size of the library is the number of clones growing after transformation of *E.coli* with the whole phagemid population. A much more important factor is the functional size of the library, which is

the number of clones that carry correctly assembled genes without deletions, frame-shifts, or improper stop-codons. The functional size of the library is always smaller than the initial size, but it is the key parameter that determines the properties of the antibodies to be selected [16].

Selection of the target antibodies requires a biopanning procedure, an affine enrichment of the library by antibodies specific to the desired antigen (Fig. 5). This involves the incubation of an immobilized antigen with the phage library. The unbound phage antibodies are removed, and the bound ones are eluted and used for infecting *E.coli*, where they replicate. These phages are then extracted and used for the next round of biopanning. Ideally, one round of biopanning should be enough, but nonspecific binding limits the enrichment during each single round; so, several rounds of selection are usually required. Then, individual antibodies directed against a specific antigen are selected from the enriched library.

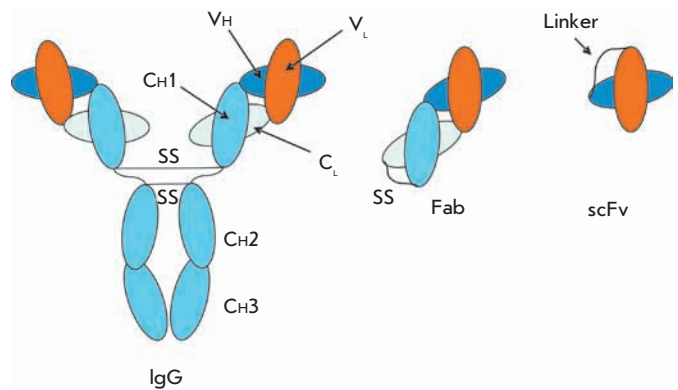


Fig. 3. Natural immunoglobulin G molecule and the antigen-binding fragments of the immunoglobulin

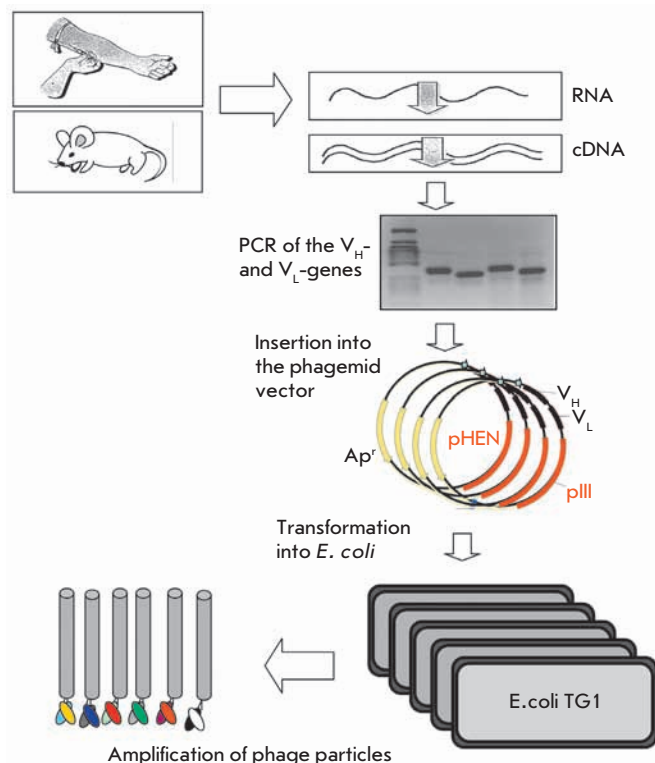


Fig. 4. A schematic representation of the construction of an antibody fragment phage library

TYPES OF PHAGE ANTIBODY LIBRARIES

There are two types of *in vitro*-made phage antibody libraries: the so-called “natural” and “synthetic” libraries, which are further subdivided according to the used gene repertoire.

Currently, a number of phage display antibody libraries have been fully or partially obtained from natural sources; thus mRNA was extracted from peripheral lymphocytes, bone marrow or the spleen [17, 18]. Libraries of human antibodies are of special interest, since the selected antibodies can be used for the development of therapeutics. Since most of the families encoding human V-genes are rarely used in the immune response, only the often-occurring gene families are usually used for library construction: V_{H1} – V_{H3} (from the heavy-chain gene family) and V_{K1} – V_{K4}, and also V_{λ1} – V_{λ3} (from the light-chain gene family). Both random recombination of the light- and heavy-antibody chains and the variability of the hyper-variable regions (complementarity determining region – CDR) of the V_H and V_L chains play an important role in the variability of the library.

Natural antibody libraries are divided into “immune” and “naïve” libraries. Immune libraries based on peripheral lymphocytes extracted from humans immunized by a certain antigen are valuable for medical research, since they provide increasing probability of selecting antibodies that could be used for therapy [19].

“Naïve” libraries based on lymphocyte mRNA extracted from unimmunized healthy people are used for obtaining antibodies directed against a wide multitude of antigens, including autoantigens [20]. These “naïve” libraries mostly represent the germline diversity of antibodies.

Synthetic libraries were introduced in order to increase the variety and size of the library and to enhance the antibody’s characteristics. Synthetic phage antibody libraries are divided into two groups:

- 1) Synthetic libraries based on a single-core V-gene;
- 2) Synthetic libraries based on a multiple-core V-genes.

Libraries of the first type are based on a single gene, which is then mutagenized in all the CDRs or just the CDRs in the V_H-domains. In this case, library diversity is limited by the degeneracy of the synthetic DNA which is used to change the sequence of the CDR-loops.

Synthetic libraries based on a single-core sequence present several advantages. They are easier to construct, and the obtained antibodies can be analyzed in less time. But the issue remains as to whether a single core sequence allows the correct folding of various antigen-binding CDRs providing a wide range of high-affinity antibodies.

Synthetic libraries of the second type are based on tens of genes, which are mutated in all the CDRs or in the CDRs in the variable domains of the heavy chains.

Different types of genes (Fab and scFv) can be used for creating both natural and synthetic phage libraries. Thus, combinatorial phage libraries of Fab- or scFv- antibody fragments can be obtained.

After MacCafferty and his collaborators demonstrated the possibility of creating a scFv library on the surface of a filamentous phage [5] in 1990, much research has been aimed at developing such libraries and studying the selected recombinant antibodies. In order to create a scFv library, populations of V_H- and V_L-genes are combined into a single DNA sequence by using an oligonucleotide encoding a flexible hydrophilic peptide (Fig. 3). This linker consists of glycine and serine residues (Gly₄ Ser)₃, which make it flexible and resistant to proteases. Then, the obtained scFv-genes are cloned into an appropriate vector (pHEN1, pHEN2, pSEX, etc.), which expresses the scFv-antibodies as part of the chimeric pIII protein. The heavy chain is most often immediately downstream from the leading sequence, while the light chain is fused to the N-terminal pIII sequence. The peptide linker facilitates the association between the V_H- and V_L domains necessary for forming the antigen binding surface, so disulfide bonds between the chains are not needed.

For further purification and characterization, it is convenient to produce the scFvs as secreted proteins. If pHEN or pSEX vectors are used, an amber stop codon inserted between the 3’-terminus of the scFv gene and the 5’-terminus of the pIII gene provides production of scFv. in nonsuppressor *E. coli* strains transformed by the phagemids [21].

Fab-fragments are heterodimers consisting of a light immunoglobulin chain (V_L-C_L) combined with a variable domain and the first constant domain of a heavy chain (V_H-C_{H1}). The chains interact after both have been synthesized, and this interaction stabilizes the formation of the antigen-binding site (Fig. 3). Creating Fab-libraries involves cloning a light chain and an Fd-domain (V_H + C_{H1}) into the appropriate vector. The

gene encoding the Fd-domain is combined with the region encoding the C-terminus of the pIII phage protein. Thus, the VI-gene and the Vh-gene, which is fused with the pIII phage protein, are transcribed as a polycistronic mRNA under the control of the chosen *E.coli* promoter. The N-terminus of each polypeptide consists of a leading sequence, which directs it towards the cellular membrane where virion assembly takes place.

A wider range of Fab-fragments can be obtained by transforming cells with plasmid bearing $V_H C_{H1}$ -chains and then transfecting the same cells with a phage population exposing $V_L C_L$ -chains. In this case, complete chain recombination must be assured, so that the chains are all present in the phage particle for the following selection procedure. In order to achieve the recombination of the H- and L-chains *in vivo*, both the plasmid and the phagemid contain loxP-sites for the phage recombinase P1, and the *E.coli* strain is modified to produce phage recombinase [22].

Fab-fragments are usually more stable after purification as compared to scFvs. They also have a longer elimination half-life and thus better pharmacokinetic and pharmacodynamic qualities [14]. In addition, Fab-fragments can be turned into full-sized immunoglobulins relatively easily by combining the C-terminus of the C_{H1} domain with the Fc-fragment.

LIBRARY CONSTRUCTION

Different sources of gene repertoire are used for the construction of phage display antibody libraries, depending on the library type. Naïve and immune libraries are constructed using naturally reorganized genes, which encode the variable immunoglobulin domains of healthy donors or donors immunized with a certain antigen, respectively (Fig. 4). mRNA from the antibody-producing lymphoid cell line is isolated for this purpose: peripheral blood lymphocytes are mainly used, but splenocytes [23, 24], tonsil cells, and B-lymphocytes from the bone marrow, have been used as well [25, 26]. A preliminary *in vitro* immunization of peripheral lymphocytes can also be used to construct an immune antibody library [27, 28].

cDNA is then synthesized on the base of isolated mRNA, and both oligo-dT and statistically devised hexanucleotides and specific primers can be used that yield cDNA copies of all the possible variable domain genes [29]. Also, one or several primers can be used to limit the range of amplified genes to one or several variable domain gene families or antibody isotypes [30]. Information from the DataBase Kabat [31] or VBASE can be used for primer development. The primer sequence usually includes restriction sites for cloning the PCR-products into the appropriate vectors.

Naïve library construction. Phage display library of human scFv was constructed by Marks and his colleagues [30]. Peripheral blood lymphocytes isolated from two volunteers were used as a gene source. The primers used for cDNA synthesis were complementary to the conservative regions of the genes encoding the κ - and λ -type light chains and the IgM and IgG heavy chains. The cDNAs were then cloned into the pHEN1 vector, yielding two libraries: $V_{H\mu} - V_L$ ($2.9 \cdot 10^7$ clones) and $V_{H\gamma} - V_L$ ($1.6 \cdot 10^8$ clones) [30]. Various protein-, hapten- [30], and autoantigen-specific antibodies were selected from these libraries [30, 32], although the affinity constants of the selected scFv's did not exceed $10^7 M^{-1}$. Still, high-affinity an-

tibodies can be extracted from relatively small-sized libraries. For instance, a scFv-library containing $4 \cdot 10^7$ clones yielded scFvs specific to steroid hormones, and their affinity constants were $10^7 - 10^8 M^{-1}$ [33]. Single-chain antibodies specific to the tumor necrosis factor ($K_a = 10^7 - 10^8 M^{-1}$) were selected from a naïve library of $2 \cdot 10^8$ of independent clones. In order to increase the representativity of this library, the protocol for obtaining scFv-genes was modified. In contrast to previous work [30], in which the scFv-genes were constructed by 3-component PCR (V_H - linker - V_L), a more efficient two-component reaction was used, involving a restriction site in the linker sequence [16].

Further experimental efforts were aimed at constructing the wider variety of antibody fragments, since antibodies with nanomolar affinity to any antigen can be selected from a large repertoire ($10^{10} - 10^{11}$). Such a library ($1.4 \cdot 10^{10}$ clones) was developed on the base of the genetic material extracted from the lymphoid cells (peripheral blood lymphocytes, tonsil cells, and bone marrow B-lymphocytes) of 43 unimmunized donors [25]. V_H - and V_L -gene repertoires were preliminarily cloned into pCantab 6 and pCantab 3His₆ vectors, respectively. The affinity of the antibodies obtained from this scFv library exceeded $10^9 M^{-1}$. It was known that antibodies with such affinities usually appeared after a secondary immune response. This library also yielded antibodies specific to cell surface antigens [34] and an antibody that could specifically bind carbohydrates [35]. Also, antibodies specific to various toxic agents were obtained from this library. This result demonstrated an advantage for large libraries when compared to the hybridoma technique, because toxic agents were impossible to use for immunization [10]. Naïve scFv libraries were also used to obtain scFv specific to neurotoxins [36], to apoptosis proteins [37], etc.

The creation of a large naïve library of Fab fragments involved a two-step cloning strategy [24]. First, amplified V_H -, $V_{L\kappa}$ - and $V_{L\lambda}$ -genes were cloned into vectors bearing the C_{H1} , $C_{L\kappa}$ - and $C_{L\lambda}$ -genes, respectively. After that, restriction fragments which were obtained from the initial $V_H C_{H1}$, $V_{L\kappa} C_{L\kappa}$, and $V_{L\lambda} C_{L\lambda}$ gene repertoires were used instead of the usual PCR products. This approach made the cloning procedure more effective. Similarly to the creation of scFv libraries, the oligonucleotides for the amplification of the variable domains were chosen to maximize the output of the gene repertoire. The resulting Fab-fragment library contained $3.7 \cdot 10^{10}$ independent clones, and the affinity of the antibodies selected for a wide range of antigens varied from $2.7 \cdot 10^7$ to $3.7 \cdot 10^8 M^{-1}$.

Synthetic library construction. Combinatorial libraries are characterized not only by size, which is determined by the number of clones, but also by representativity, which is the overall number of different $V_H - V_L$ combinations. One way to extend the representativity of a library is to increase the number of donors, and this way is particularly used for natural libraries. However, the antibody-encoding gene repertoire still does not exceed the variability of the lymphoid cell genes available for library construction. The potential antibody repertoire in the human organism is supposed to reach 10^{12} , but the antibody-producing cells represent only a portion of that amount at a moment [37, 38]. A second way to extend the representativity of a library is to replace natural CDRs with synthetic ones. In this case, it is possible to construct a combi-

natorial library including all the possible antigen-binding sequences of antibodies. Such phage display libraries are usually termed semi-synthetic or synthetic, depending on whether one or both variable domains contain synthetic CDRs. An ideal way to produce a repertoire that would encompass all the possible antibody sequences would be to chemically synthesize all 6 randomized CDRs and then connect them via various structural domains. But usually CDR3 of the heavy chain is randomized, since it is the most significant component of the antigen binding site [39].

The first semi-synthetic library was based on one heavy chain in which CDR3 was replaced with 10^{20} different sequences and one intact light chain [37]. This library contained $5 \cdot 10^7$ clones, and the affinity of selected antibodies varied from 10^7 to 10^8 M^{-1} .

Later, single-core semi-synthetic libraries were constructed using mutagenesis to change all CDRs in the V_H -domain [40]. The positions for mutagenesis were chosen according to the special variability of natural antibody sequences. These positions were randomized using the codons which are mostly found in natural antibodies that provided good solubility of selected antibodies.

Besides libraries based on a single-core sequence, some libraries based on a repertoire of V_H -genes have been constructed [41, 42]. Thus, the Nissim library was based on 50 V_H -genes encoding most of the human V-segments and random nucleotide sequences encoding the V_H CDR3. These sequences varied from 12 to 36 b.p.. This approach yielded 9 V_H -gene repertoires with different lengths of V_H CDR3 and, thus, 9 phagemid libraries. The size of the overall library was 10^8 independent clones, and this library yielded antibodies specific to a whole range of antigens [42].

Another commonly used library is the Griffin 1 library. It was constructed in a similar way to the previous library, based on 50 germline V_H -genes. In addition, 6 V_L -genes, corresponding to the 6 main subtypes of κ - and λ -chains, were used. The library contained 10^8 clones, and it yielded antibodies specific to the soluble CD4 receptor [43], human interleukin 6 [44], orthopoxviruses [45], and the Ebola virus nucleoprotein [46]. This library was also used to select catalytic antibodies [47] and other ones [48].

The affinity of the selected antibodies yielded by these semi-synthetic and synthetic libraries was not high due to several factors. It is known that V_H CDR3 is highly variable in length in natural antibodies and can consist of up to 24 amino acid residues. Creation of synthetic CDR3 sequences with varying lengths and completely random structures is inefficient. Moreover, the structural conformation of antibodies with synthetic CDRs could be incorrect (compared to natural conformations) precisely at the CDR loops [49]. In order to avoid this difficulty, limited randomization and introduction of flanking structures that enclose the completely random amino acid sequences were used [49]. The resulting library contained $3.6 \cdot 10^8$ clones, and scFv's specific to different antigens were selected from it, although the affinity of the selected antibodies was still low ($4 \cdot 10^5 - 10^7 \text{ M}^{-1}$).

Further steps for avoiding the above-mentioned limitations depend on the use of so-called "master-genes," which are the genes encoding variable domains with specific frame regions flanking the various randomized CDRs. The HuCAL

library was constructed on the base of V_H - and V_L -gene families, which are mostly used in the immune response. Hence, 7 V_H -genes and 7 V_L -genes yielded a basic collection of 49 different scFv genes. All the genes were synthesized to avoid the codons for amino acid residues that promote aggregation [50]. This library was constructed according to the position of key amino acid residues in CDRs and the frame regions, the length of the CDRs, and the level of their variability. The synthesis of oligonucleotides encoding CDR3 involved the use of cassette trinucleotide mutagenesis, which removed the TAG termination codon and uncommonly used codons. This library contained $2 \cdot 10^9$ clones, and antibody fragments were selected from this library for a wide range of antigens, including peptides, proteins, and whole cells. The affinity constants were near those of a secondary immune response (10^9 M^{-1}). 3D structures have been created for all the used consensus structures. That has allowed further study of the source for the variability of natural structural motifs of human antibodies, and also the correlations between antibody structure, its affinity, specificity, and the bound antigen class [50, 51].

The n-CoDeR™ library was constructed in a similar way [52]. A single master gene modified by introduction of different *in vivo*-formed CDRs have been used. Since the genes contained natural CDRs, an increased amount of correctly assembled and functional molecules was guaranteed. Moreover, computer analysis showed that the antibodies produced from the CoDeR™ library were less immunogenic. This library yielded antibodies that were specific to carbohydrates and human autoantigens, with an affinity exceeding 10^9 M^{-1} .

Construction of immune libraries. Despite the fact that many universal naïve synthetic and semi-synthetic libraries have been constructed in recent years, it has become clear that the main problem is not the creation of an enormous repertoire of antibody fragments, but conservation of this repertoire for extended periods of time, and keeping it without failure in antibody structures. Also, the larger the library, the more practically difficult it is to work with it. For instance, amplification of a library with 10^{10} clones takes tens of liters of culture medium. This is one of the main reasons why smaller libraries containing a narrowed repertoire of antibodies are preferable.

Immune libraries have two main features: they are enriched with antigen-specific antibodies, and the affinity of some of the antibodies has increased during the immune response. So, these libraries contain an extended amount of clones producing high-affinity antibodies specific to the antigen used for the immunization and that appeared as a result of a secondary immune response. It is assumed that immune libraries containing 10^6 clones can yield some antibody fragments that specifically bind the antigen used for immunization. In comparison, naïve libraries must contain at least 10^8 individual clones in order to possess the original variability of antibodies [53].

The first immune libraries were constructed against HIV [54], the respiratory syncytial virus [55], hepatitis B virus [56], herpes simplex virus, and cytomegalovirus [57, 58]. Libraries have also been constructed against human autoantigens [59–61] and against antigens that cause allergic reactions [62].

Libraries based on the lymphocyte genetic material isolated from patients were constructed to select antibodies against

specific tumor markers [63 – 68]. Immune libraries have also been made against the hepatitis A virus [19], chicken pox virus [69], orthopoxviruses [70], etc. Recently, virus-neutralizing scFv fragments specific to the flu virus H5N1 [26, 71] and virus-neutralizing Fab-fragments specific to the rabies virus [72] were extracted from immune libraries. Full-size human antibodies against orthopoxviruses [73, 74], neutralizing antibodies against hepatitis A [75], and neutralizing antibodies specific to the B glycoprotein of cytomegalovirus [76] have been made using antibody fragments isolated from immune libraries.

A serious advantage of immune libraries is the possibility to select high-affinity antibodies, which appear after viral infections or cancer, and antibodies specific to autoantigens, which are present in patients with autoimmune diseases. Analysis of these antibodies can help in the identification of the antigen epitopes responsible for the humoral immune response. Another advantage of immune libraries is that weakly immunogenic antigens can be used for antibody selection.

AFFINITY SELECTION OF ANTIBODIES FROM LIBRARIES

The next important step after constructing a library or choosing an available one is the enrichment of the original antibody repertoire with the antibodies specific to a target antigen (Fig. 5). This procedure is called “biopanning” or affinity enrichment. Below, we present several biopanning strategies.

Biopanning with immobilized antigens. Traditionally, biopanning has been performed using an antigen adsorbed on a plastic surface, for instance immunotubes (Maxisorb tubes; Nalge Nunc Intl., Naperville, IL) or enzyme immunoassay plates [30, 78]. Biopanning can also be performed using affinity chromatography, which involves the immobilization of the target antigen in a column [43, 79]. The column is then washed to remove unspecific antibodies, and the bound specific phage antibodies are eluted from the column and amplified in *E.coli* cells. BIAcore chip sensors can also be used as an antigen-binding medium for affinity selection [80]. Using this method for biopanning, it is important to keep the conformational stability of the antigen. Some phage antibodies selected against adsorbed antigens cannot bind the antigen in its native conformation. One method to avoid this problem is to use indirect antigen binding by incorporating antigen-specific antibodies [81].

Elution of the specifically bound antibodies is performed by using acid solutions, such as HCl or a glycine buffer [78, 82], or basic (alkaline) solutions, such as triethylamine [30]. It is very important to neutralize the phage antibody eluate right after elution, adjusting the pH to 7.2–7.4. Antibodies can also be eluted by cleaving a specially introduced site between the antibody and the pIII protein with a protease [83]. Moreover, antibodies can also be eluted by adding an excess of antigen, since this antigen will compete with the immobilized one binding the antibodies [16, 43].

Biopanning using antigen solution. Immobilizing the antigen on a solid surface often causes conformational changes of the antigen molecules. To avoid conformational changes, a set of methods are used to bind antigens to antibodies in a solution. The use of tagged soluble antigens allows precise quantification of the antigen concentration [84], and thus the use of minimal concentrations, which leads to selection of high

affinity antibodies. After incubating the antibodies with an antigen conjugated with biotin, phages bound to the tagged antigen are collected with avidin- or streptavidin-covered paramagnetic beads. Then, the specifically bound phages can be eluted from the antigen and characterized. A drawback of this method is the co-selection of anti-streptavidin antibodies. This problem can be solved by adding one more step; incubation of the antibodies with streptavidin-covered beads so as to remove any streptavidin-specific antibodies.

Biopanning on cells. Direct antibody selection against cell surface markers can be performed using either a cell monolayer or a cell suspension. The unbound phage antibodies are removed by washing the culture dishes in the case of a monolayer and by centrifugation in the case of a suspension. In order to optimize the selection of antigen-specific antibodies and minimize the selection of nonspecific antibodies, a negative selection procedure can be performed before, after or simultaneously with the positive selection [49]. Simultaneous positive and negative selection creates competitive conditions: a small number of antigen-positive (target) cells and an excess of antigen-negative adsorbing cells, which bind the unspecific antibodies from the phage library. In order to collect the cells with the bound phage antibodies, fluorescently labeled antibodies specific to another antigen on the target cell surface are added to the suspension. The cells are then collected by a FACS-sorter.

Using a similar method, anti-Rh(D) Fab-antibodies, which are very promising for clinical use, have been extracted from a Fab-fragment library [85]. Similar approaches can be used for identifying tumor-specific antigens and also as a highly productive protocol for selecting antibody fragments specific to conformation-dependent cell surface markers. A scFv-library was subjected to 3 rounds of positive selection using human melanoma cells and negative selection using human

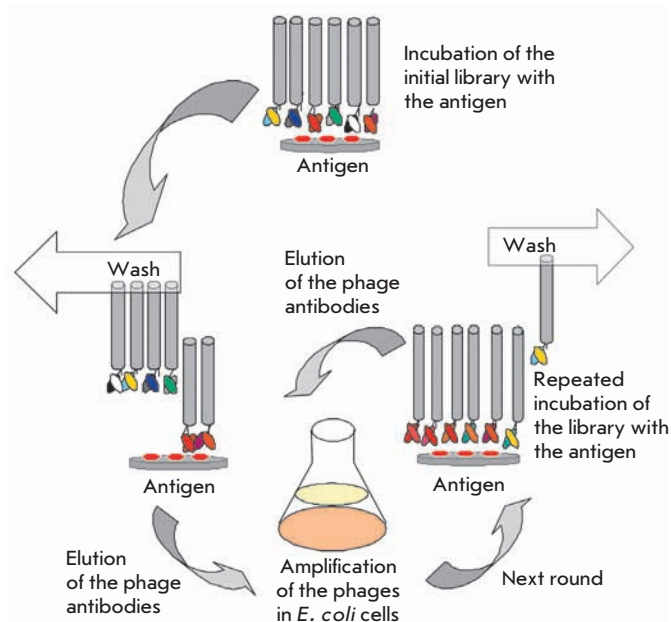


Fig. 5. A schematic representation of the biopanning procedure

peripheral blood mononuclear cells [64]. Two scFv, which could bind melanoma cells according to the enzyme immunoassay and FACS, were selected with this procedure [64]. Similar selection can be performed using fragments of various tissues.

A novel method has been developed for selecting antibodies which can penetrate eukaryotic cells via receptor-dependent endocytosis. In this case, selection is performed under conditions that stimulate active endocytosis. Thus, after lysis of the used cells, a population of phage antibodies which are able to penetrate eukaryotic cells can be obtained [86, 87]. It is suggested that such antibodies can act as vehicles for introducing various drugs into cells.

In vivo biopanning. This method involves a direct injection of the antibody repertoire into an animal's organism. Tissues are then removed and phage antibodies bound to tissue-specific cell markers can be extracted. A similar approach was used for peptide libraries [88]. *In vivo* biopanning has several advantages: first, the extracted antibodies specifically bind intact cell targets; second, antibodies binding nontarget cell surface proteins and blood plasma are immediately eliminated. Antibodies obtained via *in vivo* biopanning or cell biopanning can be very useful for the functional analysis of newly discovered receptors and in the search for potential targets of novel drugs.

Notably, most of the existing selection procedures involve only one or several antigens. However, there are biopanning strategies that allow parallel selection for antibodies against a population of antigens; the so-called "2D display." Thus, it has been proposed to incorporate multiplex flow cytometry for high-output simultaneous selection of individual antibodies against a wide range of antigens present on the surface of cells [89]. An alternative method of selection against a population of antigens is the combinatorial selection of antigen-antibody pairs which are able to replicate. Such an approach was used for selecting a scFv phage library for antibodies specific to antigens from a yeast library [90].

PROBLEMS AND SUCCESSES OF DISPLAYS BASED ON FILAMENTOUS PHAGES

Many researchers have faced a number of problems while constructing and using libraries based on filamentous phages. For instance, the low concentration of the original antigen-specific lymphocytes in the population can be a problem while constructing immune libraries. This and similar problems can be avoided by preliminary enrichment of the B-lymphocyte population, using antigen molecules conjugated with magnetic beads [91, 92].

Another difficulty is selective loss of high-affinity antibodies during biopanning cycles. Another problem is the

decrease of the share of phages with full-sized inserts during library amplification and propagation. Analysis of the antibody-encoding gene's size showed that the number of clones with defective inserts increased after even one round of biopanning [16]. Phages with smaller inserts (usually with an absent V_H region) tend to grow more quickly than phages with full-sized inserts, which leads to their more effective amplification. That is why using the appropriate biopanning strategy is so important. Varying the elution and screening conditions during selection can help to select high-affinity antibodies.

Another problem is the big difference between the theoretically possible and the actual variety provided by the phagemid population. This difference cannot be fully explained by cell transformation efficiency. It can also be caused by cell toxicity of the antibodies, incorrect protein folding or assembly, competition between the chimeric and wild-type pIII proteins, proteolysis of the antigen-binding site on the surface of the phage, and several other factors.

Nevertheless, despite the existing methodological problems, phage display techniques are showing considerable progress. This technology is a very promising instrument for several reasons. Human antibody fragments selected from phage display libraries could be useful in the construction of fully human antibodies that have lower immunogenicity than chimeric or humanized antibodies and would be more valuable for therapeutic use. Antibody fragment libraries allow the rapid isolation of specific antigen-binding domains without any of the limitations caused by the availability of donor lymphocytes or cell-fusion difficulties. Another advantage of phage display libraries is that one library can be used for multiple screening against different antigens. One more advantage of this technology is the absence of laboratory animal use, the relative simplicity of the protocols, and the possibility of screening a large amount of candidate molecules in a short period of time. A very important feature is the possibility of selecting antibodies against a wide range of antigens, including toxic substances or highly dangerous viruses, which cannot be used for immunization because of ethical reasons. Phage display technology has proved its effectiveness in the development of therapeutic antibodies. In recent years, more than 14 antibodies on the U.S. pharmaceutical market have been obtained using phage display methods [15]. All these facts explain the interest of large companies, such as Morphosys GmbH (Germany; <http://www.morphosys.com>); Cambridge Antibody Technology (United Kingdom; <http://www.catplc.co.uk>), Dyax (USA; <http://www.dyax.com>) and others, in the development and application of antibody fragment phage display libraries. ●

REFERENCES

1. Smith G. // *Science*. 1985. V. 228. P. 1315–1317.
2. Parmley S., Smith G. // *Gene*. 1988. V. 73. P. 305–318.
3. Ilyichev A., Minenkova O., Tat'kov S., et al. // *Doklady AN USSR (rus)*. 1989. V. 307. P. 481–483.
4. Minenkova O., Ilyichev A., Kishchenko G. // *Gene*. 1993. V. 128. P. 85–88.
5. MacCafferty J., Griffiths A., Winter G., et al. // *Nature*. 1990. V. 348. P. 552–554.
6. Van Wezenbeek P., Huloenmakers J.G. // *Gene*. 1980. V. 11. P. 129–148.
7. Beck E., Zink B. // *Gene*. 1981. V. 16. P. 35–58.
8. Maneewannakul K., Maneewannakul S., Ippen-Ihler K. // *J. Bacteriol.* 1993. V. 175. P. 1384–1391.
9. Model P., Russel M. *The bacteriophages*. V. 2. N.Y.: Plenum, 1988. 456 p.
10. Kay B., Winter G., McCafferty J. *Phage display of peptides and proteins: a laboratory manual*. N.Y.: Academic press. 1996. 306 p.
11. Mead D., Kemper B. // *Biotechnology*. 1988. V. 10. P. 85–102.
12. Mullen L., Nair S., Ward J., et al. // *TRENDS in Microbiol.* 2006. V. 14. P. 141–147.
13. Griffiths A., Duncan A. // *Curr. Opin. Biotech.* 1998. V. 9. P. 102–108.
14. Deev S., Lebedenko E., // *Acta Naturae*. 2009. V.1. P. 32–50.

REVIEWS

15. Azzazy H, Highsmith W. // *Clin. Biochem.* 2002. V. 35. P. 425–445.
16. Batanova T, Ulitin A, Morozova V, et al. // *Mol. Genet. Microbiol. Virusol. (rus)*. 2006. V. 3. P. 35–41.
17. Blazek D, Celer V. // *Folia Microbiol.* 2003. V. 48. P. 687–698.
18. Chingwei V, Blazek D, Celer V, et al. // *J. Virol. Methods.* 2004. V. 115. P. 83–82.
19. Kim S, Jang M, Stapleton J, et al. // *Virology.* 2004. V. 318. P. 598–607.
20. Nagano K, Imai S, Mukai Y, et al. // *Pharmazie.* 2009. V. 64. P. 238–241.
21. Barbas C. // *Curr. Opin. Biotechnol.* 1993. V. 4. P. 526–530.
22. Zahra D, Vancov T, Dunn J, et al. // *Gene.* 1999. V. 227. P. 49–54.
23. Sheets M, Amersdorfer P, Finner R, et al. // *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. P. 6157–6162.
24. de Haard H, van Neer N, Reurs A, et al. // *J. Biol. Chem.* 1999. V. 274. P. 18218–18230.
25. Vaughan T, Williams A, Pritchard K, et al. // *Nat. Biotechnol.* 1996. V. 14. P. 309–314.
26. Kashyap A, Steel J, Oner A, et al. // *Proc Natl Acad Sci USA.* 2008. V. 105. P. 5986–5991.
27. Yamashita M, Katakura Y, Shim S, et al. // *Cytotechnology.* 2002. V. 40. P. 161–165.
28. Yamashita M, Katakura Y, Shirahata S. // *Cytotechnology.* 2007. V. 55. P. 55–60.
29. Ulitin A, Kapralova M, Laman A, et al. // *Doklady AN (rus)*. 2005. V. 405. P. 1–4.
30. Marks J, Hoogenboom H, Bonnert T, et al. // *J. Mol. Biol.* 1991. V. 222. P. 581–597.
31. Johnson G, Wu T. // *Nucleic Acids Res.* 2001. V. 29. P. 205–206.
32. Griffiths A, Malmqvist M, Marks J, et al. // *EMBO J.* 1993. V. 12. P. 725–734.
33. Dorsam H, Rohrbach P, Kurscher T, et al. // *FEBS Lett.* 1997. V. 1. P. 7–13.
34. Hoogenboom H, Lutgerink J, Pelsers M, et al. // *Eur. J. Biochem.* 1999. V. 260. P. 774–784.
35. Foy B, Killen G, Frohn R, et al. // *J. Immunol. Methods.* 2002. V. 261. P. 73–83.
36. Cardoso D, Nato F, England P, et al. // *Scand. J Immunol.* 2000. V. 51. P. 337–344.
37. Barbas C, Bain J, Hoekstra D, et al. // *Proc. Natl. Acad. Sci. USA.* 1992. V. 89. P. 4445–4457.
38. Winter G, Griffiths A, Hawkins R, et al. // *Annu. Rev. Immunol.* 1994. V. 12. P. 433–455.
39. Xu J, Davis M. // *Immunity.* 2000. V. 13. P. 37–45.
40. Lee C, Sidhu S, Fuh G. // *J. Immunol. Meth.* 2004. V. 284. P. 119–132.
41. Tomlinson I, Walter G, Marks J, et al. // *J. Mol. Biol.* 1992. V. 227. P. 776–798.
42. Nissim A, Hoogenboom H, Tomlinson I, et al. // *EMBO J.* 1994. V. 13. P. 692–698.
43. Griffiths A, Williams S, Hartley O, et al. // *EMBO J.* 1994. V. 13. P. 3245–3260.
44. Krebs B, Griffin H, Winter G, et al. // *J Biol Chem.* 1998. V. 273. P. 2858–2865.
45. Tikunova N, Morozova V, Batanova T, et al. // *Hum. Antibodies.* 2001. V. 10. P. 95–99.
46. Tikunova N, Batanova T, Chepurinov A. // *Vopr. Virusol. (rus)* 2005. V. 50. P. 25–29.
47. Reshetnyak A, Armentano M, Ponomarenko N, et al. // *J Am. Chem. Soc.* 2007. V. 129. P. 16175–16182.
48. Morozova V, Tikunova N, Batanova T, et al. // *Vestnik RAMN (rus)*. 2004. V. 8. P. 22–27.
49. de Kruif J, Boel E, Logtenberg T. // *J. Mol. Biol.* 1995. V. 248. P. 97–105.
50. Knappik A, Ge L, Honegger A, et al. // *J. Mol. Biol.* 2000. V. 296. P. 57–86.
51. Rothe C, Urlinger S. // *J. Mol. Biol.* 2008. V. 376. P. 1182–1200.
52. Carlsson R, Soderlind E. // *Expert. Rev. Mol. Diagn.* 2001. V. 1. P. 102–108.
53. Gavilondo J, Larrick J. // *Biotechniques.* 2000. V. 29. P. 128–138.
54. Burton D, Barbas C, Persson M, et al. // *Proc. Natl. Acad. Sci. USA.* 1991. V. 88. P. 10134–10137.
55. Barbas C, Crowe J, Cababa D, et al. // *Proc. Natl. Acad. Sci. USA.* 1992. V. 89. P. 10164–10168.
56. Zebedee S, Barbas C, Hom Y, et al. // *Proc. Natl. Acad. Sci. USA.* 1992. V. 89. P. 3175–3179.
57. Williamson R, Burioni R, Sanna P, et al. // *Proc. Natl. Acad. Sci. USA.* 1993. V. 90. P. 4141–4145.
58. Burioni R, Williamson R, Sanna P, et al. // *Proc. Natl. Acad. Sci. USA.* 1994. V. 91. P. 355–359.
59. Welschof M, Terness P, Kolbinger F, et al. // *J. Immunol. Methods.* 1995. V. 179. P. 203–214.
60. Chapal N, Chardes T, Bresson D, et al. // *Endocrinology.* 2001. V. 142. P. 4740–4750.
61. Raats J, Roeffen W, Litjens S, et al. // *Eur. J. Celliol.* 2003. V. 82. P. 131–141.
62. Padavattan S, Flicker S, Scrimmer T, et al. // *J Immunol.* 2009. V. 182. P. 2141–2151.
63. Mao S, Gao C, Lo C, et al. // *Proc. Natl. Acad. Sci. USA.* 1999. V. 96. P. 6953–6958.
64. Kupsch J, Tidman N, Kang N, et al. // *Clin. Cancer Res.* 1999. V. 5. P. 925–931.
65. Li J, Pereira S, Van Belle P, et al. // *J. Immunol.* 2001. V. 166. P. 432–438.
66. Schmidt A, Muller D, Mersmann M, et al. // *Eur. J. Biochem.* 2001. V. 268. P. 1730–1738.
67. Dantas-Barbosa C, Brigido M, Maranhao A. // *Genet. Mol. Res.* 2005. V. 4. P. 126–140.
68. Figini M, Obici L, Mezzanzanica D, et al. // *Cancer Immunol. Immunother.* 2009. V. 58. P. 531–546.
69. Kausmally L, Waalen K, Lobersli I, et al. // *J. Gen. Virol.* 2004. V. 85. P. 3493–3500.
70. Dubrovskaya V, Ulitin A, Laman A, et al. // *Mol. Biol. (rus)*. 2007. V. 41. P. 173–185.
71. Throsby M, van der Brink E, Jongeneelen M, et al. // *PLoS One.* 2008. V. 3. P. 3942.
72. Houimel M, Dellagi K. // *J Virol. Methods.* 2009. V. 161. P. 205–215.
73. Schmaljohn C, Cui Y, Kerby S, et al. // *Virology.* 1999. V. 258. P. 189–200.
74. Yun T, Tikunova N, Shingarova L, et al. // *Doklady RAN (rus)*. 2006. V. 407. P. 98–101.
75. Cao J, Meng S, Li C, et al. // *J. Med. Virol.* 2008. V. 80. P. 1171–1180.
76. Ohta A, Fujita A, Murayama T, et al. // *Microbes Infect.* 2009. V. 11. P. 1029–1036.
77. Cai X, Garen A. // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 6537–6541.
78. Kang A, Barbas C, Janda K, et al. // *Proc. Natl. Acad. Sci. USA.* 1991. V. 88. P. 4363–4366.
79. Garrard L, Henner D. // *Gene.* 1993. V. 128. P. 103–109.
80. Malmborg A, Duenas M, Ohlin M, et al. // *J. Immunol. Methods.* 1996. V. 198. P. 51–57.
81. Sanna P, Williamson R, De Logu A, et al. // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 6439–6443.
82. Roberts B, Markland W, Siranosian K, et al. // *Gene.* 1992. V. 121. P. 9–15.
83. Ward R.L., Clark M.A., Lee J, et al. // *J. Immunol. Methods.* 1996. V. 189. P. 73–82.
84. Hawkins R, Russell S, Winter G. // *J. Mol. Biol.* 1992. V. 226. P. 889–896.
85. Siegel D, Chang T, Russell S, et al. // *J. Immunol. Methods.* 1997. V. 206. P. 73–85.
86. Nielsen U, Marks J. // *Pharm. Sci. Technol. Today.* 2000. V. 3. P. 282–291.
87. Poul M. // *Meth. Mol. Biol.* 2009. V. 562. P. 155–163.
88. Pasqualini R, Ruoslahti E. // *Nature.* 1996. V. 380. P. 364–366.
89. Ayriss J, Valero R, Bradbury A., et al. // *Methods mol. Biol.* 2009. V. 525. P. 241–260.
90. Bowley D, Jones T, Burton D, et al. // *Proc. Natl. Acad. Sci. USA.* 2009. V. 106. P. 1380–1385.
91. Koefoed K, Farneas L, Wang M., et al. // *J. Immunol. Methods.* 2005. V. 297. P. 187–220.
92. Ditzel H. // *Methods Mol. Biol.* 2009. V. 562. P. 37–43.

Chemical and Functional Aspects of Posttranslational Modification of Proteins

D.G. Knorre, N.V. Kudryashova, T.S. Godovikova*

Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, Russian Academy of Sciences

*E-mail: godov@niboch.nsc.ru

ABSTRACT This paper reviews the chemical and functional aspects of the posttranslational modifications of proteins, which are achieved by the addition of various groups to the side chain of the amino acid residue backbone of proteins. It describes the main prosthetic groups and the interaction of these groups and the apoenzyme in the process of catalysis, using pyridoxal catalysis as an example. Much attention is paid to the role of posttranslational modification of proteins in the regulation of biochemical processes in live organisms, and especially to the role of protein kinases and their respective phosphotases. Methylation and acetylation reactions and their role in the "histone code," which regulates genome expression on the transcription level, are also reviewed. This paper also describes the modification of proteins by large hydrophobic residues and their role in the function of membrane-associated proteins. Much attention is paid to the glycosylation of proteins, which leads to the formation of glycoproteins. We also describe the main non-enzymatic protein modifications such as glycation, homocysteination, and desamidation of amide residues in dibasic acids.

Keywords: proteins, enzymes, posttranslational modification, prosthetic groups, phosphorylation, regulation, signal transduction, acylation, alkylation, ubiquitination, histone code, non-fermentative modification.

Abbreviations: CoA – coenzyme A, EGFR – epidermal growth factor receptor, JNK – Jun N-terminal kinase, SAPK – stress activated protein kinase, MAPK – mitogen-activated protein kinase, IP – inositoltriphosphate, DAG – diacylglycerol, JAK – Janus kinase, STAT – signal transducer and activator of transcription, Fyn, Lck – non-receptor tyrosinekinases of the Src family, Ub – ubiquitin residue, ULP – ubiquitin-like protein, Ras, Rab, Rho – protein products of the protooncogenes *ras*, *rab*, *rho*, which play a role in cell growth and differentiation, SAM – S-adenosylmethionine, PARP – poly(ADP-ribose)polymerase, VRAP – telomerase, found to be a part of vault particles, GSH – glutathione, HIF – hypoxia inducible factor, Gla – γ -carboxyglutamic acid, AGE – advanced glycation end products, CML – N_c-carboxymethyl-lysine, CEL – N_c-carboxyethyl-lysine, HSA – human serum albumin, GFP – green fluorescent protein, PIMT – protein isospartyl-O-methyltransferase, DNT – dermonecrotic toxin.

INTRODUCTION

Template biosynthesis of polypeptide chains on ribosomes most often does not immediately produce a fully functional protein. The newly formed polypeptide chain must undergo certain chemical modifications outside the ribosome. These modifications are most often driven by enzymes and take place after all the information supplied by the template RNA (mRNA) has been read, that is after mRNA translation: thus, these additional processes are called posttranslational modifications.

Posttranslational protein modification processes can be divided into two main groups. The first group unites proteolytic processes, which are mainly cleavages of certain peptide bonds, resulting in the removal of some of the formed polypeptide fragments. The second group consists of the processes that modify the side chains of the amino acid residues and usually do not interfere with the polypeptide backbone. The chemical nature and function of these modifications is diverse. Moreover, each type of modification is characteristic of certain groups of amino acid residues. The result of these processes is that the proteome of the cell or organism consists of several orders more components than there are genes encoding these components of the proteome. This paper is a review of the second group of posttranslational protein modifications.

There are four main groups of protein functions that require posttranslational modification of amino acid residue

side chains. The functional activity of a wide number of proteins requires the presence of certain prosthetic groups covalently bound to the polypeptide chain. These are most often complex organic molecules which take a direct part in the protein's activity. The transformation of inactive apoproteins into enzymes is one of these modifications. Another important group of posttranslational modifications regulates biochemical processes by varying (sometimes switching on and off) enzymatic activity. Another large group of modifications are protein tags, which provide intracellular localization of proteins, including marking the proteins for transport to the proteasome, where they will be hydrolysed and proteolysed. And finally, some posttranslational modifications directly or indirectly influence the spatial structure of newly synthesized proteins.

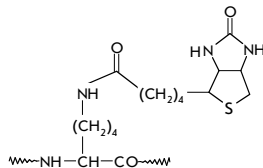
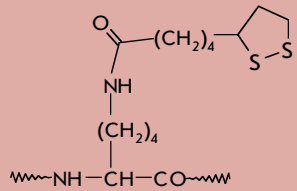
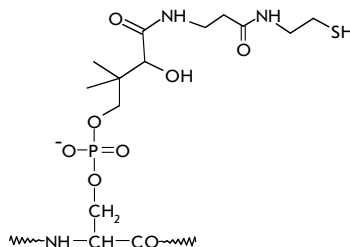
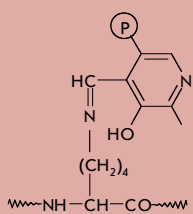
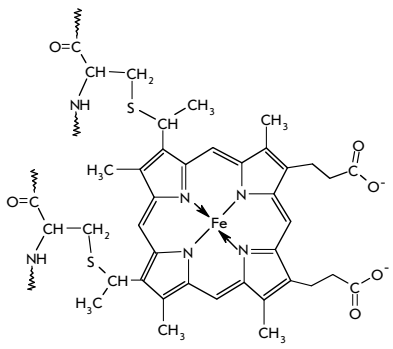
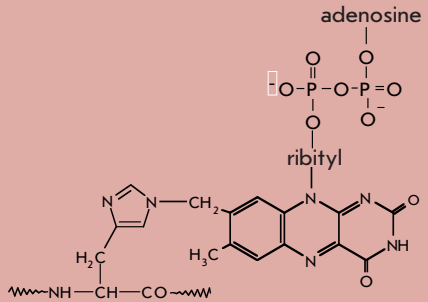
MODIFICATION OF PROTEINS BY ADDITION OF PROSTHETIC GROUPS

In some cases, the last step in the biosynthesis of a functional protein is the covalent binding of a prosthetic group, which forms part of the active site [1, 2]. Table 1 shows the structural formulas of side chain modification products after the covalent binding of certain cofactors to proteins, as well as the types of reactions in which the corresponding prosthetic groups take part.

Most of the listed prosthetic groups remain covalently bound to the apoenzyme through the whole catalytic process.

REVIEWS

Table 1. The main prosthetic groups involved in biocatalytic reactions

Coenzyme name	Structure of prosthetic group derivative	Classes of enzymes. Type of reaction, which involves the prosthetic group
Biotin		Carboxylases. E.C. 6.4.1.2; 6.4.1.3. Carboxylation. Transfer of a single carbon fragment (CO ₂) onto acetyl-CoA, propionyl-CoA, and other organic molecules
Lipoate		Acytransferases. E.C. 2.3.1.12. Reduction-oxidation. Transfer of carbon fragments onto CoA via reductive acylation of lipoamide during oxidative decarboxylation of α-ketoacids.
Panthotenate		Acytransferases. E.C. 2.3.1.85. Transacylation. Transfer of an acyl fragment from one enzyme of a multi-enzyme complex to another.
Pyridoxal phosphate		Aminotransferases. E.C.2.6.1. Transamination of amino acids.
Heme		cytochrome c oxidase. E.C. 1.9.3.1. Reduction-oxidation. Transfer of electrons on the mitochondrial membrane during oxidative phosphorylation.
FAD		Oxidoreductases. E.C. 1.3.99.1. Reduction-oxidation. Oxidation of the -CH ₂ -CH ₂ - group down to <i>trans</i> -CH=CH-

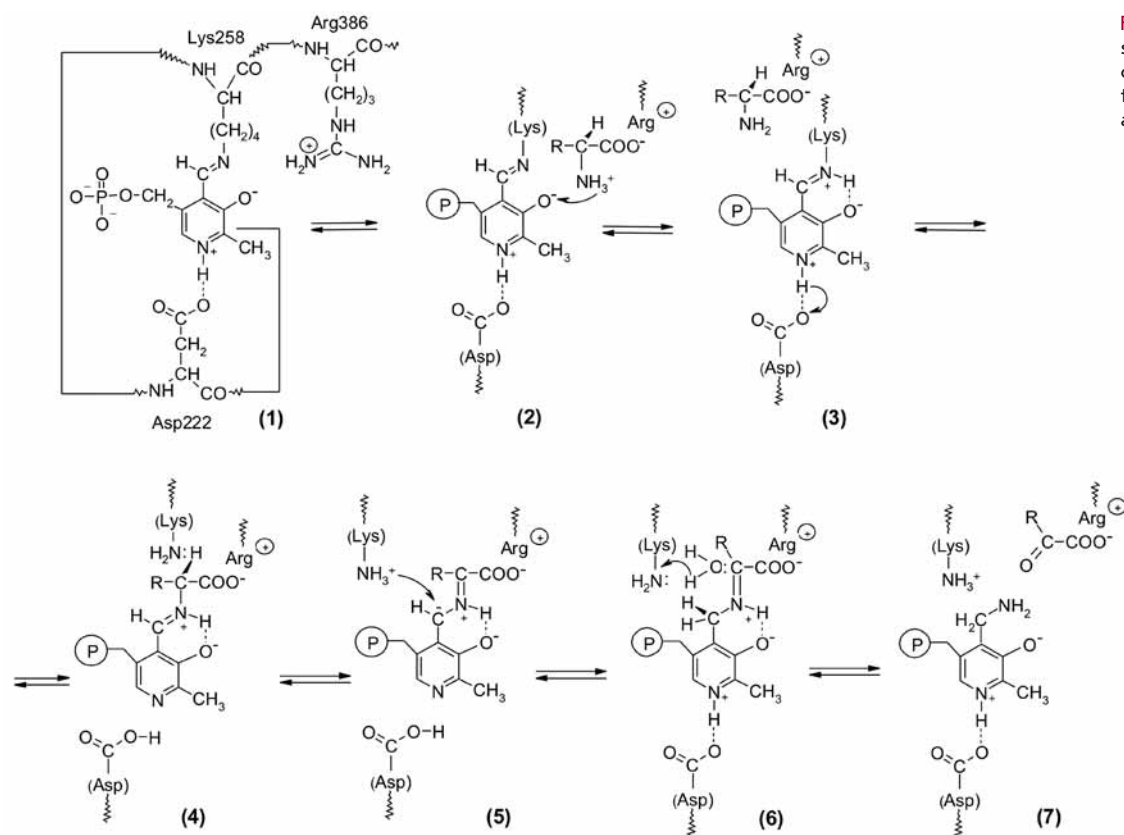


Fig. 1. A schematic representation of the first stage of the transamination reaction catalyzed by aspartate aminotransferase

The only exceptions are pyridoxal enzymes, which experience a demodification of the protein during catalysis; namely, a conversion of the bond between pyridoxal phosphate and the lysine amino group of the apoenzyme into the bond between the coenzyme and the substrate amino acid. A dynamic model of the reaction processes, catalyzed by transaminases, was suggested by M.Y. Karpeisky and V.I. Ivanov in 1969 [3]. Later [4], the authors suggested that the phosphate and methyl groups of the coenzyme act as a sort of axis around which pyridoxal can rotate, thus forming either enzyme-imine or substrate-imine covalent compounds. X-ray analysis data confirmed and detailed the conclusion about pyridoxal phosphate multi-point binding.

Aspartate aminotransferase (C.E. 2.6.1.1), which catalyses the transamination of oxalacetate and glutamate, can be used to illustrate the mechanism of action of pyridoxal enzymes (Fig. 1).

The coenzyme of the transaminase is not present as a free aldehyde, rather it is an intramolecular aldimine with the lysine side chain amino-group (Lys-258). The enzyme-bound imine assures the high rate of the reaction, as compared to the free pyridoxal phosphate [2–4]. It is this structure that causes the higher activity of imines as compared to aldehydes. The more basic nitrogen of imines is protonated much more efficiently than the carbonyl group oxygen atom (Fig. 1, (3)). The resulting transfer of the proton from the α -NH₃⁺-group of the substrate to the atom of N-aldimine pyridoxal phosphate creates the required cationic form of the coenzyme and, simulta-

neously, a deprotonated amino acid (3). Moreover, the imine carbon is more electrophilic than the carbonyl one, which means that it is more easily attacked by the deprotonated amino group of the α -amino acid (Fig. 1, (4)). An increase of the electrophilicity of this site is also achieved by the interaction of heterocycle nitrogen with an aspartate residue of the enzyme (hydrogen bond with Asp-222). Thus, the transitional imine-enzyme promotes the rapid formation of a transient bond between the substrate and the coenzyme.

The described example of pyridoxal catalysis illustrates the fact that the apoenzyme plays as important a role in catalysis as the prosthetic group; that is, the former cannot simply be called a carrier of the catalytic group. This is also the case for other prosthetic groups.

REGULATION OF ENZYME ACTIVITY BY PHOSPHORYLATION

The central role in reactions responsible for rearrangement of all intracellular processes eventually signaling either cell division or cell death is played by a large group of enzymes called protein kinases (phosphotransferases, EC 2.7). These enzymes can add phosphate groups to the side chains of amino acids in various proteins [5–12]. γ -phosphate ATP is the donor of a phosphate group in all such reactions. Kinases are grouped according to the amino acid to which they add the phosphate into tyrosine kinases (E.C. 2.7.10.2) and serine/threonine kinases (E.C. 2.7.11.1) [5]. Also, histidine kinases are often found in bacteria, plants, and fungi. The latter enzymes function in a two-step signal transduction system [13]. The inorganic

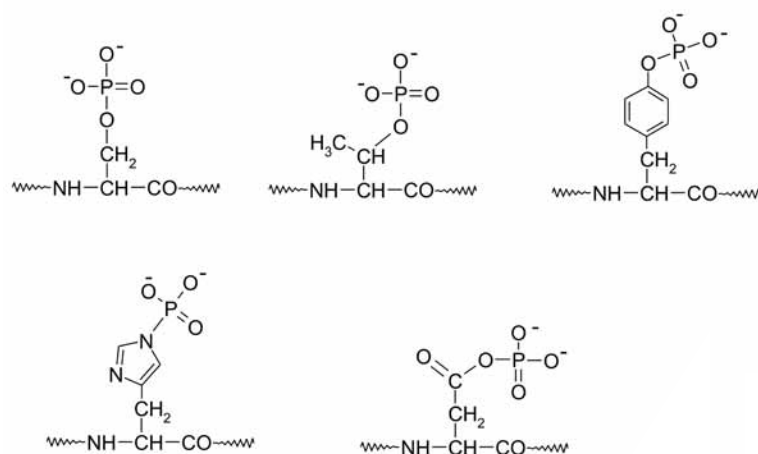


Fig. 2. The structure of phosphorylated amino acid fragments

phosphate residue, which is attached to a histidine in the enzyme itself, is then transferred onto an aspartate residue in the target protein. Phosphorylation of the aspartate results in further signal transduction [13]. Figure 2 shows the structures of amino acid phosphorylation products in proteins [1].

Concerted regulation of interactions in a multicellular organism is achieved by the release of specialized molecules (hormones, cytokines, etc.) which activate a signaling cascade in target cells. In cases where the signal causes alterations in the expression level of certain genes, the final links in the signaling chains are transcription factors [14–18]. Target cells can identify the signaling molecule amongst a multitude of others with the help of a receptor protein present on the target cell. This protein receptor has a specific binding site for the appropriate signaling molecule. Some receptors are localized on the surface of the cellular membrane, while others are intracellular receptors and are localized in the cytoplasm or inside the nucleus. A schematic representation of the main stages of, for example, hormone signal transduction via membrane receptors is presented in Fig. 3. At some of these stages, the activity of enzymes is regulated by phosphorylation.

Membrane receptors can be divided into three functionally distinct structural regions. The first domain (the recognition domain) is situated in the N-terminal region of the polypeptide chain and is located on the outside of the cellular membrane. This region carries glycosylated sites and recognizes and binds the signaling molecule. The second domain is the transmembrane domain. In some receptors, which are coupled to G-proteins, this domain consists of 7 tightly packed α -helix polypeptide regions. Another type of receptor has a transmembrane domain that consists of a single α -helix region. The third (cytoplasmic) domain creates a chemical signal inside the cell, which couples the binding of a signal molecule (a ligand) to a specific intracellular signal.

The cytoplasmic regions of a number of receptors which face onto the inner side of the membrane exhibit tyrosine kinase activity. For instance, the binding of the insulin hormone to its membrane receptor, which is a tyrosine kinase

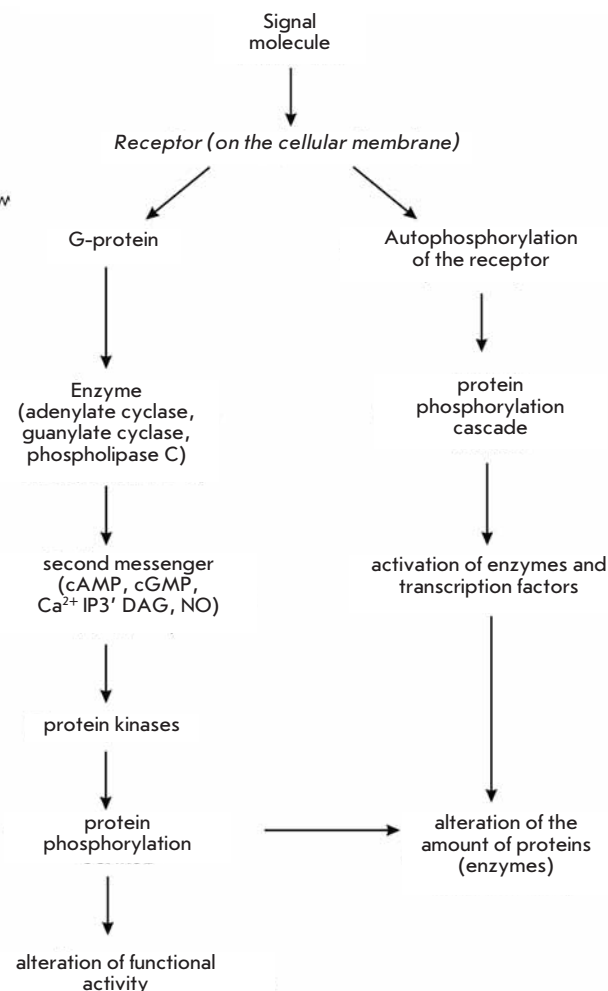


Fig. 3. The basic stages of signal transduction via protein phosphorylation. IP – inositoltriphosphate, DAG – diacylglycerine

and has a phosphorylation site, causes autophosphorylation and leads to phosphorylation of the receptor's substrates and also of other proteins [10]. The epidermal growth factor receptor (EGFR) belongs to a family of growth factor receptors which bind protein ligands and also exhibit tyrosine kinase activity [14]. After binding the appropriate ligand, the receptor forms a dimer, five tyrosine residues are autophosphorylated on the C-terminus of the receptor, and the protein acquires intracellular tyrosine kinase activity. Further EGFR activity is involved in the initiation of the signal transduction cascade, which includes the activation of mitogen-activated protein kinases, protein kinase B, JNK (Jun N-terminal kinase), or Stress Activated Protein Kinase (SAPK) – the so-called MAP-kinase family. This promotes DNA synthesis and proliferation [11, 12, 18–20].

Cytoplasmic domains of other receptors (somatotropin, prolactin, cytokines, etc.) do not exhibit tyrosine kinase activity themselves but are instead associated with other cytoplasmic protein kinases (the so-called «Janus kinases» or JAK family kinases), which phosphorylate the receptors and thus

REVIEWS

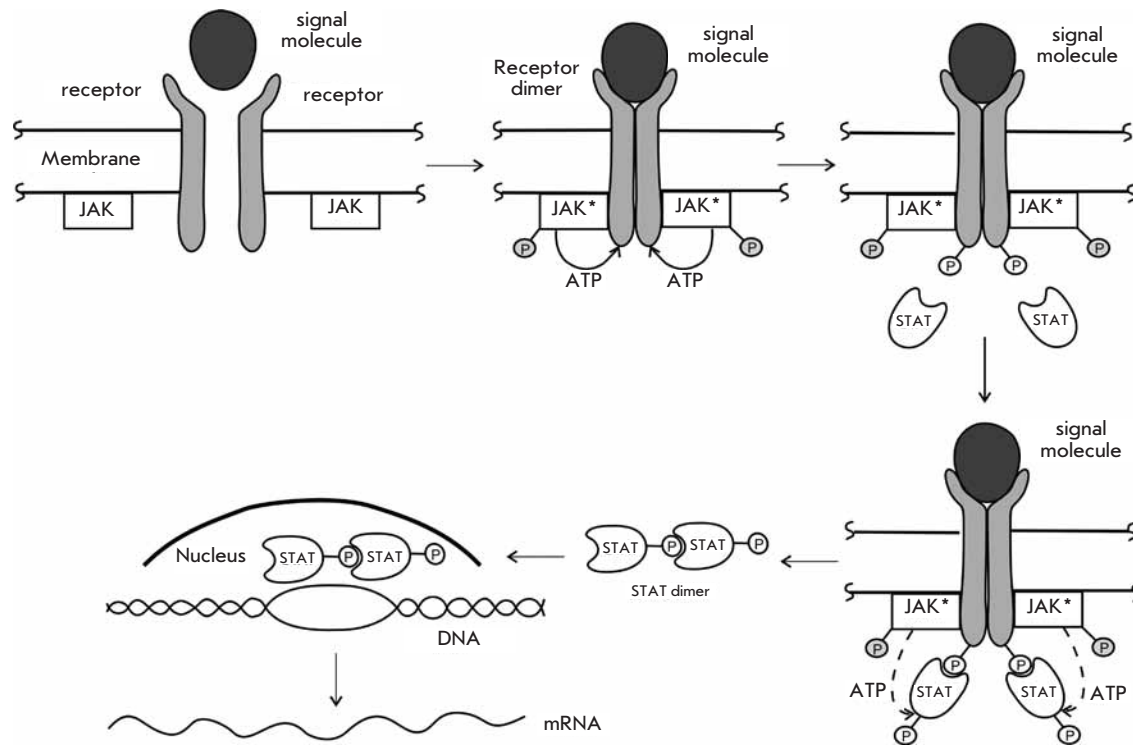


Fig. 4. A schematic representation of signal transduction via JAK-associated membrane receptors

activate them [11, 18]. The defining feature of Janus kinases among all the other mammalian tyrosine kinases is their tandem kinase (JH1) and pseudokinase (JH2) domains. The latter is the cause for the name “Janus kinase,” since they are the only mammalian tyrosine kinases with a pseudokinase domain; thus, they have two “faces” just as the two-faced god Janus. The pseudokinase domain, though it is very similar to kinase domains, does not possess any of the residues responsible for phosphotransferase activity. Apparently, the function of this domain is the regulation of catalytic activity.

Binding of the signaling molecule by a receptor is thought to activate signaling via homo- and heterodimerization of the receptor subunits, which then bind to Janus kinases. This leads to autophosphorylation of the kinases and increases their catalytic activity. The activated Janus kinases phosphorylate tyrosine residues in the subunits of the receptor, which allows the receptor to bind other proteins, for instance the Signal Transducer and Activator of Transcription proteins (STAT). These STAT proteins are then phosphorylated by the Janus kinases, form dimers, and are transported into the nucleus, where they bind specific DNA motifs, thus regulating transcription (Fig. 4).

Mitogen-activated kinases (MAPK, E.C. 2.7.11.24) respond to extracellular stimuli (mitogens) and regulate a range of cellular processes (gene expression, cell division, differentiation, and apoptosis) [11, 17–20]. This MAP signal cascade is conservative in eukaryotes, from yeast to mammals.

The activity of serine/threonine protein kinases is influenced by a number of factors, for instance damage to DNA, and also a range of chemical signals, including cAMP, cGMP, diacylglycerol, and Ca^{2+} calmodulin [5, 8, 21–24]. This type

of protein kinases phosphorylates serine or threonine residues in consensus sequences, which form a phosphoacceptor site. This amino acid sequence in the substrate molecule allows contact between the catalytic groove of the protein kinase with the phosphoacceptor site, which creates kinase specificity not towards a certain substrate but towards a certain family of proteins sharing the same consensus sequence. While the catalytic domains of the protein kinases are highly conservative, the recognition sites vary, which allows the recognition of various substrates. Protein kinases A, B, C, G, calmodulin-dependent protein kinases, etc. are all regulated by hormone signal second messengers.

The phosphorylation reaction can take place not only at a single site in the protein molecule, but also at multiple sites, which causes the phosphorylation of the functional groups of various amino acid residues [25–28]. Multiple phosphorylation is characteristic of several enzymes; for instance eukaryotic RNA polymerase II (E.C. 2.7.7.6) [28]. The C-terminus of this enzyme’s major subunit carries a large number (52 for mammals, 26–27 for yeast) of repeated heptapeptide consensus sequences (Tyr-Ser-Pro-Thr-Ser-Pro-Ser). Multiple phosphorylation of these repeats at the serine and threonine residues enhances the binding of a large number of transcription elongation factors and their associated proteins. This is a vital step in conversion of the enzymatic transcription preinitiation complex into a stable elongation complex [29], which allows the RNA polymerase to move along the chromatin DNA.

PROTEIN ACETYLATION

One of the widely spread types of posttranslational modification that plays an important role in living organisms is acety-

lation [30–38]. The reaction takes place at the ϵ -aminogroups of lysine residues, and acetyl coenzyme A acts as a donor of acetyl groups. The positive charge of the amino group disappears after this reaction, causing a redistribution of charge in the whole protein molecule, and also increasing the hydrophobicity and size of the modified amino acid's side chain. Among other things, histones use this as a binding signal for transcription factors and associated proteins, i.e. transcription initiation. A very important feature of the proteins that can be acetylated is a so-called bromodomain, a conservative 110 amino acid module [30, 31].

The acetylation process has been well studied on histone proteins [32–38]. Selective acetylation of several lysine residues creates specific chromatin affinity towards certain transcription factors, which predetermines which genes will be expressed. This is why the distribution of acetylation sites between histones and among their amino acid residues is an important factor in the regulation of chromatin expression and is usually considered as one of the elements of the “histone code,” which governs the above-mentioned process. In general, the “histone code” includes the whole range of amino acid modifications in the N- and C-terminal sequences of histones (phosphorylation, acetylation, methylation, and ADP-ribosylation), which determines the functional status of the gene with respect to replication and transcription [33–38].

Various forms of the histone acetyltransferase (E.C. 2.3.1.48) catalyze the acetylation of lysine residues located at specific positions in the protein molecule. For instance, the octamer core of a nucleosome, which consists of two copies of H2A, H2B, H3, and H4 histones, contains 30 conservative lysine residues available for acetylation in the N-terminal domains of the proteins (residues in positions 5 and 9 in H2A; residues 5, 12, 15 and 20 in H2B; residues 9, 14, 18, 23 and 27 in H3; and residues 5, 8, 12 and 16 in H4) [39]. Since the number of modified amino acid residues and their size can vary, this creates a multitude of combinations for acetylated residue distribution, which plays an important role in chromatin function. For instance, acetylation of Lys-18 in *Saccharomyces cerevisiae* yeast histone H3 is the main indicator of active chromatin transcription. This modified residue binds the largest number of transcription factors. Activation of β -interferon genes in humans requires acetylation of Lys-8 in the H4 histone and Lys-14 in the H3 histone [39].

It was discovered that acetylation of lysine residues in the C-terminal domains of proteins protects the protein from modification by ubiquitin, thus increasing the lifespan and active functioning time of this protein. [40].

ACYLATION OF PROTEINS BY HIGHER FATTY ACID RESIDUES

The most widespread modifications by addition of fatty acid residues are myristoylation, which is the addition of a $\text{CH}_3-(\text{CH}_2)_{12}-\text{CO}-$ residue to the amino group of an N-terminal glycine [1, 41, 42], and palmitoylation, which is the addition of a $\text{CH}_3-(\text{CH}_2)_{14}-\text{CO}-$ residue at the SH-group of a cysteine residue [1, 43, 44]. In both cases, the acylation is accomplished by the appropriate acyl coenzyme A, which is produced during oxidative decay of longer fatty acids.

An N-terminal glycine residue [42, 45] appears in proteins after the N-terminal methionine residue, used to signal the start of translation, is cleaved away. Addition of the myristil

group is catalyzed by the myristoyl CoA: protein N-myristoyltransferase (E.C. 2.3.1. 97) [46, 47]. The formation of an amide bond between glycine and myristate is an irreversible process. Introduction of the myristoyl residue alters the lyphophilic qualities of the protein molecule and promotes weak and reversible interactions of the protein with the phospholipid membranes or hydrophobic domains of other proteins. Such an interaction is vital for cell signaling, apoptosis, and extracellular protein transport activities. Protein kinase A and GAG, one of the main structural proteins of HIV, are examples of myristoylated proteins [45, 48]. Usually, modification by myristic acid acts in conjunction with other protein regulatory mechanisms.

Often, myristoylation of the N-terminal glycine is followed by addition of a palmitic acid residue to a cysteine residue, thus forming a thioester bond [1, 43, 45, 49]. Unlike myristoylation, this modification is reversible: there are several enzymatic mechanisms that catalyze palmitoylation of cysteine residues, as well as their depalmitoylation [50].

Introducing a palmitic acid residue has the same results as glycine modification by myristate, and the lyphophilicity of the protein molecule increases. This enhances the interactions with membranes and promotes transport through them, while the possibility of the reverse depalmitoylation reaction allows the regulation of the protein activity on various stages of the cell cycle and cell signaling. Palmitoylation is usually seen in proteins that participate in signaling: G-proteins (small G-proteins from the Ras-family, α -subunit of heterotrimeric G-proteins) and non-receptor tyrosine kinases of the Src-family (Fyn, Lck) [43, 45, 47, 51].

PROTEIN UBIQUITINYLATION

Acylation of proteins by the activated C-terminal carboxyl group of glycine in ubiquitin, an 8kDa peptide consisting of 76 amino acid residues, is of great biological importance [52–59]. The main, although not the only, purpose of this reaction is the marking of proteins for degradation. These include various damaged proteins, as well as ordinary proteins which fulfill their functions in certain phases of the cell cycle and whose activity is unfavorable during other phases.

Conjugation of the target protein and ubiquitin is a three-stage process. The first stage is the activation of the carboxyl group of ubiquitin, performed by the ubiquitin-activating enzyme E1 using ATP, thus forming ubiquitinyl-AMP. The second stage is the transfer of the ubiquitin residue onto the SH-group of the ubiquitin-transporting protein E2. In the third stage, the ubiquitin-protein ligase E3 catalyses the transfer of ubiquitinyl residue onto the protein substrate, forming an amide bond between the C-terminal glycine of ubiquitin (G76) and a lysine residue in the target protein (substrate). A thus-modified protein is a target for proteolysis in proteasomes or lysosomes [57].

Whereas E1 is the single such enzyme in the cell, E2 has 20–40 isoforms, and the E3 enzyme has hundreds of isoenzymes, which differ by the nature of the protein substrate. Preliminary modification of the target protein is often needed in order for the E3 enzyme to recognize its substrate (phosphorylation (Ser/Thr, Tyr), hydroxylation (Pro), glycosylation (Asn), and N-terminal aminoacylation) [54].

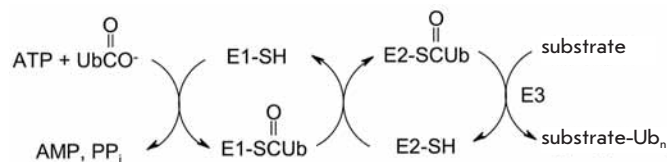


Fig. 5. Addition of an ubiquitin residue (residues) onto a substrate protein. E1-SH – ubiquitin-activating enzyme, E2-SH – ubiquitin transport protein, E3 – ubiquitin-protein ligase. Ub – ubiquitin residue

The target protein molecule can be modified by one or several molecules of ubiquitin. The scheme (Fig. 5) denotes such a product as substrate-Ub_n. Polyubiquitinylation of the substrate involves the acylation of the ubiquitin fragment already bonded to the target protein (Lys-29, Lys-48 or Lys-63) with the C-terminal glycine residue of the other ubiquitin molecule [53, 60–63]. The formation of the ubiquitin-protein covalent adduct does not interfere with the conjugation of the above-named lysine residues with another ubiquitin; thus, this process eventually leads to polyubiquitinylation of the substrate protein. (Fig. 6).

The degree to which the conjugate has been ubiquitylated defines its biological function. Thus, effective proteasome degradation of proteins requires tetraubiquitinylation at Lys29 or Lys48, depending on the target protein. Misfolded proteins and the majority of short-lived proteins form tandem chains of ubiquitin residues connected by bonds at Lys48 [59]. Monoubiquitinylation usually takes place on random multiple lysine residues in the target protein. This happens during the metaphase anaphase transition in mitosis, when metaphase proteins need to be “switched off.” Monoubiquitylation of the human H2B is required for the methylation of histone H3, which in turn is very important for chromatin remodeling and for the transcription activation of “silent genes” [35]. Tandems of several ubiquitin residues connected via Lys63 and bonded with PCNA (Proliferating Cell Nuclear Antigen) play an important role in postreplicative DNA repair [59, 61].

Currently, several ubiquitin-like proteins (ULP) are known, and they are all grouped into the ubiquitin family including ubiquitin itself, Nedd8, Sumo, Fat10, ISG15, Urm1, Hub1, etc. [53, 56–59, 62, 64]. These proteins are variously homologous to ubiquitin in their amino acid sequence and share a similar spatial structure. A large number of ULP in cells indicates their involvement in a wide range of different cellular processes. Thus, Sumo is involved in nuclear transport, transcription regulation and chromosome segregation; ISG15

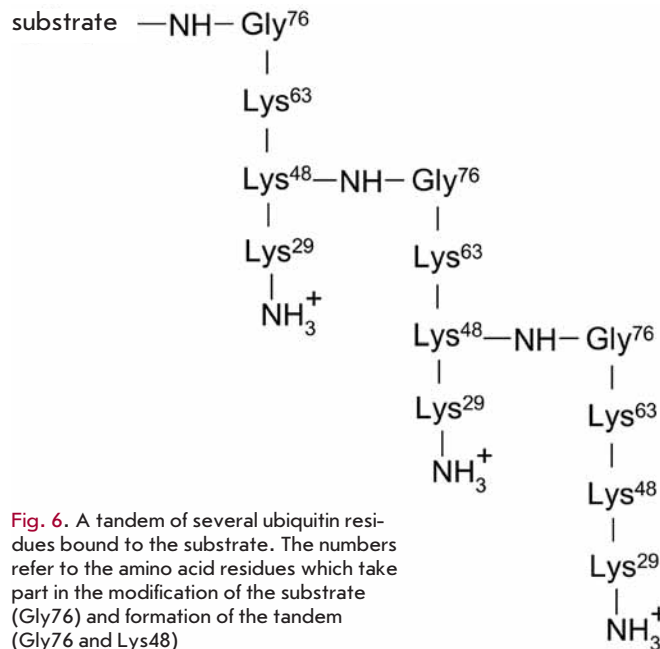


Fig. 6. A tandem of several ubiquitin residues bound to the substrate. The numbers refer to the amino acid residues which take part in the modification of the substrate (Gly76) and formation of the tandem (Gly76 and Lys48)

is part of the immune response cascade; Nedd8 is involved in the meiosis-mitosis switch; and Urm1 is implicated in cellular growth at elevated temperatures [59].

Chaperones interact with newly synthesized, misfolded polypeptides and act as cofactors for ubiquitinylation enzymes, since they possess an ubiquitin-recognition domain. After the target protein has been tagged by ubiquitin, the chaperones escort the ubiquitylated protein into the proteasome, where they dissociate from the protein complex. The ubiquitin chains are unbound, and the target protein is denatured via an ATP-dependent process and then broken down into short peptides by proteases.

PROTEIN ALKYLATION

Another often-seen posttranslational modification is alkylation. This type of modification includes the methylation of lysine and arginine residues [26, 30, 33–38, 39, 65–72] and prenylation (addition of farnesyl and geranyl-geranyl moieties to cysteine side chains) [47, 73–80] (Fig. 7).

Protein methylation in living organisms is catalyzed by methyltransferases [1, 65, 67] and involves the transfer of a CH₃-group from *S*-adenosylmethionine according to the depicted reaction (Fig. 8).

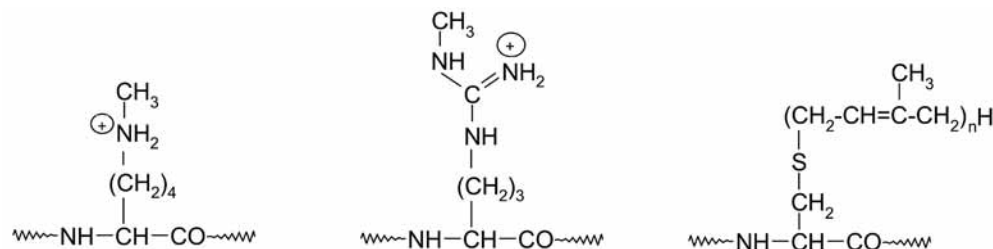


Fig. 7. Structures of alkylated amino acid side chains in proteins

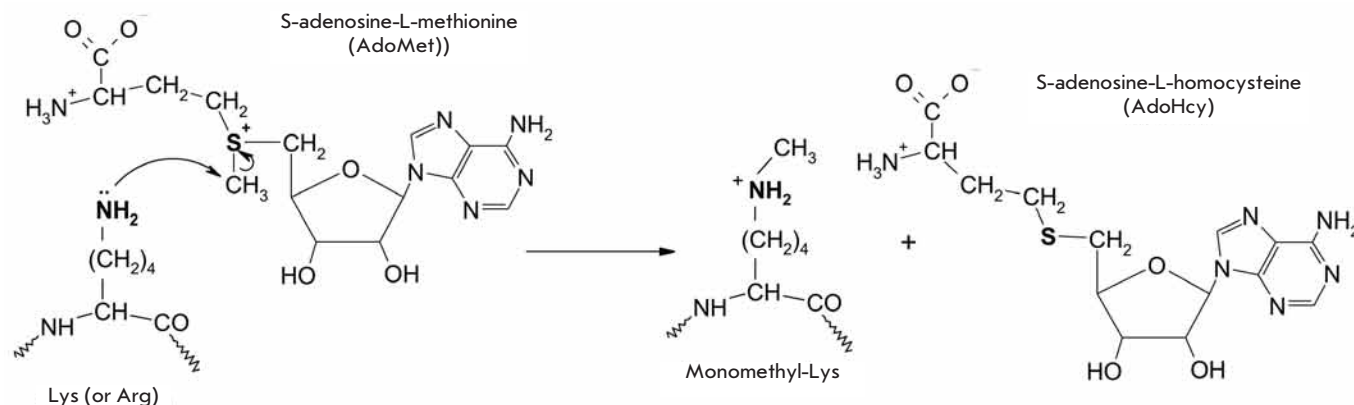
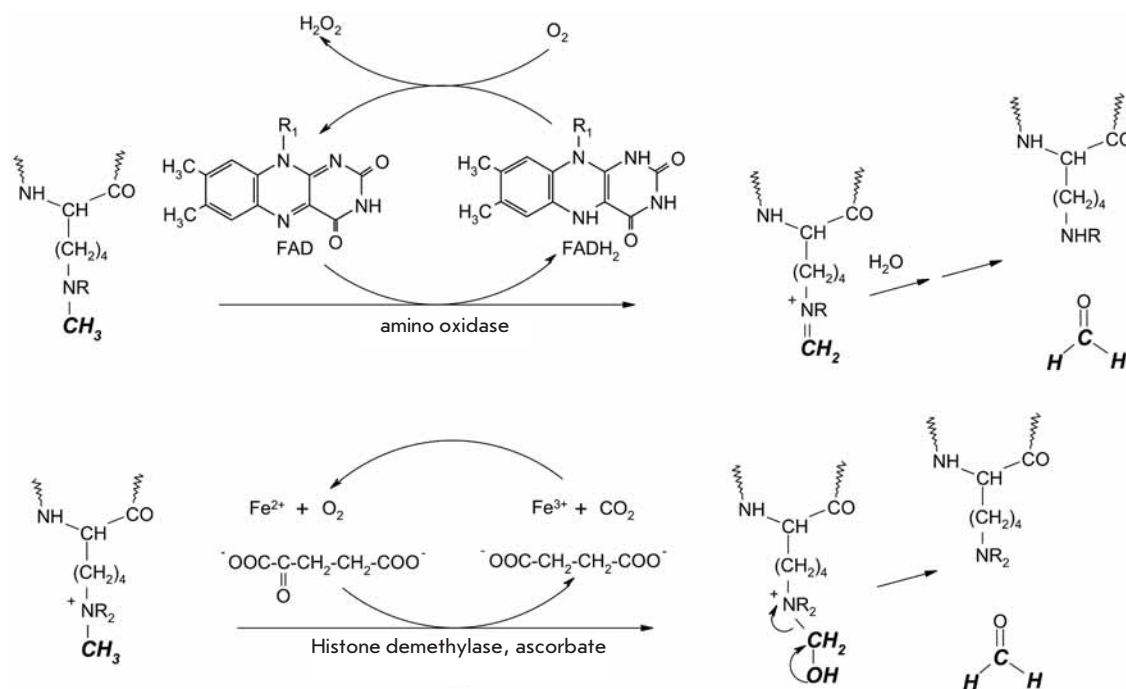


Fig. 8. Methylation of lysine residues by methyltransferases

Fig. 9. Demethylation reaction of di- and monomethylated lysine residues in histones catalyzed by the FAD-dependent aminooxidase (*top*), and tri-, di- and monomethylated lysine residues in histones catalyzed by histone demethylase, which functions in the presence of cofactors, Fe^{2+} ions, α -ketoglutarate and ascorbate (*bottom*)



Lysine can form mono-, di- and trimethyllysines in methyltransferase-catalyzed reactions, while arginine can form mono- and dimethylarginines [65]. These compounds differ by size and hydrophobicity from the original residue.

The mechanics of protein methylation have been best studied in histone modification. Histone methyltransferases are highly specific towards the nature of the amino acid residue (histone-lysine methyltransferases (E.C. 2.1.1.43) and histone-arginine methyltransferases (E.C. 2.1.1.125)) and the position of this residue in the polypeptide chain [1, 65]. Lysine residue methylation in histones is a very important element of the aforementioned “histone code” [33–36, 38]. The best characterized methylation positions in histones are Lys4 and Lys9 in the H3 histone. Besides the mentioned residues, Lys27, Lys36, Arg2, Arg17 and Arg26 residues in H3 can also be modified, as well as Arg3 in the H4 histone [33, 34, 67, 70].

It was demonstrated that the trimethylated Lys4 in the H3 histone is necessary for transcription activation, while dimethylated Lys4 is found both in the active and silent gene [33, 34, 70]. The heterochromatin protein 1 (HP1) interacts with the trimethylated Lys9 of H3 via its chromodomain (a recognition domain for alkylated amino acid residues), initiates local chromatin condensation, and recruits other protein factors into the assembly of an active transcription complex [26, 30, 33, 67, 70].

Until recently, it was thought that the methylation of lysine residues was an irreversible process [1]. But a short while ago, researchers managed to extract enzymes that catalyzed the cleavage of methyl groups from lysine and arginine residues, which means that this type of posttranslational modification is also dynamic. Demethylation of lysine is an oxidative process and can be catalyzed either by the FAD-de-

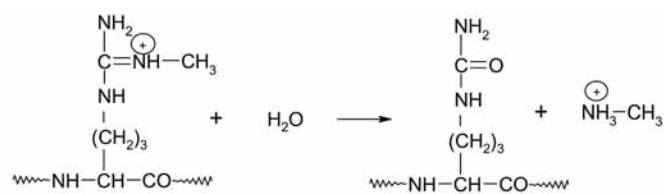


Fig. 10. Demethylation of modified arginine residues catalyzed by the nuclear peptidylarginine deiminase (PAD4) [58]

pendent polyamine oxidase, or a lysine-specific demethylase, which functions as a dioxygenase in the presence of cofactors, such as Fe^{2+} ions, α -ketoglutarate, and ascorbate (E.C. 1.5.3.4) [37, 65, 66, 82, 83]. For schematic representation of this process see Fig. 9.

A nuclear peptidylarginine deiminase (E.C. 3.5.3.15) can demethylate arginine residues, turning methylated arginine into citrulline [66] (Fig. 10).

Thus, methylation-demethylation and acetylation-deacetylation of specific residues in histones are major factors in gene repression and activation.

PROTEIN PRENYLATION

Some cases of posttranslational modification are the addition of isoprenoid moieties onto a cysteine residue. These moieties are formed from isoprene residues – farnesyl and geranyl-geranyl (Fig. 11). Modification of proteins with these radicals is catalyzed by proteinfarnesyl and proteingeranyl-geranyl transferases, respectively (E.C. 2.5.1.58 and E.C. 2.5.1.59 or E.C. 2.5.1.60; Type I and II geranyl-geranyl transferases). Type I enzymes catalyze the transfer of a geranyl-geranyl residue onto a cysteine residue in a Cys-A-A-X sequence, while type II use the Cys-Cys-X-X, X-X-Cys-Cys or X-Cys-X-Cys sequences [47, 73–80], where A is a small aliphatic amino acid, and X are various amino acids.

Ras-, Rab- and Rho-family proteins (products of the *ras*, *rab* and *rho* proto-oncogenes, involved in cellular growth and differentiation); centromeric proteins; and γ -subunits of heterotrimeric G-proteins, chaperones tyrosine phosphatases are all subjected to prenylation [47, 73, 75, 78, 79, 81]. The C-terminal sequence of Ras-family proteins includes a Cys-A-

A-X motif, in which X is the amino acid that determines the enzyme specificity: Leu, Phe, and Met in case of the type I geranyl-geranyl transferase; and Ala, Gln, Ser, Met, and Phe in the case of the farnesyltransferase [47, 74, 78, 79]. Enzymes that transfer the isoprenyl residues are metalloenzymes, and they carry a single Zn^{2+} ion for each dimeric enzyme molecule. The zinc ion activates the cysteine thiol group for nucleophilic attack by the isoprenyl moiety [73]. The addition of the isoprenyl group to the Cys-A-A-X motif is usually not the last modification of the target protein (Ras, Rho), further processing occurs via proteolytic cleavage of A-A-X tripeptide from the C-terminus by a Cys-A-A-X-specific protease; and carboxymethylation of the isoprenylcysteine residue, by the isoprenyl-cysteine-carboxymethyl transferase (E.C. 2.1.1.100) [84–87] (Fig. 12).

GTPases of the Rab family carry a Cys-Cys-X-X motif near the C-terminus. Both these cysteines can be modified by geranyl-geranyl residues with the help of type II protein geranyl-geranyl transferase, which creates two lipid anchors on the protein molecule [74, 75]. Such a protein exhibits increased affinity towards lipid membranes, and it can thus act as a unique recognition site for specific protein-protein interactions.

Proteins of the Rab family are involved in intracellular vesicle transport circulating between the cellular membrane and the cytosol. Reversible association of the protein with the cellular membrane is achieved through the isoprenyl residues decorating these proteins [75, 84].

Since 20–30% of all human oncological conditions are caused by mutations in Ras family proteins, enzymes that modify these proteins with isoprenyl residues can serve as targets for anti-tumor drugs [73, 79].

PROTEIN GLYCOSYLATION

Glycosylation of proteins plays a very important role in the functioning of eukaryotic cells. Glycosylation modifies the OH-groups of serine and threonine residues (O-glycosylation) and the functional groups of asparagine residue side chains (N-glycosylation) (Fig. 13).

N-glycosylation of proteins happens at the carboxamide nitrogen atom of an asparagine residue in the context Asn-X-Ser/Thr. N-glycoside formation begins in the endoplasmic reticulum. The oligosacaryl transferase enzyme (E.C. 2.4.1.119)

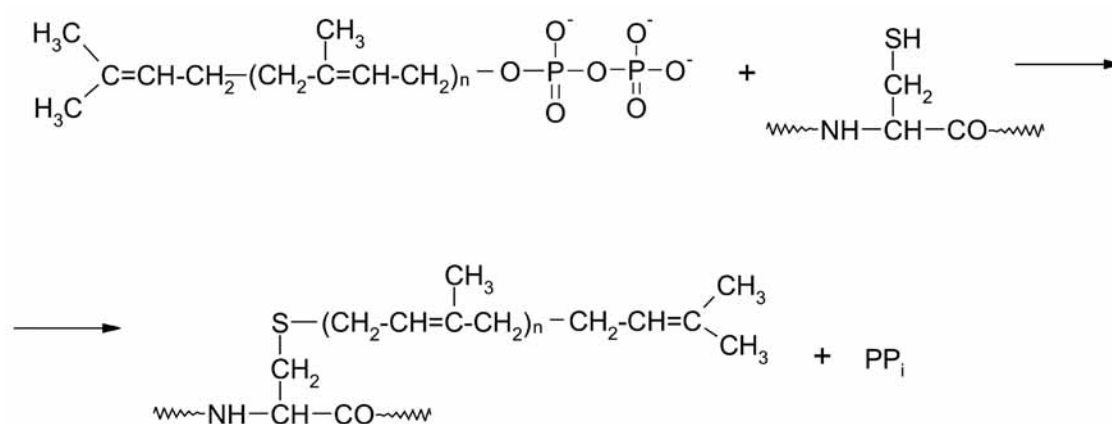


Fig. 11. Transfer of an isoprenoid residue from pyrophosphate to a cysteine residue in an apoprotein. $n = 2$ – is a farnesyl residue, $n = 3$ – geranyl-geranyl residue

Fig. 12. Prenylation of the Ras protein: 1 – addition of a farnesyl residue onto the Cys-A-A-X sequence (A- a small aliphatic amino acid residue, X is Leu, Phe or Met); 2 – Cleaving of the A-A-X tripeptide by the Ras-converting enzyme, which is a CysAAX-endopeptidase; 3 – carboxymethylation of the isoprenylcysteine residue catalyzed by the isoprenylcysteine carboxymethyltransferase [86]

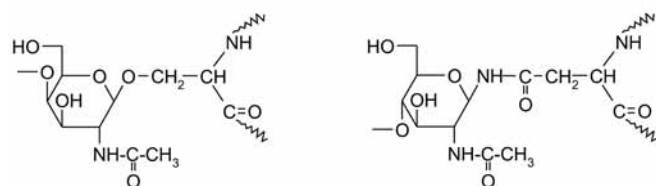
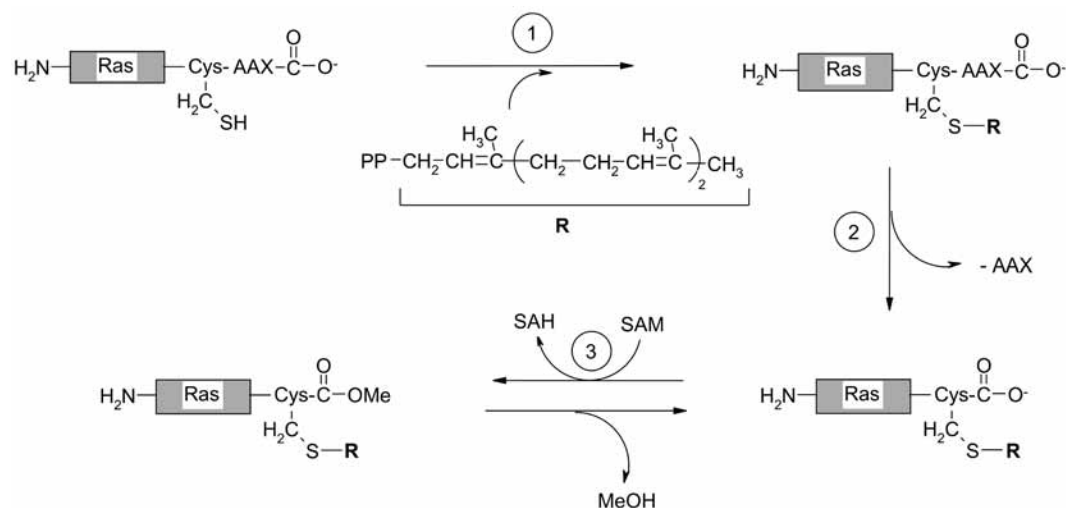


Fig. 13. Structures of the products of *N*-acetylglucosamine addition onto serine and asparagine side chains in proteins

transfers a branched tetradecasaccharide fragment onto the target protein. This fragment is (Glc₃Man₉(GlcNAc)₂), and it comes from the carbohydrate donor molecule dolichol pyrophosphate.

The vast variety of glycoproteins is assured by the processing of the protein-bound tetradecasaccharide residue, which is accomplished by a set of glycosidases and glycosyl transferases.

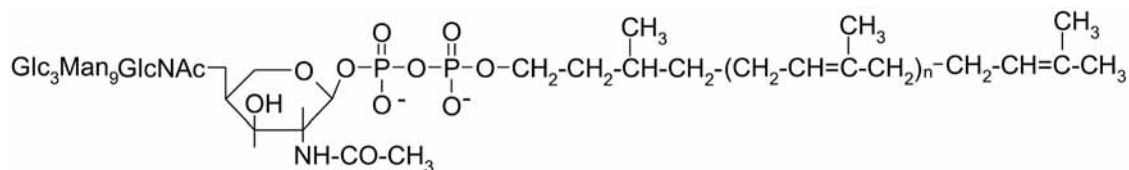
Figure 15 presents the structure of a bound tetradecasaccharide and the products of the first stages of processing, which are catalyzed by glucosidases I and II (E.C. 3.2.1.106) that cleave away two glucose residues, and mannosidases (E.C. 3.2.1.130) that cleave away 6 mannose residues. The glycoprotein formed after separation of the two glucose residues, and thus bearing an *N*-bound dodecasaccharide residue, is then recognized by the chaperones calnexin and calreticulin, which facilitate correct folding of the protein while it is being transported from the location of synthesis on the membrane-bound ribosomes to the inside of the endoplasmic reticulum [1, 88, 89, 90–93]. After a third glucose residue is cleaved away by an endoplasmic reticulum glucosidase, the chaperones lose

their affinity towards the undecasaccharide and dissociate from the glycoprotein complex. UDP-glucose:glycoprotein glucosyltransferase (E.C. 2.7.8.19) returns a glucose residue back onto the undecasaccharide, which makes canexin and calreticulin continue the glycoprotein folding. This is a mechanism for maintaining the functional structure in secreted glycoproteins.

If a glycoprotein is not folded correctly during several rounds of deglycosylation-reglycosylation, then it is transported into the cytosol. There, it is polyubiquitylated by the E3-ligase, which is a part of the degradation system for misfolded proteins in the endoplasmic reticulum and is hydrolyzed in the proteosomes [1, 88, 89, 90–94].

Correctly folded Man₉(GlcNAc)₂*N*-glycoprotein loses 6 mannose residues with the help of endoplasmic reticulum and Golgi apparatus mannosidases and forms a protein conjugated with a core pentasaccharide (Man₃(GlcNAc)₂). The latter can receive various monosaccharides with the help of a number of glycosyl transferases, of which there is a great many in the endoplasmic reticulum and the Golgi apparatus. Thus, the variety of glycoproteins is numbered in tens of thousands [1, 88, 89, 95].

Glycoprotein *O*-glycoside chains are much shorter and simpler than *N*-glycoside chains. Numerous proteins, including transcription factors, nuclear pore proteins, oncoproteins, etc., contain a monosaccharide residue of *N*-acetylglucosamine, which is introduced into the protein by an *O*-GlcNAc-transferase (E.C. 2.4.1.94) and can be cleaved by the appropriate hydrolase [1, 88, 89, 96, 97–100]. There are also di-, tri- or tetraglycoside fragment bearing *O*-glycosides.



n = 9 - 22

Fig. 14. Structure of the carbohydrate-bearing dolichol pyrophosphate

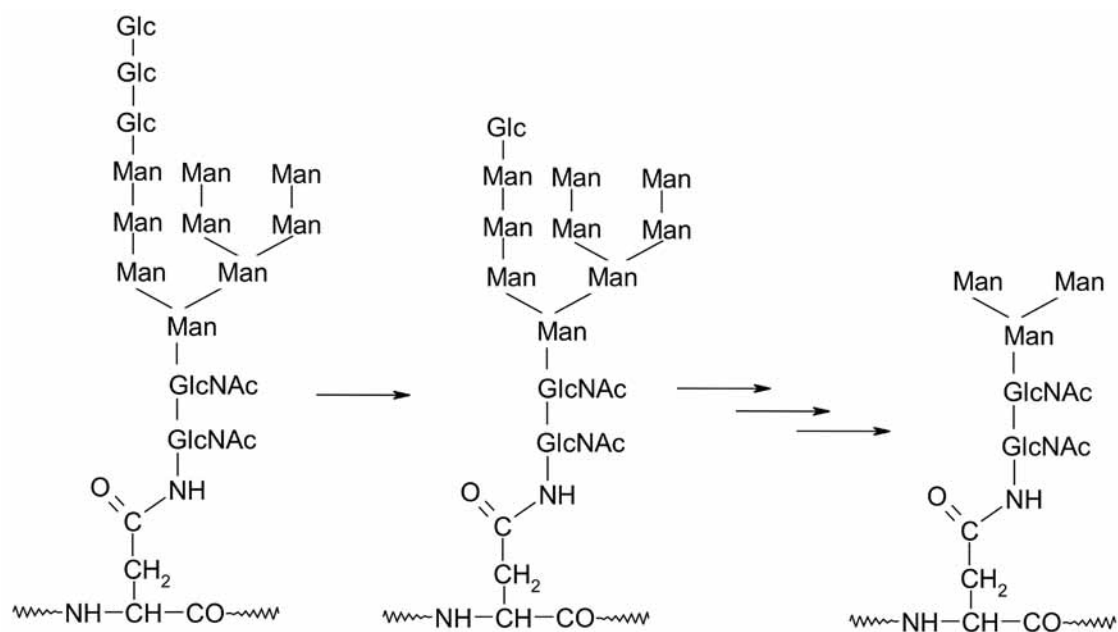


Fig. 15. Structure and first stages of the processing of oligosaccharide fragment $\text{Glc}_3\text{Man}_3(\text{GlcNAc})_2$, as a part of a glycoprotein. Monosaccharides Glc – glucose, GlcNAc – N-acetylglucosamine

Short O-glycoside chains in O-glycoproteins are important for transcription activation, and they act as recognition sites during interaction with cell membrane receptors, which are involved in the transduction of signals into the cell [1, 88, 89, 100–102].

PROTEIN SULFATION

Another posttranslational modification of protein molecules is the addition of a sulfate residue at the OH-group of tyrosine. Phosphoadenylylphosphosulfate acts as a sulfate donor (Fig. 16). The reaction is catalyzed by the sulfotransferase enzyme (E.C. 2.8.2.20) [103, 104].

For instance, three tyrosine residues in the N-terminal region of the human chemokine cell membrane receptor (a regulator of anti-inflammatory immune reactions), which plays an important role in embryo development and in the immune response, are subject to posttranslational sulfation in the Golgi apparatus. This increases the affinity of the receptor towards its ligand, the SDF-1 α chemokine. An enzyme called sulfatase (E.C. 3.1.5.6) was found in lysosomes and was able to catalyze the hydrolysis of sulfoesters [103, 105, 106].

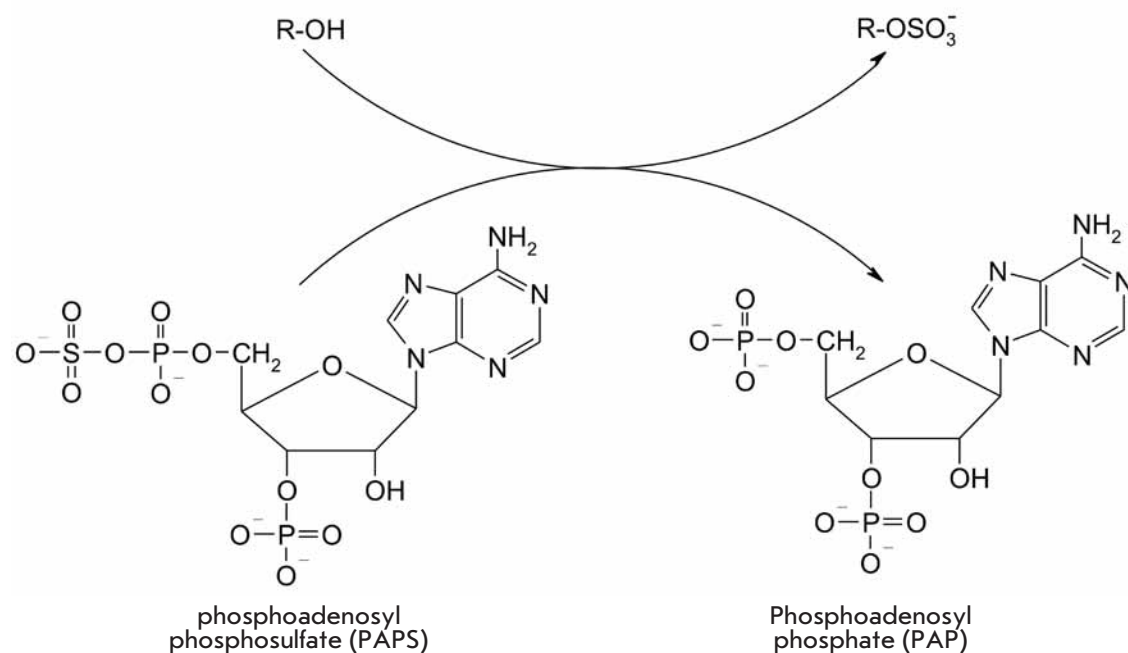


Fig. 16. Sulfation reaction catalyzed by sulfotransferase

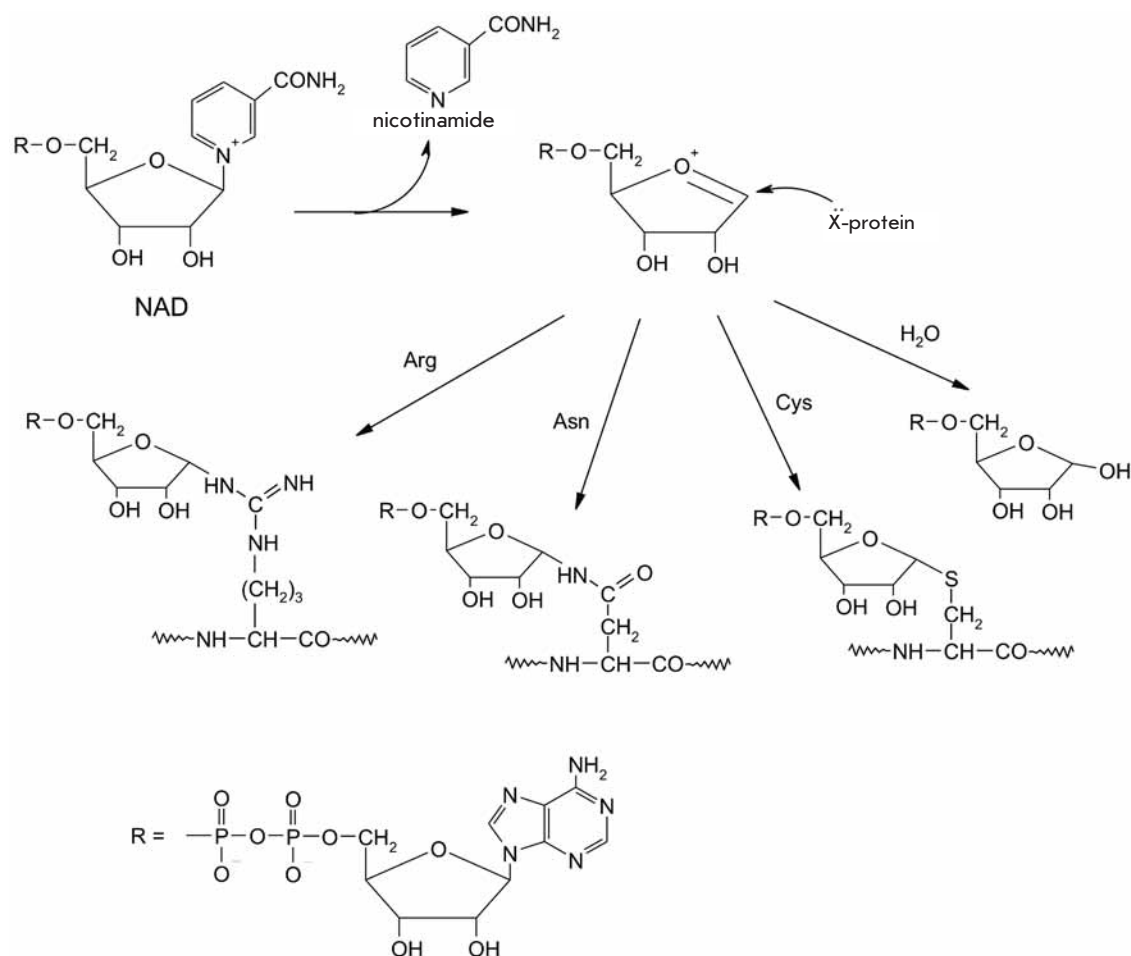


Fig. 17. (ADP-ribosylation) of nucleophilic amino acid residues (X) present in the protein (cysteine, arginine and asparagine) [1]

MONO- AND POLY(ADP-RIBOSYL)ATION

Many cellular processes, such as DNA repair, apoptosis, and the functioning of the spindle during cell division, use mono- and poly(ADP-ribosylation) as an important regulating mechanism [107]. Various pathogenic bacteria secrete toxins that ADP-ribosylate human proteins, thus causing severe diseases, such as cholera, diphtheria, pertussis, and botulism [108–111].

NAD⁺ acts as a donor of the ADP-ribosyl residue. The positively charged nicotinamide bond is cleaved by the ADP-ribosyltransferase (E.C. 2.4.2.31) and forms a ribo-oxocarbene cation, which interacts with various nucleophilic groups in protein active sites and leads to their (ADP-ribosylation) (Fig. 17) [108, 109].

For instance, pertussis toxin transfers the created cation to the thiolate chain of a cysteine residue in the active site of the human G_i-protein α -subunit. This protein regulates synthesis of the second messenger cAMP [1, 111, 112]. Cholera toxin transfers an ADP-ribosyl residue onto the arginine residue in the human G_s-protein α -subunit ([1, 111, 113]. The ADP-ribosyl residue can also be transferred by the C3 toxin of *Clostridium botulinum* onto the nucleophilic Asn41 residue of the minor GTPase of the Rho protein superfamily, which leads to actin depolymerization and impairment of the metabolic processes of the host cell [1, 111].

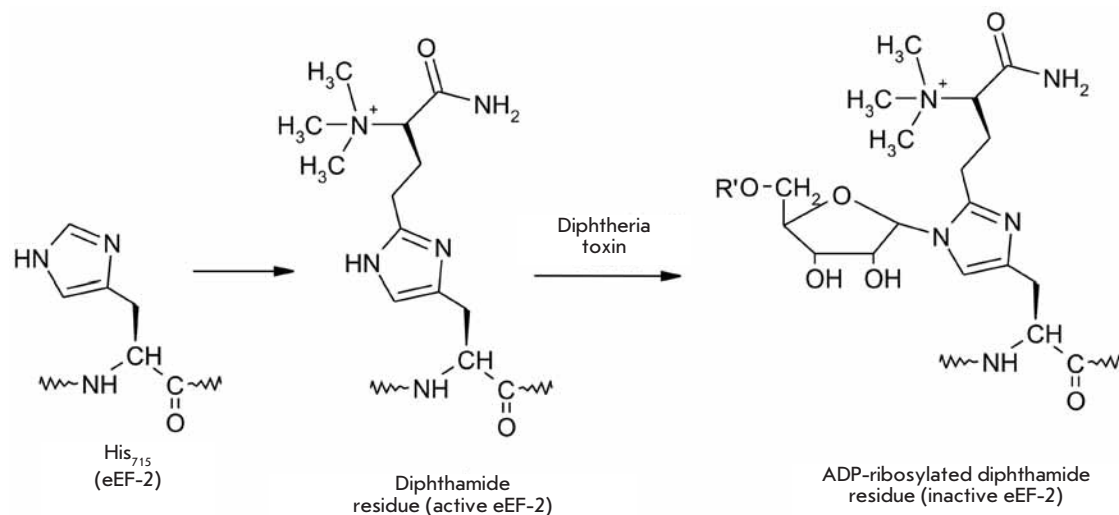
Diphtheria toxin ADP-ribosylates His715 in the eEF-2 elongation factor and, therefore, blocks the translocation of peptides on ribosomes and the whole translation process in human cells [114].

In reality, His715 is subjected to stepwise complex modification: first, an aminocarboxypropyl residue is transferred from *S*-adenosylmethionine (SAM), then SAM-dependent *N,N,N*-trimethylation takes place, then the carboxyl group is amidated in a glutamine-mediated fashion, thus forming a dipthamide residue, and only then does the toxin ADP-ribosylate the dipthamide residue at the N3 atom of the imidazole ring (Fig. 18) [115–117].

During the lifetime of the organism, the genome constantly suffers the effects of genotoxic agents of both exogenic and endogenic nature [118]. An approximate estimate demonstrated that every day the genomes of human cells experience up to 104–106 instances of DNA damage [119]. Under these circumstances, the stability of cell genome is one of the most important factors in maintaining the survival of a multicellular organism, since any uncorrected damage to DNA can promote the emergence of mutator cell phenotypes [120]. Poly(ADP-ribose) (PAR) synthesis is one of the immediate reactions of the cell in response to DNA breaks under the influence of ionizing radiation, or alkylating or oxidizing agents [121, 122]. This process is catalyzed by enzymes poly(ADP-ribose)poly-

REVIEWS

Fig. 18. Modification of the His715 residue in the structure of the human eEF-2 elongation factor results in the blocking of protein synthesis in human cells



merases (PARPs), which are constantly and abundantly expressed in the cell [123]. PARPs are activated in response to DNA breaks and catalyze the posttranslational modification of a range of DNA-binding proteins by covalently adding a polymer poly(ADP-ribose) to the carboxyl groups of glutamic and aspartic acid in the acceptor proteins [124]. Currently, approximately 30 nuclear proteins that are poly(ADP-ribosylated) *in vivo* and *in vitro* have been described [123, 125]. All these proteins exhibit DNA-binding activity and are involved in DNA metabolism (replication, transcription, reparation) or in chromatin formation (histones). Several enzymes of the poly(ADP-ribose) polymerase class have been found in eukaryotes, including PARP1, PARP2, and PARP3, which have nuclear localization; tankyrases 1 and 2, which interact with telomere proteins and are thought to regulate telomere function; VRAP (193 kDa), found in cytoplasmic ribonucleoprotein vault-particles [126]; sPARP – a truncated form of PARP1, which does not require activation by DNA breaks [127]; and macro PARPs (BAL/PARP-9, PARP14, PARP15), which are involved in the epigenetic modification of chromatin [124, 128]. Ninety percent of the nuclear poly(ADP-ribose) synthesis is caused by PARP1 activity [129]. This protein is expressed at a constant level throughout the cell cycle, and each cell carries around $1.0 \cdot 10^6$ of the protein molecules, which amounts to 1 protein molecule for each 6000 nucleotide pairs [130]. Catalytically inactive PARP1 is present in the nucleoplasm and is activated by DNA breaks. It then binds to the damaged area and catalyzes PAR synthesis [128]. PARP synthesizes poly(ADP)-ribose in three stages: initiation, elongation, and branching of the polymer (Fig. 19).

The first stage involves the formation of the ester bond between the ADP-ribose and the carboxyl group in a glutamate residue in the acceptor protein [131, 132]. The second stage involves the formation of an *O*-glycoside bond between the C2' and C1'' atoms of the ADP-ribose, thus creating a linear polymer of ADP-ribose molecules [133, 134]. In the third stage, the glycoside bond links the C2'' and C1''' atoms of the

ADP-ribose, forming branches in the polymer structure [135, 136] (Fig. 19).

The rate of the chemical reaction at the mono(ADP-ribosylation) stage is approximately 200 times slower than at the elongation stage [137]. Based on the measurement of kinetic parameters of the *in vitro* PARP-catalyzed poly(ADP-ribosylation) reaction, the authors of [138] hypothesize that the latter reaction is inter-molecular, meaning that PARP1 functions as a homodimer at the DNA break site. Two molecules react with the DNA break at once, and during the reaction both molecules simultaneously synthesize PAR and function as acceptors. The covalent modification of PARP1 by the addition of a charged poly(ADP-ribose) residue leads to alterations in the enzyme's physicochemical characteristics and its dissociation from the DNA-complex [139]. Thus, regulation of PARP1 DNA-binding activity can be achieved through self-modification [140].

Discovery of poly(ADP-ribosylation) modifications in chromatin remodeling proteins, histones *in vivo*, and topoisomerases *in vitro* leads to the assumption that PARP1 is involved in chromatin remodeling during DNA repair [123, 133, 141]. It was demonstrated that the kinetic parameters of DNA repair reactions were influenced by the presence of histones on the damaged DNA [123]. *In vivo* poly(ADP-ribosylation) of the H1 histone and the histones forming the nucleosome core during DNA damage can play an important role in DNA repair, especially if the DNA is structured as chromatin, since histone modification can lead to their dissociation from the DNA molecule, thus allowing the repair enzymes easy access to the damaged site [123, 140].

Therefore, the current overall notion is that the cell response to damaged DNA can be modulated by the activity of PARP1. On one hand, PARP1 activates repair processes, thus promoting cell survival; on the other hand, when DNA damage is irreparable and the emergence of a mutator phenotype is highly probable, "overactivation" of PARP1 induces cell death [142]. This is why the PAR synthesis catalyzed by

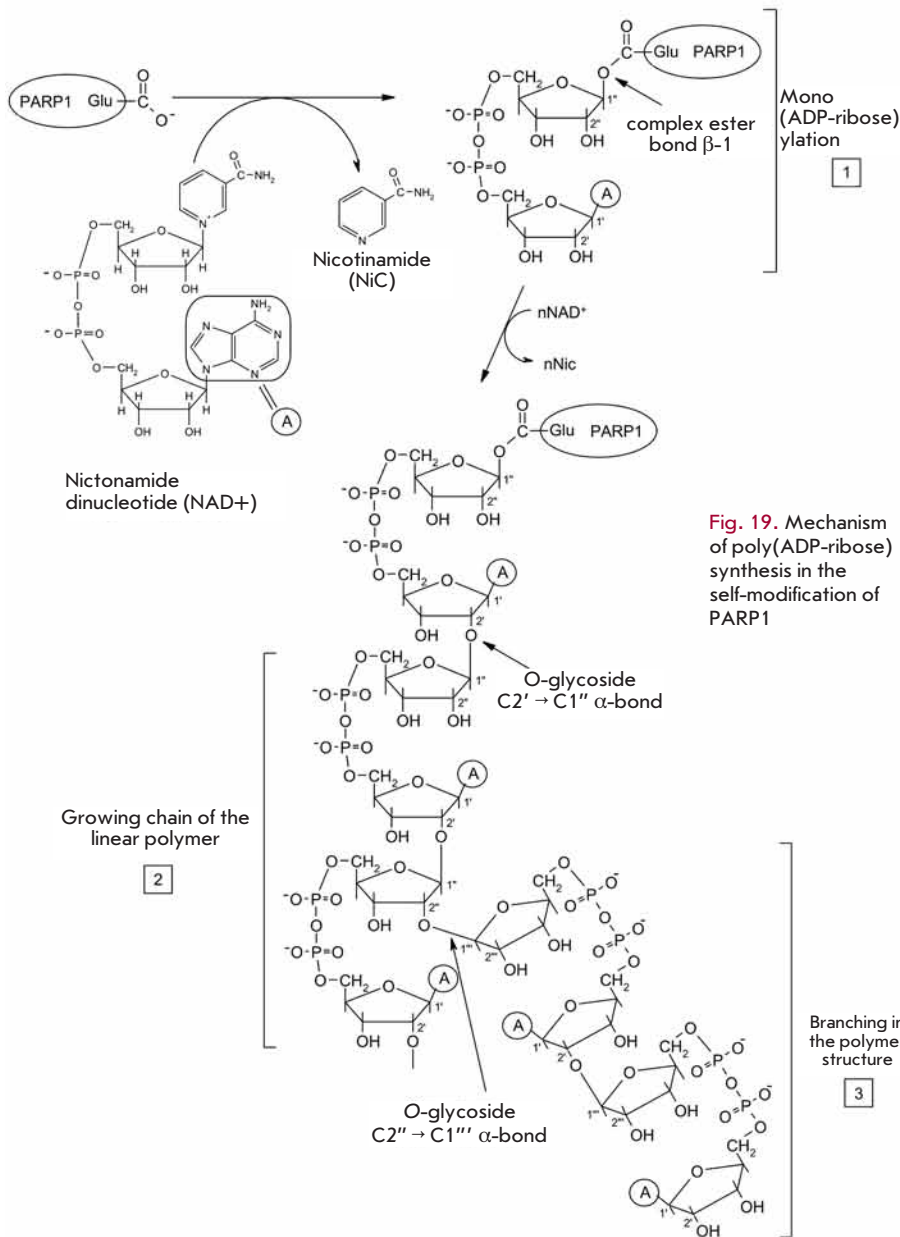


Fig. 19. Mechanism of poly(ADP-ribose) synthesis in the self-modification of PARP1

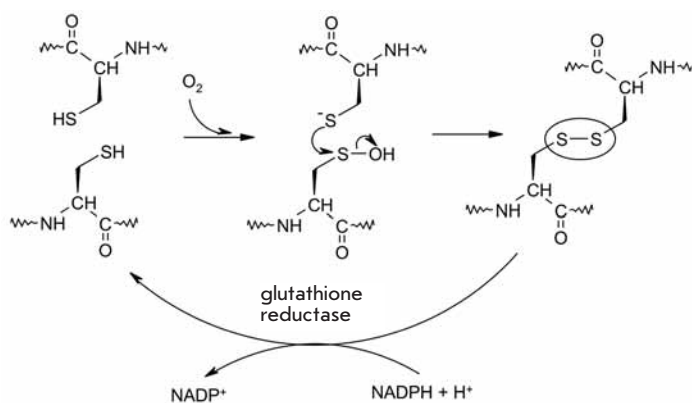


Fig. 20. Oxidation of the sulfhydryl group of the cysteine residue, which can be reduced again into a thiol group with the help of NAD(P)H and a glutathione reductase [147]

PARP in the process of interacting with DNA breaks can be regarded as a signal of the DNA damage level, which is used to determine the cell's future functional strategy.

OXIDATION OF THE SULFOHYDRIDE MOIETY OF THE CYSTEINE RESIDUE IN PROTEINS

A large number of proteins are characterized by the formation of disulfide bonds in a reaction between cysteine residues either inside a single polypeptide chain or between different polypeptide molecules. Such bonds fulfill a structural function and determine the tertiary and quaternary structure of the protein, which are vital for the protein's metabolic functions in the organism. This modification is also involved in the regulation of the cell's reduction-oxidation status, which affects numerous aspects of cellular processes, such as proliferation, differentiation, and apoptosis by changing the functioning of proteins via a reversible modification of cysteine residues [143–147].

Oxidation of cysteine residues involves the following processes: formation of a disulfide bond, the formation of sulfi- and sulfoacids, and binding of glutathione [145]. Formation of a disulfide bond is accomplished via the oxidation of the electron-rich sulfhydryl moiety (or of the thiolate anion, which is generated from the former after proton dissociation) of the cysteine residue side chain. One-electron oxidation of the sulfhydryl moiety leads to the formation of a thiyl radical, which can dimerize into a disulfide [147].

Under physiological conditions, most of the sulfhydryl groups are in oxidized form and thus involved in disulfide bonds. Reduction of the disulfide bonds *in vivo* is accomplished by the glutathione tripeptide γ -Glu-Cys-Gly (GSH), which converts into oxidized glutathione (GSSG). High levels of NAD(P)H and of the glutathione reductase (E.C. 1.8.1.7) and thioredoxinreductase (E.C. 1.8.1.9) enzymes lead to the reduction of oxidized glutathione [143–147] (Fig. 20). As proteins move down the secretory pathways of eukaryotic cells, the levels of glutathione and NAD(P)H decrease, which is why most proteins exist in structures stabilized by disulfide bonds [148].

Oxidizing agents (hydrogen peroxide, hydroxide radical) can oxidize the

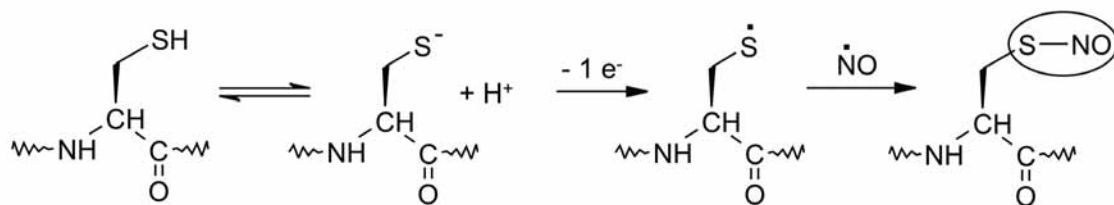


Fig. 21. Oxidation of the thiolate ion in the presence of nitrogen oxide results in the formation of cysteinyl-nitroxide [154]

cysteine sulfhydryl group into the cysteine-sulfenic acid (-SOH) [147]. Interaction of the cysteine-sulfenic acid residue with the closest Cys-S⁻ group also results in the formation of the disulfide bond.

Reduction of the disulfide bond can result either from thiol-disulfide exchange with either glutathione or thioredoxin (TSH), a low-molecular-weight (12 kDa) protein which contains catalytically active sulfhydryl groups in its active center (Cys-Gly-Pro-Cys) and plays the central role in the regulation of the reduction-oxidation status of disulfide bonds in proteins, which in turn governs a wide range of cellular processes. Oxidized forms of these compounds are reduced by NAD(P)H and glutathione reductase/thioredoxin reductase [146–149].

Both the thiolate ion and the thiyl radical can interact with other oxidizing agents and radicals (such as NO[•]) (Fig. 21). The resulting CysSNO molecule is involved in oxidation signaling in the cell [150–154].

HYDROXYLATION OF PROTEIN FUNCTIONAL GROUPS

Another type of posttranslational modification is the oxidative hydroxylation reaction. This reaction takes place at non-nucleophilic amino acid residue side chains: the CH₂-groups of proline, lysine and asparagine form 3-hydroxyproline, 4-hydroxyproline, 5-hydroxyproline, and 3-hydroxyasparagine, and this process is catalyzed by iron-containing monooxygenases of the E.C. 1.14.16 subclass [155, 156, 157] (Fig. 22).

Oxidized proline and lysine residues play an important role in the formation of hydrogen bonds in the tri-strand spatial structure of the connective tissue protein collagen. Oxidation takes place at the Pro-Gly and Lys-Gly sequences. 4-hydroxyproline is found about 10 times more often than 3-hydroxyproline [155–160].

Besides the above said, hydroxylation of specific amino acid residues plays a role in the function of the HIF transcription factor (hypoxia inducible factor) [156, 159–161]. This pro-

tein is activated under conditions of insufficient oxygen. It induces the transcription of a wide range of genes, including the gene encoding erythropoietin, which stimulates erythrocyte differentiation from precursor cells, thus increasing the transport of oxygen to cells suffering from hypoxia [160].

The α-subunit of the human HIFαβ is posttranslationally hydroxylated in the central region of the molecule at two proline residues, Pro402 and Pro564, forming 4-OH-Pro, and also in the C-terminal region at Asn803, forming 3-OH-Asn [156]. A molecule bearing hydroxylated proline residues is subjected to ubiquitylation by the E3 ligase, and the lifespan of HIF is determined by the rate of hydroxylation, ubiquitylation, and proteolysis in the proteasomes. Low O₂ pressure causes slow hydroxylation of proline. High oxygen pressure causes the Pro-hydroxylase to efficiently hydroxylate Pro residues, which increases affinity towards the E3 ligase 1 000-fold and causes rapid ubiquitylation and decay in the proteasomes, while at low oxygen pressures, HIF is fairly stable and can exist for a long time [162, 163].

The hydroxylation of proline and asparagine side chains is catalyzed by a family of oxygenases that contain non-heme iron [163]. The active site of the enzyme (Fig. 23) contains two histidines and one asparagine, which take up three of the six coordination spaces around the Fe²⁺ atom, while two spaces are occupied by the α-ketoglutarate co-substrate; and the sixth, by oxygen. Interaction of the α-ketoglutarate and oxygen results in oxidative decarboxylation and yields CO₂ and succinate, which accepts one of the oxygen atoms of the molecular oxygen. The second oxygen atom takes part in the generation of the high-valence Fe⁴⁺=O complex. The latter group is an effective oxidizing agent, which cleaves the unactivated C-H bond at the C3 or C4 atom of proline, C5 of lysine and C3 of asparagine, thus forming •C-H and Fe³⁺-OH radicals.

Transfer of the hydroxyl radical •OH from Fe³⁺-OH to •C-H results in the hydroxylation of the amino acid side chain, which by itself is not a donor of electrons and does not

Fig. 22. Structure of monooxygenated proline, lysine, and asparagine residues

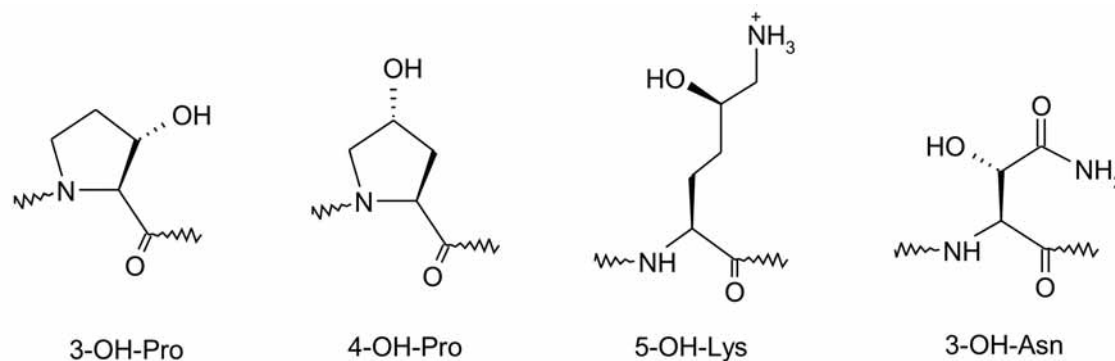
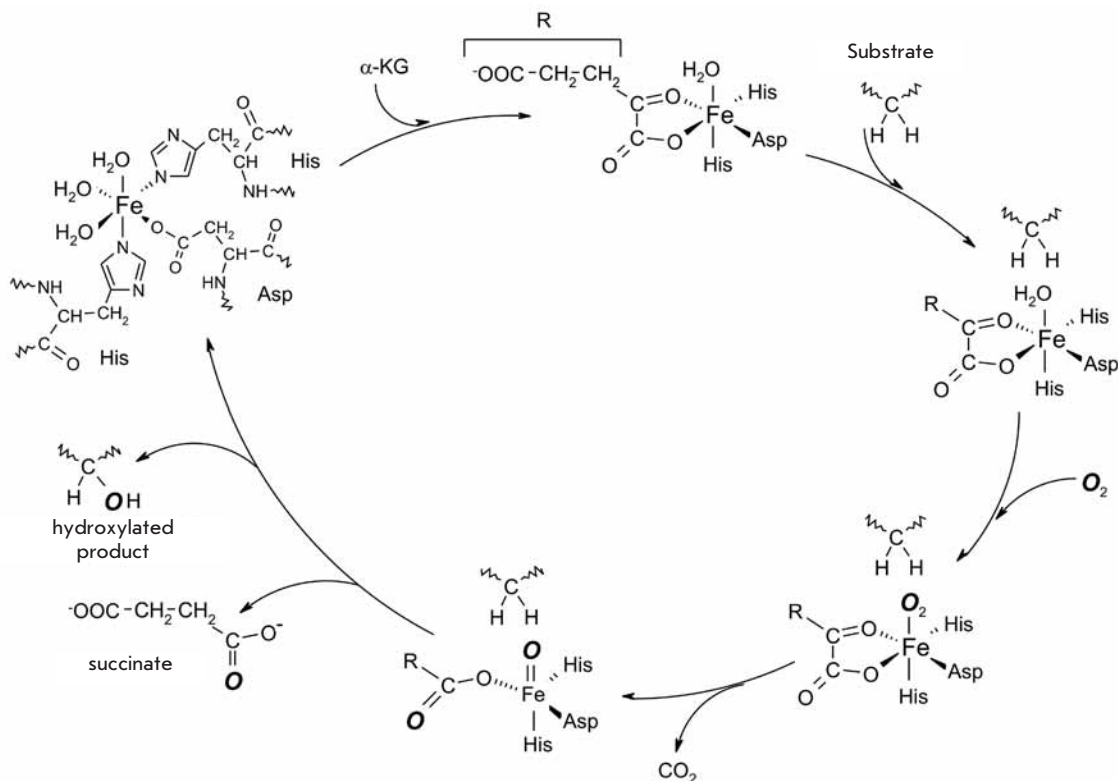


Fig. 23. Mechanism of the hydroxylation reaction



act as a nucleophile in this reaction. Monooxygenases, which catalyze hydroxylation reactions, attach the hydroxide radical in a stereospecific manner.

POSTTRANSLATIONAL CARBOXYLATION OF THE GLUTAMIC ACID RESIDUE

Most protein factors, which are involved in blood clotting in mammals, contain several residues of γ -carboxyglutamic acid (Gla). This residue appears in blood clotting factors as a result

of posttranslational modification; namely the fixation of CO_2 by the γ -methylene carbon atom of glutamic acid (Glu) during the factor's progress down the secretion pathways [164–166]. The Gla residue side chain, which bears two negatively charged carboxyl groups, has a capacity to form chelate complexes with bivalent cations, which is especially important for interaction with the Ca^{2+} ion [164].

Gla can be found in such proteins as prothrombin and blood clotting factors IX and X, which are proenzymatic

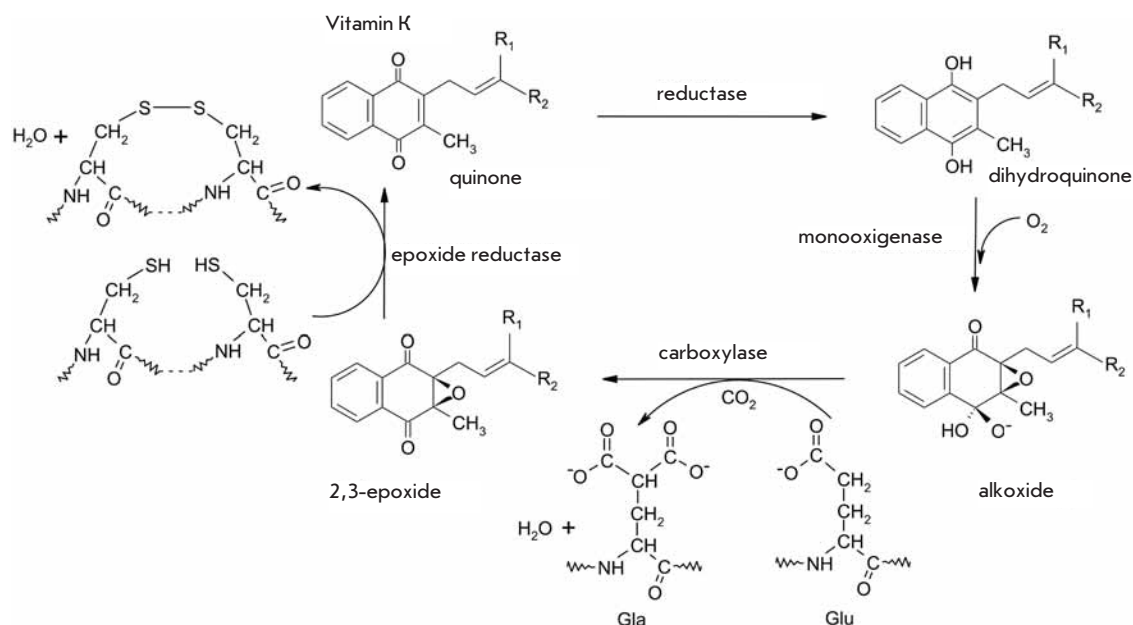


Fig. 24. Vitamin K-dependent carboxylation of a glutamic acid residue catalyzed by γ -glutamylcarboxylase. The 2,3-epoxide of vitamin K is reduced by vitamin K 2,3-epoxide reductase

REVIEWS

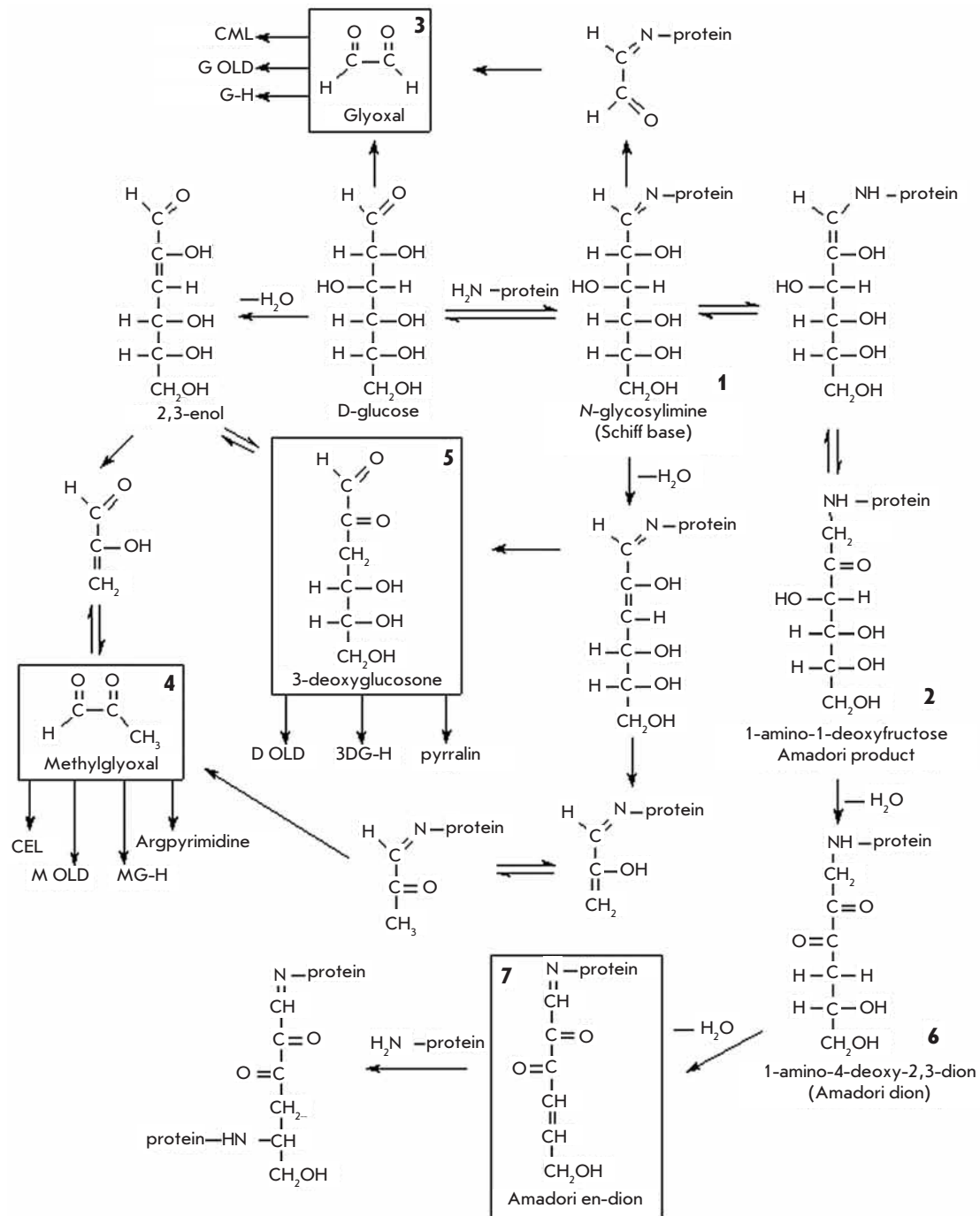


Fig. 25. Glycation of proteins in the presence of D-glucose. The rectangles show the main precursors of AGEs, which are formed during glycation

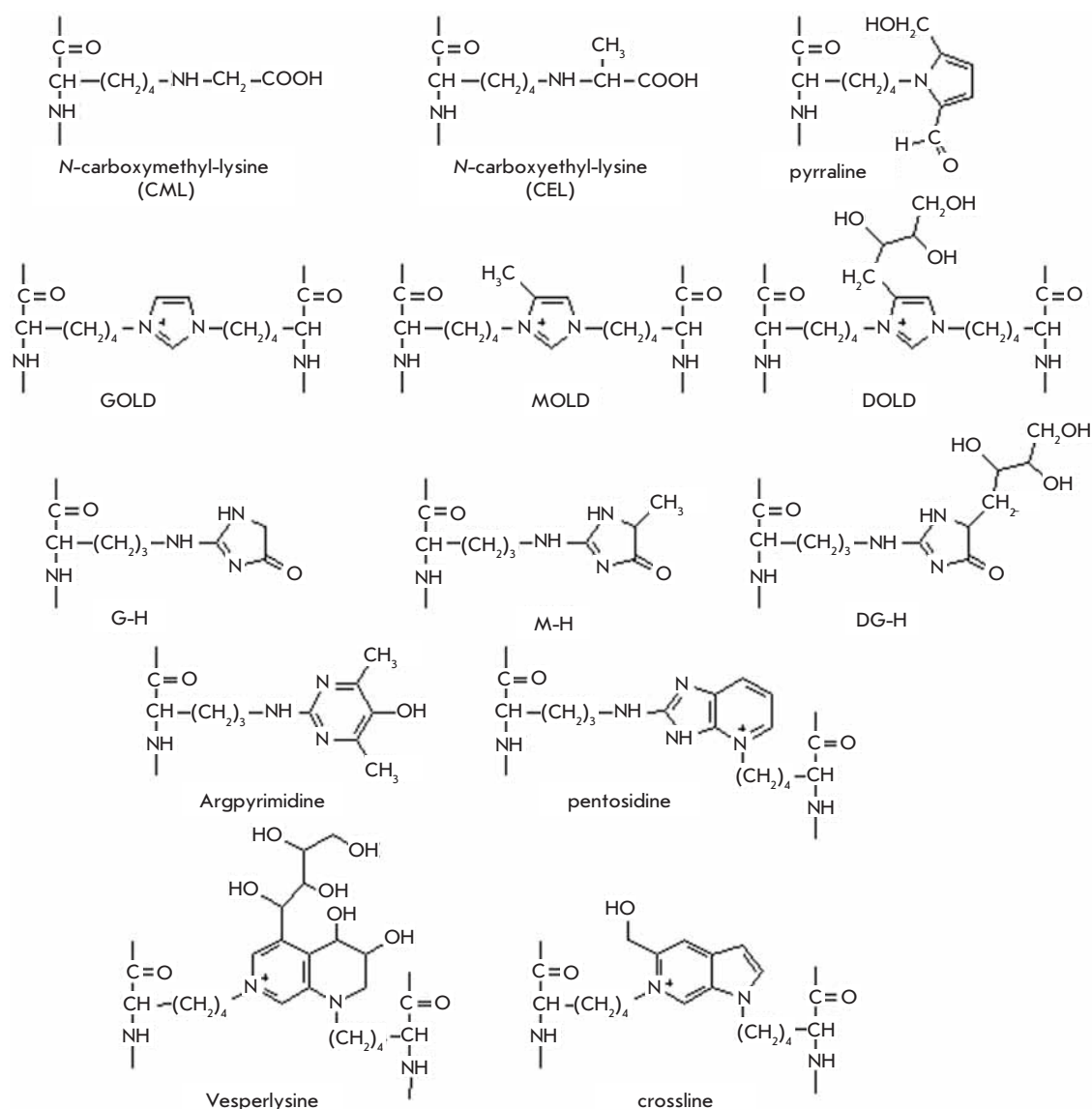
forms of proteases [164]. Carboxylation of 10–12 Glu residues in the *N*-terminal region of the proenzymes in a sequence of up to 40 amino acids leads to the binding of several Ca^{2+} ions and to conformation alteration of the blood clotting factors, which then associate on the surface of platelets adjacent to the proteases, which activate the factors by partial proteolysis and initiate the blood clotting cascade [164–166].

Carboxylation of the glutamic acid residue is catalyzed by the γ -glutamylcarboxylase (E.C. 1.14.99.20), which uses

the reduced (dihydronephthochinol) form of vitamin K (Fig. 24) [1, 164–166]. The oxidation of the reduced form of vitamin K by oxygen results in the formation of a hyperperoxide adduct of vitamin K, which forms a cyclic alkoxide anion, 2,3-epoxide of vitamin K, and generates a strong base, which captures a proton from the γ -methylene carbon atom of glutamic acid. The formed carbanion attacks the carbon atom of CO_2 and forms a new C-C bond in the malonyl side chain of the Gla residue. Reduction of the 2,3-epoxide of vi-

REVIEWS

Fig. 26. Structure of certain AGEs formed as a result of *in vivo* protein modification by D glucose



tamin K into its original form is catalyzed by 2,3-epoxydere-ductase (E.C. 1.1.4.1), which is associated in a complex with protein disulfide isomerase in the endoplasmic reticulum (E.C. 1.8.4.2) [167].

NON-ENZYMATIC MODIFICATION OF FUNCTIONAL GROUPS IN PROTEINS

PROTEIN GLYCATION

Protein glycation is an endogenous non-enzymatic addition of reducing sugar residues present in the bloodstream to the side chains of either lysine or arginine residues in proteins. A schematic representation of the glycation process, which can be divided into the early and late stages, is shown on Fig. 25. The first stage of glycation involves the nucleophilic attack of the glucose carbonyl group by an ϵ -amino group of lysine or a guanidine moiety of arginine, which results in the formation of a labile Schiff base - *N*-glycosylimine (1). The formation

of the Schiff base is a relatively rapid and reversible process [168]. Next, the glycosylimine regroups and forms an Amadori product, 1-amino-1-deoxyfructose (2). This process happens more slowly than the formation of glycosylimine, but much quicker if compared to the rate of Schiff base hydrolysis. This is why proteins bearing 1-amino-1-deoxyfructose residues tend to accumulate in blood. Modification of lysine residues at the early glycation steps is thought to be facilitated by the close proximity of histidine or lysine residues, which catalyze this process [169].

The late stage of glycation, which involves transformations of the *N*-glycosylimine and the Amadori product, is a slower and less studied process. It results in the formation of stable, advanced glycation end-products (AGEs) (Fig. 26). There are published data [170] on the direct involvement of α -dicarbonyl compounds in AGE formation (glyoxal (3), methylglyoxal (4), and 3-deoxyglucosone (5)). These compounds form *in vivo* both during glucose degra-

REVIEWS

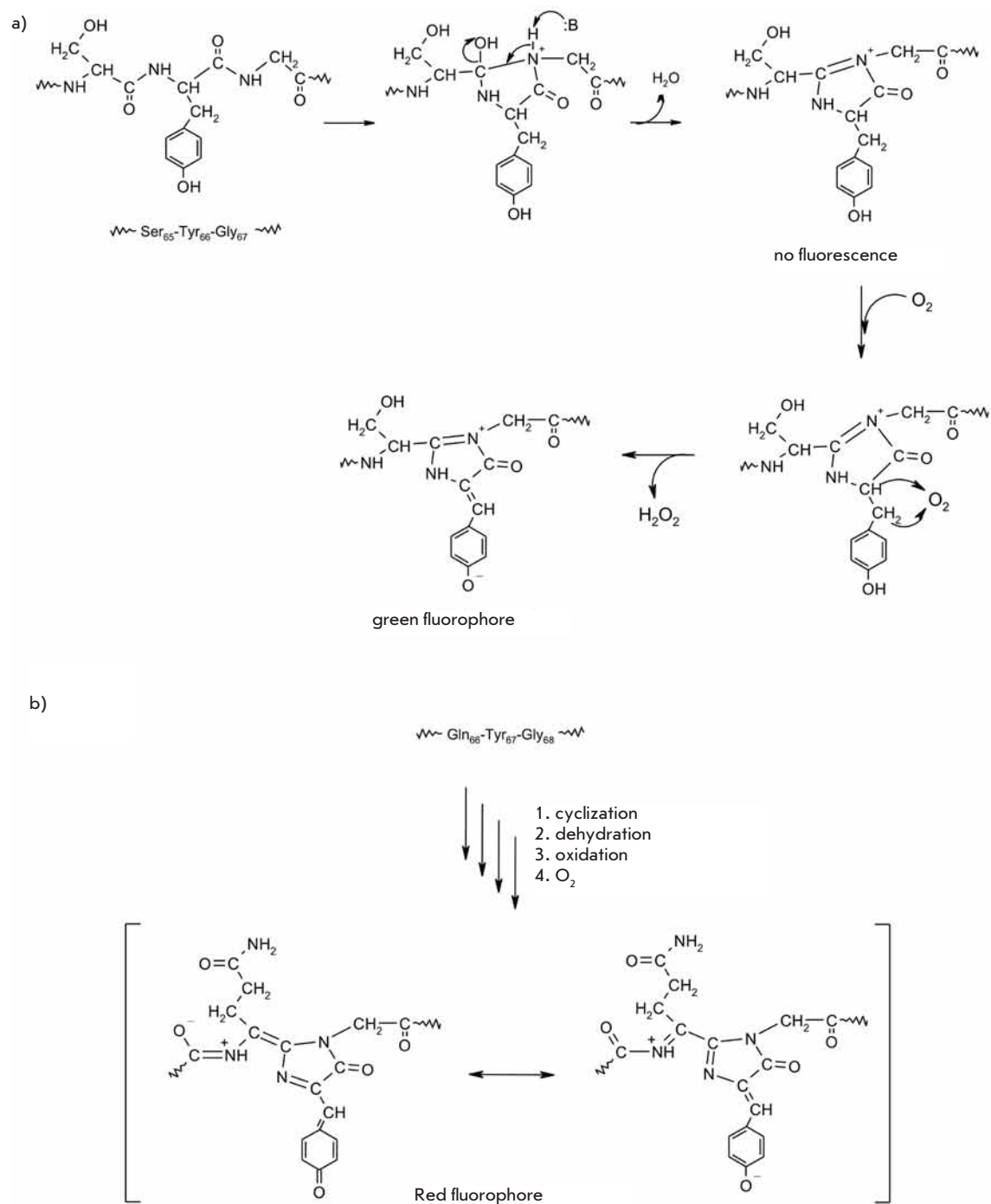


Fig. 27. Formation of (a) green and (b) red chromophores in proteins from tripeptides by intramolecular posttranslational autocatalytic cyclization

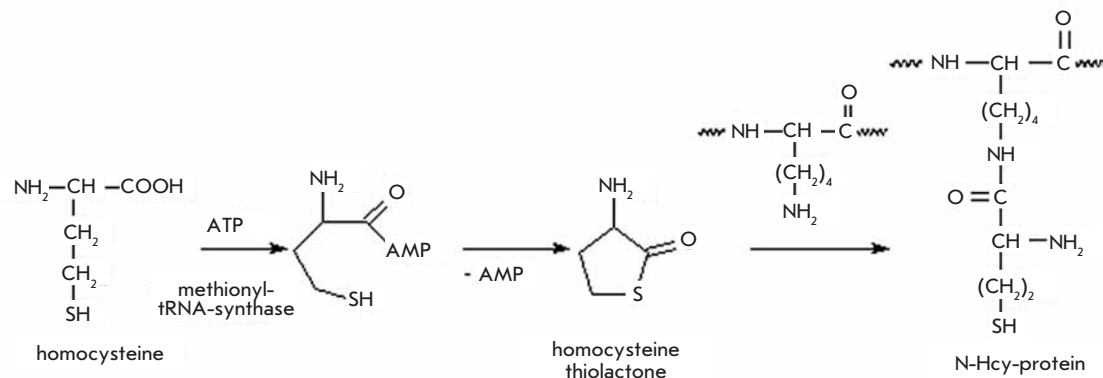
ation and in the transformations of the Schiff base during the modification of lysine residues in proteins by glucose (Fig. 25).

Reactions between α -dicarbonyl compounds and the ϵ -amino groups of lysine residues or the guanidinium groups of arginine in proteins result in the formation of protein crosslinks, which lead to complications caused by the protein glycation seen in diabetes and other diseases. Moreover, sequential dehydration of the Amadori product results in the formation of a 1-amino-4-deoxy-2,3-dion (**6**) and en-dion (**7**) at the C4

and C5 atoms, respectively (Fig. 25). These side chains can form intra- and intermolecular protein crosslinks [170].

Some AGEs have been characterized, including N_ϵ -carboxymethyl-lysine (CML) and N -carboxyethyl-lysine (CEL) [171], bis(lysyl)imidazole adducts (GOLD, MOLD and DOLD) [172], imidazolones (G-H, MG-H и 3DG-H) [173, 174], pyrrole [175], argpyrimidine [176], pentosidine [177], crossline [178], and vesperlysine [179] (Fig. 26). Among these pentosidine, crossline and vesperlysine are fluorophores, and their fluorescence emission maximum ($\lambda_{em} = 440$ nm) is shift-

Fig. 28. N-homocysteinylation of proteins by the homocysteine thiolactone



ed into the long-wave region, compared to tryptophan residue fluorescence in proteins [180]. This property of AGEs allows to monitor the glycation reaction progress by measuring the fluorescence at the excitation wavelength characteristic of the forming fluorophore glycation products (glucophores).

INTRAMOELCULAR POSTTRANSLATIONAL AUTOCATALYTIC CYCLIZATION

A very impressive type of posttranslational modification is the autocatalytic restructuring of the peptide backbone in the folded protein during GFP (green fluorescent protein) maturation. This protein is encoded by a single gene, and the chromofore is made up of three amino acid residues, Ser65-Tyr66-Gly67, capable of posttranslational autocatalytic cyclization, which does not require any cofactors or substrates [181–183].

Formation of the chromophore requires that the precursor take on the form of a β -barrel. This folded and colorless GFP-precursor bears the Ser65-Tyr66-Gly67 tripeptide in a spatially squeezed conformation in which the amide of Gln-67 can attack the peptide carbonyl and form a pentatomic tetrahedral adduct (Fig. 27, a). Then, this adduct is dehydrated, and the stable cyclic intermediate product slowly autooxidizes, forming a double bond coupled to the phenol ring of Tyr-66. This last oxidation reaction produces a chromophore with an excitation maximum of 506 nm.

GFP is used as an *in vivo* vital marker, which allows the study of various processes taking place in live cells and organisms [184–186]. Fusion proteins based on GFP are used in novel drug screenings [187, 188], apoptosis detection [189], in the visualization of chromosome dynamics [190], and in many other applications [191, 192]. Several volumes of Methods in

Enzymology [193] and Methods in Cell Biology [194] are dedicated to GFP. The discovery of fluorescent genetic markers was awarded the Nobel Prize in 2008.

During the last decade, the number of studies with other colored proteins similar to GFP but extracted from coral has been steadily growing [195–197]. A drawback of these proteins is their marked propensity to aggregate, which however can be rectified by mutagenesis [198]. A schematic representation of the formation of a red fluorophore from the Gln66-Tyr67-Gly68 tripeptide in a protein molecule is shown in Fig. 27, b.

PROTEIN HOMOCYSTEINYLATION

The majority of methylation processes in live organisms use S-adenosylmethionine, thus forming S-adenosylhomocysteine. The latter is hydrolyzed by the adenosylhomocysteinase (E.C. 3.3.1.1) enzyme into adenosine and homocysteine. This reaction catalyzed by methionyl-tRNA synthetase (E.C. 6.1.1.1) turns homocysteine into thiolactone (this is a side reaction for this enzyme) [199]. Homocysteine thiolactone is an acylating agent and can react with the functional groups of lysine residues [200–203]. The ϵ -amino group of lysine performs a nucleophilic attack of the carbonyl carbon atom of the thiolactone, which results in decyclization of the lactone and the formation of an additional sulfhydryl moiety (Fig. 28).

This type of modification is characteristic of blood proteins (albumin, hemoglobin, transferrin, and globulins) [204–207]. Ninety percent of the homocysteine in human blood plasma is incorporated into N-homocysteinyated serum albumine (HSA) [201]. It is known that the main HSA homocysteinylation site both *in vitro* and *in vivo* is the Lys-525 residue [208]. Further-

Fig. 29. Protein S-homocysteinylation

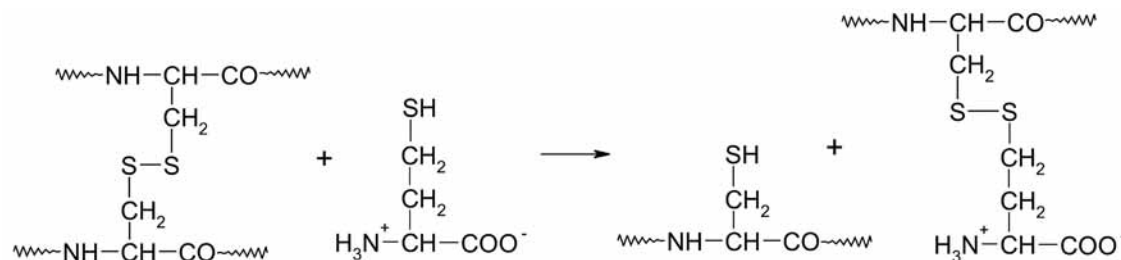
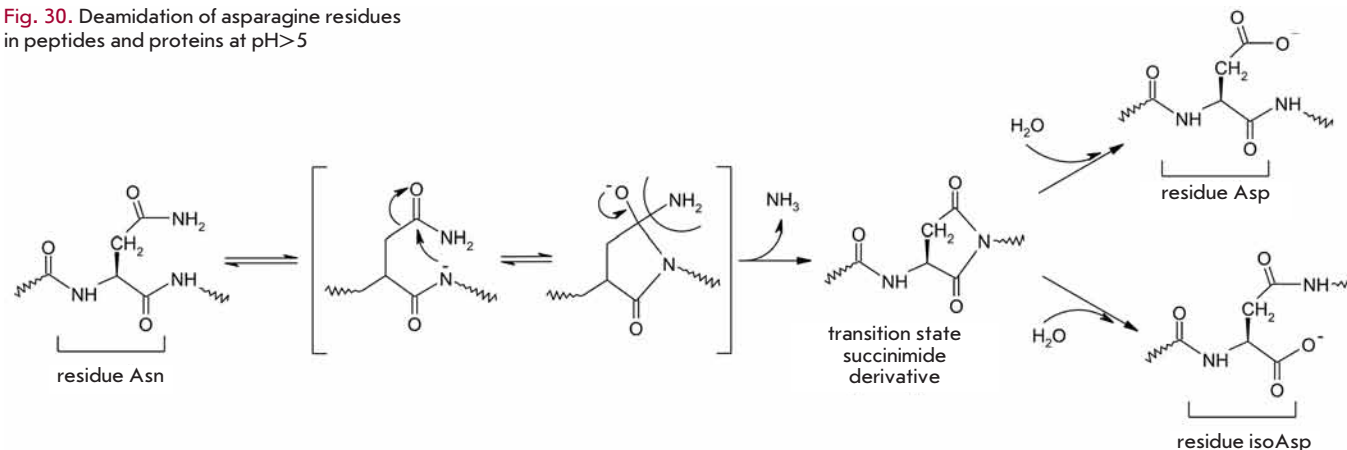


Fig. 30. Deamidation of asparagine residues in peptides and proteins at pH>5



more, two additional albumin modification sites were discovered at Lys-4 and Lys-12 [209].

Homocysteine can take part in disulfide exchange reactions with S-S bonds in proteins, thus forming S-homocysteinylated proteins (Fig. 29) [200, 202, 206, 207, 210].

Homocysteinylation of proteins has a considerable effect on their biological activity, including increased sensitivity to oxidation and increased propensities for oligomerization, denaturation, and sedimentation. The introduction of 8–9 homocysteine residues into the methionyl-tRNA-synthase and 11–12 residues into trypsin completely deactivates these proteins [207]. N-homocysteinylation of human serum albumin lowers its RNA-hydrolyzing activity considerably [205]. Multiple homocysteinylation of cellular proteins can eventually result in cell apoptosis [200, 201, 203, 206, 210].

DEAMIDATION AND TRANSAMIDATION

One of the types of posttranslational modification, which plays an important role in cellular functions, is the deamidation of the amides of dicarbonic acids. Many authors believe these reactions to be non-enzymatic cleavage of ammonia from the amide group of asparagine or glutamine, resulting in an intermediate product, a cyclic imide (Fig. 30) [211–215]. The rate of this product's formation is determined by the local amino acid surroundings and the characteristics of the solution (pH and ingredients) [213, 214]. Asparagine residues in proteins are deamidated 40 times more often than glutamine residues. Furthermore, the rate of asparagine deamidation is 100-fold greater than the rate of glutamine deamidation [214].

The cyclic imide decays forming either aspartate residue, which forms in the largest quantities (3:1), or an isoaspartate residue, in which the peptide bond involves the β-carboxyl group of the aspartate side chain [216, 217]. In the latter case, the length of the protein increases by one methylene group (CH₂), which can influence the structure and the functioning of the protein, including its stability [214, 216, 217].

Deamidation reactions result in the formation of an ionizable carboxyl group charged negatively under physiological conditions, which alters the overall charge of the protein molecule and its spatial structure [214].

The β-isopeptide bond formed by lysine and glutamine side chains is considered by the organism to be an aberration

of a normal peptide bond, which is formed by the α-amino groups and carboxyl groups of amino acids, and is corrected by the protein isoaspartyl-O-methyltransferase (PIMT) (E.C. 2.1.1.77), a widespread cellular enzyme [211, 212, 216]. The deamidation reaction of Asn/Gln and a deficit of PIMT cause serious illnesses in humans, such as cataract [218], Alzheimer's disease [219], autoimmune diseases [220], and prion-dependent encephalopathy [214, 221, 222].

According to Robinson's hypothesis, the instability of the asparagine and glutamine residues in cellular proteins under physiological conditions determines a key biological function, which is a programmed biological clock mechanism limiting the lifespan of proteins and peptides [212, 223, 224].

Deamidation, as well as ADP-ribosylation, can be caused by bacterial toxins. The cytotoxic necrotic factor 1 from *Escherichia coli* (CNF1) and the dermonecrotic toxin (DNT) from *Bordetella* deamidate small GTPases in the human organism, such as Rho A (Gln63), Rac1, and Cdc42 (Gln61), which results in blockage of GTP hydrolysis and disorders in the regulation of cytoskeleton remodeling [225–228].

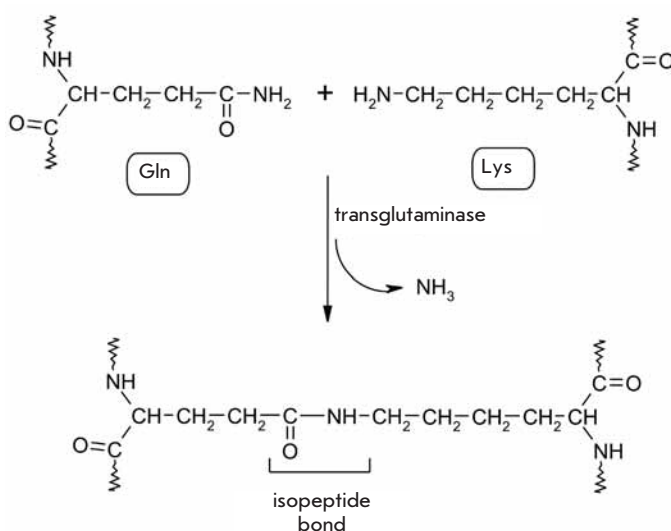


Fig. 31. Transamidation catalyzed by transglutaminase (E.C. 2.3.2.13)

Deamidation is often coupled with subsequent transamidation (interaction of the ϵ -amino group of a lysine residue with the side chain of a glutamine residue in the same protein molecule), which is one of the types of crosslinks characteristic of posttranslational modification (Fig. 31) [228–232].

This process leads to the formation of multiple bonds between glutamine and lysine residues in protein molecules, which results in a massive protein aggregate whose subunits

are cross-linked. This is an important process in the metabolism of skin and hair and also during the healing of wounds [233]. ●

This work on the effects of chemical modification of human serum albumin on the RNA-hydrolytic activity of the protein was performed with support from the Interdisciplinary Integration Project for Basic Research Siberian Branch RAS №88 and the Russian Foundation for Basic Research (grant № 09-04-01483a).

REFERENCES

1. Walsh C.T., Garneau-Tsodikova S., Gatto G.J. *Angew. Chem. Int. Ed.* 2005. V. 44. № 45. P. 7342–7372.
2. Lehninger A. *Principles of Biochemistry*. New York: W.H. Freeman and Company. 2008.
3. Karpesky M.Y., Ivanov V.I. *Nature*. 1966. V. 210. № 30. P. 493–496.
4. Lowe J.N., Ingraham L.L. *An Introduction to Biochemical Reactions Mechanisms*. Chap. 3. *Foundation of Molecular Biology Series*. New Jersey: Prentice-Hall, Englewood Cliffs. 1974.
5. Hubbard S.R., Miller W.T. *Curr. Opin. Cell Biol.* 2007. V. 19. № 2. P. 117–123.
7. Übersax J.A., Ferrell J.E.Jr. *Rev. Mol. Cell. Biol.* 2007. V. 8. № 7. P. 530–541.
6. Beene D.L., Scott J.D. *Curr. Opin. Cell Biol.* 2007. V. 19. № 2. P. 192–198.
9. Alemany R., Perona J.S., Sanchez-Dominguez J.M., Montero E., Canizares J., Escriba P.V. *Biochim. Biophys. Acta.* 2007. V. 1768. № 4. P. 964–975.
10. Saltiel A.R., Pessin J.E. *Trends Cell Biol.* 2002. V. 12. № 2. P. 65–71.
11. Maures T.J., Kurzer J.H., Carter-Su C. *Trends Endocrinol. Metab.* 2007. V. 18. № 1. P. 38–45.
12. Patwardhan P., Miller W.T. *Cell. Signal.* 2007. V. 19. № 11. P. 2218–2226.
13. Lieser S.A., Aubol B.E., Wong L., Jennings P.A., Adams J.A. *Biochim. Biophys. Acta.* 2005. V. 1754. № 1–2. P. 191–199.
14. Bublib E.M., Yarden Y. *Curr. Opin. Cell Biol.* 2007. V. 19. № 2. P. 124–134.
15. Dorsam R.T., Gutkind J.S. *Rev. Cancer.* 2007. V. 7. № 2. P. 79–94.
16. Viallard J.F., Lacombe F., Belloc F., Pellegrin J.L., Reiffers J. *Cancer Radiother.* 2001. V. 5. № 2. P. 109–129.
17. Syeed A.S., Vohra H., Gupta A., Ganguly N. *Curr. Science.* 2001. V. 80. № 3. P. 349–360.
18. Nakagami H., Pitzschke A., Hirt H. *Trends Plant Sci.* 2005. V. 10. № 7. P. 339–346.
19. Chau B.N., Wang J.Y.J. *Nat. Rev. Cancer.* 2003. V. 3. P. 130–138.
20. Johnson L.N., Lewis R.J. *Chem. Rev.* 2001. V. 101. № 8. P. 2209–2242.
21. Birnbaumer L. *Biochim. Biophys. Acta.* 2007. V. 1768. № 4. P. 756–771.
22. Cooper D.M.F., Crossthwaite A.J. *Trends Pharmacol. Sci.* 2006. V. 27. № 8. P. 26–431.
23. Willoughby D., Cooper D.M.F. *Physiol. Rev.* 2007. V. 87. P. 965–1010.
24. Deng X., Mercer P.F., Scotton C., Gilchrist A., Chambers R.C. *Mol. Biol. Cell.* 2008. V. 19. № 6. P. 2520–2533.
25. Xu Y. *Cell Death Differ.* 2003. V. 10. № 4. P. 400–403.
26. Lindner H.H. *Electrophoresis*, 2008. V. 29. № 12. P. 2516–2532.
27. Sarg B., Chwatal S., Talasz H., Lindner H.H. *J. Biol. Chem.* 2009. V. 284. № 6. P. 3610–3618.
28. Dahmus M.E. *J. Biol. Chem.* 1996. V. 271. № 32. P. 19009–19012.
29. Lee T.I., Young R.A. *Annu. Rev. Genet.* 2000. 34. P. 77–137
30. Bottomley M.J. *EMBO Rep.* 2004. V. 5. № 5. P. 464–469.
31. Owen D.J., Ornaghi P., Yang J.-C., Lowe N., Evans P.R., Ballario P., Neuhaus D., Filetici P., Travers A.A. *EMBO J.* 2000. V. 19. № 22. P. 6141–6149.
32. Mukherjee S., Hao Y.-H., Orth K. *Trends Biochem. Sci.* 2007. V. 32. № 5. P. 210–216.
33. Margueron R., Trojer P., Reinberg D. *Curr. Opin. Gen. Develop.* 2005. V. 15. № 2. P. 163–176.
34. Shen S., Casaccia-Bonnel P.J. *Mol. Neurosci.* 2008. V. 35. № 1. P. 13–22.
35. Iizuka M., Smith M.M. *Curr. Opin. Gen. Develop.* 2003. V. 13. № 2. P. 154–160.
36. Couture J.F., Trievel R.C. *Curr. Opin. Struct. Biol.* 2006. V. 16. № 6. P. 753–760.
37. Cole P.A. *Nat. Chem. Biol.* 2008. V. 4. № 10. P. 590–597.
38. Strahl B.D., Allis C.D. *Nature*. 2004. V. 403. № 6765. P. 41–45.
39. Khorasanizadeh S. *Cell.* 2004. V. 116. № 2. P. 259–272.
40. Feng L., Lin T., Uranishi H., Gu W., Xu Y. *Mol. Cell. Biol.* 2005. V. 25. № 13. P. 5389–5395.
41. Desmeules P., Penney S.-E., Desbat B., Saless C. *Biophys. J.* 2007. V. 93. № 6. P. 2069–2082.
42. Farazi T.A., Waksman G., Gordon J.I. *J. Biol. Chem.* 2001. V. 276. № 43. P. 39501–39504.
43. Smotrys J.E., Linder M.E. *Annu. Rev. Biochem.* 2004. V. 73. P. 559–587.
44. Tanimura N., Saitoh S., Kawano S., Kosugi A., Miyake K. *Biochem. Biophys. Res. Commun.* 2006. V. 341. № 4. P. 1177–1183.
45. Resh M.D. *Biochim. Biophys. Acta.* 1999. V. 1451. № 1. P. 1–16.
46. Adams J.A. *Chem. Rev.* 2001. V. 101. № 8. P. 2271–2290.
47. Pechlivanis M., Kuhlmann J. *Biochim. Biophys. Acta.* 2006. V. 1764. № 12. P. 1914–1931.
48. Selvakumar P., Lakshmikuttyamma A., Shrivastava A., Das S.B., Dimmock J.R., Sharma K. *Progr. Lipid Res.* 2007. V. 46. № 1. P. 1–36.
49. Dietrich L.E.P., Ungermann C. *EMBO Rep.* 2004. V. 5. № 11. P. 1053–1057.
50. Drisdell R.C., Alexander J.C., Sayeed A., Green W.N. *Methods.* 2006. V. 40. № 2. P. 127–134.
51. Hemsley P. A., Grierson C.S. *Trends Plant Sci.* 2008. V. 13. № 6. P. 295–302.
52. Pickart C.M. *Annu. Rev. Biochem.* 2001. V. 70. P. 503–533.
53. Glickman M.H., Ciechanover A. *Physiol. Rev.* 2002. V. 82. № 2. P. 373–428.
54. Pickart C.M. *Cell.* 2004. V. 116. № 2. P. 181–190.
55. Finley D., Ciechanover A., Varshavsky A. *Cell.* 2004. V. 116. 2 Suppl. P. S29–S32.
56. Capili A.D., Lima C.D. *Curr. Opin. Struct. Biol.* 2007. V. 17. № 6. P. 726–735.
57. Pickart C. M., Eddins M. J. *Biochim. Biophys. Acta.* 2004. V. 1695. № 1–3. P. 55–72.
58. Herrmann J., Lerman L.O., Lerman A. *Circ. Res.* 2007. V. 100. № 9. P. 1276–1291.

59. Schwartz D.C., Hochstrasser M. *Trends Biochem. Sci.* 2003. V. 28. № 6. P. 321–328.
60. Li W., Ye Y. *Cell. Mol. Life Sci.* 2008. V. 65. № 15. P. 2397–2406.
61. Pickart C.M., Fushman D. *Curr. Opin. Chem. Biol.* 2004. V. 8. № 6. P. 610–616.
62. Aguilar R.C., Wendland B. *Curr. Opin. Cell Biol.* 2003. V. 15. № 2. P. 184–190.
63. Pickart C.M. *Mol. Cell.* 2001. V. 8. № 3. P. 499–504.
64. Gill G. *Curr. Opin. Genet. Dev.* 2005. V. 15. № 5. P. 536–541.
65. Smith B.C., Denu J.M. *Biochim. Biophys. Acta.* 2009. V. 1789. № 1. P. 45–57.
66. Scoumanne A., Chen X. *Histol. Histopathol.* 2008. V. 23. № 9. P. 1143–1149.
67. Marmorstein R., Trievel R.C. *Biochim. Biophys. Acta.* 2009. V. 1789. № 1. P. 58–68.
68. Berger S.L. *Curr. Opin. Genet. Dev.* 2002. V. 12. № 2. P. 142–148.
69. Loyola A., Almouzni G. *Trends Biochem. Sci.* 2007. V. 32. № 9. P. 425–433.
70. Lachner M., Jenwein T. *Curr. Opin. Cell Biol.* 2002. V. 14. № 3. P. 286–298.
71. de La Roche Saint-André C. *Biochimie.* 2005. V. 87. № 7. P. 603–612.
72. Kiefer J.C. *Develop. Dynamics.* 2007. V. 236. № 4. P. 1144–1156.
73. Lane K.T., Beese L.S. *J. Lipid Res.* 2006. V. 47. № 4. P. 681–699.
74. Leung K.F., Baron R., Seabra M.C. *J. Lipid Res.* 2006. V. 47. № 3. P. 467–475.
75. Pylypenko O., Rak A., Durek T., Kushnir S., Dursina B.E., Thomae N.H., Constantinescu A.T., Brunsfeld L., Watzke A., Waldmann H., Goody R.S., Alexandrov K. *EMBO J.* 2006. V. 25. № 1. P. 13–23.
76. Lu J.-Y., Hofmann S.L. *J. Lipid Res.* 2006. V. 47. № 7. P. 1352–1357.
77. Kinsella B.T., Erdman R.A., Maltese W.A. *Proc. Natl. Acad. Sci. USA.* 1991. V. 88. № 20. P. 8934–8938.
78. Maltese W.A. *FASEB J.* 1990. V. 4. № 15. P. 3319–3328.
79. Basso A.D., Kirschmeier P., Bishop W.R. *J. Lipid Res.* 2006. V. 47. № 1. P. 15–31.
80. Magee T., Seabra M.C. *Curr. Opin. Cell Biol.* 2005. V. 17. № 2. P. 190–196.
81. Lan F., Nottke A.C., Shi Y. *Curr. Opin. Cell Biol.* 2008. V. 20. № 3. P. 316–325.
82. Shi Y., Lan F., Matson C., Mulligan P., Whetstone J.R., Cole P.A., Casero R.A. *Cell.* 2004. V. 119. № 7. P. 941–953.
83. Tsukada Y., Fang J., Erdjument-Bromage H., Warren M. E., Borchers C.H., Tempst P., Zhang Y. *Nature.* 2006. V. 439. № 7078. P. 811–816.
84. Baron R.A., Seabra M.C. *Biochem. J.* 2008. V. 415. № 1. P. 67–75.
85. Roberts P.J., Mitin N., Keller P.J., Chenette E.J., Madigan J.P., Currin R.O., Cox A.D., Wilson O., Kirschmeier P., Der C.J. *J. Biol. Chem.* 2008. V. 283. № 37. P. 25150–25163.
86. Wright L.P., Phillips M.R. *J. Lipid Res.* 2006. V. 47. № 5. P. 883–891.
87. Leung K.F., Baron R., Ali B.R., Magee A.I., Seabra M.C. *J. Biol. Chem.* 2007. V. 282. № 2. P. 1487–1497.
88. Lehle L., Strahl S., Tanner W. *Angew. Chem. Int. Ed.* 2006. V. 45. № 41. P. 6802–6818.
89. Lowe J.B., Marth J.D. *Annu. Rev. Biochem.* 2003. V. 72. P. 643–691.
90. Parodi A.J. *Annu. Rev. Biochem.* 2000. V. 69. P. 69–93.
91. Parodi A.J. *Biochem. J.* 2000. V. 348. Pt. 1. P. 1–13.
92. Caramelo J.J., Parodi A. *J. Semin. Cell. Dev. Biol.* 2007. V. 18. № 6. P. 732–742.
93. Deprez P., Gautschi M., Helenius A. *Mol. Cell.* 2005. V. 19. № 2. P. 183–195.
94. Deigaard S., Nicolay J., Taheri M., Thomas D.Y., Bergeron J.J.D. *Curr. Issues Mol. Biol.* 2004. V. 6. № 1. P. 29–42.
95. Roth J. *Chem. Rev.* 2002. V. 102. № 2. P. 285–303.
96. Lis H., Sharon N. *Eur. J. Biochem.* 1993. V. 218. № 1. P. 1–27.
97. Zachara N.E., Hart G.W. *Biochim. Biophys. Acta.* 2006. V. 1761. № 5–6. P. 599–617.
98. Wells L., Whelan S.A., Hart G.W. *Biochem. Biophys. Res. Commun.* 2003. V. 302. № 3. P. 435–441.
99. Goto M. *Biosci. Biotechnol. Biochem.* 2007. V. 71. № 6. P. 1415–1427.
100. Gerken T.A., Gilmore M., Zhang J. *J. Biol. Chem.* 2002. V. 277. № 10. P. 7736–7751.
101. Ohtsubo K., Marth J.D. *Cell.* 2006. V. 126. № 5. P. 855–867.
102. Haines N., Irvine K.D. *Nat. Rev. Mol. Cell. Biol.* 2003. V. 4. № 10. P. 786–797.
103. Bojarova P., Williams S.J. *Curr. Opin. Chem. Biol.* 2008. V. 12. № 5. P. 573–581.
104. Chapman E., Best M.D., Hanson S.R., Wong C.-H. *Angew. Chem. Int. Ed.* 2004. V. 43. № 27. P. 3526–3548.
105. Ghosh D. *Cell. Mol. Life Sci.* 2007. V. 64. № 15. P. 2013–2022.
106. Hanson S.R., Best M.D., Wong C.-H. *Angew. Chem. Int. Ed.* 2004. V. 43. № 43. P. 5736–5763.
107. Koch-Nolte F., Adriouch S., Bannas P., Krebs C., Scheuplein F., Seman M., Haag F. *Ann. Med.* 2006. V. 38. № 3. P. 188–199.
108. Sakurai J., Nagahama M., Hisatsune J., Katunuma N., Tsuge H. *Advan. Enzyme Regul.* 2003. V. 43. P. 361–377.
109. Tsuge H., Nagahama M., Nishimura H., Hisatsune J., Sakaguchi Y., Itogawa Y., Katunuma N., Sakurai J. *J. Mol. Biol.* 2003. V. 325. № 3. P. 471–483.
110. Holbourn K.P., Sutton J.M., Evans H.E., Shone C.C., Acharya K.R. *Proc. Natl. Acad. Sci. USA.* 2005. V. 102. № 15. P. 5357–5362.
111. Krueger K.M., Barbieri J.T. *Clin. Microbiol. Rev.* 1995. V. 8. № 1. P. 34–47.
112. Kaslow H.R., Lim L.K., Moss J., Lesikar D.D. *Biochemistry.* 1987. V. 26. № 1. P. 123–127.
113. Spangler B.D. *Microbiol. Rev.* 1992. V. 56. № 4. P. 622–647.
114. Collier R.J. *Bacteriol. Rev.* 1975. V. 39. № 1. P. 54–85.
115. Jorgensen E., Merrill A.R., Andersen G.R. *Biochem. Soc. Trans.* 2006. V. 34. Pt. 1. P. 1–6.

REVIEWS

116. Jorgensen R, Merrill A.R., Yates S.P., Marquez V.E., Schwan A.L., Boesen T., Andersen G.R. *Nature*. 2005. V. 436. № 7053. P. 979–984.
117. Yates S.P., Jorgensen R., Andersen G.R., Merrill A.R. *Trends Biochem. Sci.* 2006. V. 31. № 2. P. 123–133.
118. Lindahl T. *Nature*. 1993. V. 362. № 6422. P. 709–715.
119. Sharer O.D. *Angew. Chem. Int. Ed.* 2003. V. 42. P. 2946–2974.
120. Bernstein C., Bernstein H., Payne C.M., Garewal H. *Mutat. Res.* 2002. V. 511. № 2. P. 145–178.
121. Althaus F.R., Kleczowska H.E., Malanga M., Muntener C.R., Pleschke J.M., Ebner M., Auer B. *Mol. Cell. Biochem.* 1999. V. 193. № 1–2. P. 5–11.
122. Hassa P.O., Haenni S.S., Elser M., Hottiger M.O. *Microbiol. Mol. Biol. Rev.* 2006. V. 70. № 3. P. 789–829.
123. D'Amours D., Desnoyers S., D'Silva I., Poirier G.G. *Biochem J.* 1999. V. 342. Pt. 2. P. 249–268.
124. Nguewa P.A., Fuertes M.A., Valladares B., Alonso C., Perez J.M. *Proc. Biophys. Mol. Biol.* 2005. V. 88. № 1. P. 143–172.
125. Schreiber V., Dantzer F., Ame J.C., de Murcia G. *Natl. Rev. Mol. Cell. Biol.* 2006. V. 7. № 7. P. 517–528.
126. Fan J., Wilson D.M. *Free Radic. Biol. Med.* 2005. V. 38. № 9. P. 1121–1138.
127. Sallmann F.R., Vodenicharov M.D., Wang Z.Q., Poirier G.G. *J. Biol. Chem.* 2000. V. 275. № 20. P. 15504–15511.
128. Ame J.C., Spelnhauer C., de Murcia G. *BioEssays*. 2004. V. 26. № 8. P. 882–893.
129. Kim M.Y., Zhang T., Kraus W.L. *Genes Dev.* 2005. V. 19. № 17. P. 1951–1967.
130. Yamanaka H., Penning C.A., Willis E.H., Wasson D.B., Carson D.A. *J. Biol. Chem.* 1988. V. 263. № 8. P. 3879–3883.
131. Ikejima M., Marsischky G., Gill D.M. *J. Biol. Chem.* 1987. V. 262. № 36. P. 17641–17650.
132. Kawaichi M., Ueda K., Hayaishi O. *J. Biol. Chem.* 1980. V. 255. № 3. P. 816–819.
133. Ueda K., Kawaichi M., Okayama H., Hayaishi O. *J. Biol. Chem.* 1979. V. 254. № 3. P. 679–687.
134. Alvarez-Conzalez R. *J. Biol. Chem.* 1988. V. 263. № 33. P. 17690–17696.
135. Rolli V., O'Farrell M., Menissier de Murcia J., de Murcia G. *Biochemistry*. 1997. V. 36. № 40. P. 12147–12154.
136. Miwa M., Saikawa N., Yamaizumi Z., Nishimura S., Sugimura T. *Proc. Natl. Acad. Sci. U.S.A.* 1979. V. 76. № 2. P. 595–599.
137. Mendoza-Alvarez H., Alvarez-Conzalez R. *Biochemistry*. 1987. V. 26. № 11. P. 3218–3224.
138. Mendoza-Alvarez H., Alvarez-Conzalez R. *J. Biol. Chem.* 1993. V. 268. № 30. P. 22575–22580.
139. Zahradka P., Ebisuzaki K. *J. Biol. Chem.* 1984. V. 259. № 2. P. 986–995.
140. Lindahl T. *Mutat. Res.* 2000. V. 462. № 2–3. P. 129–135.
141. Tanuma S., Yagi T., Johnson G.S. *Arch. Biochem. Biophys.* 1985. V. 237. № 1. P. 38–42.
142. Hassa P.O., Haenni S.S., Elser M., Hottiger M.O. *Microbiol. Mol. Biol. Rev.* 2006. V. 70. № 3. P. 789–829.
143. Ying J., Clavreul N., Sethuraman M., Adachi T., Cohen R.A. *Free Radic. Biol. Med.* 2007. V. 43. № 8. P. 1099–1108.
144. Giles N.M., Giles G.I., Jacob C. *Biochem. Biophys. Res. Commun.* 2003. V. 300. № 1. P. 1–4.
145. Jones D.P., Go Y.-M., Anderson C.L., Ziegler T.R., Kinkade J.M., Kirilin W.G. *FASEB J.* 2004. V. 18. № 11. P. 1246–1248.
146. Go Y.-M., Jones D.P. *Biochim. Biophys. Acta.* 2008. V. 1780. № 11. P. 1273–1290.
147. Jacob C., Giles G.I., Giles N.M., Sies H. *Angew. Chem. Int. Ed.* 2003. V. 42. № 39. P. 4742–4758.
148. Giles N.M., Watts A.B., Giles G.I., Fry F.H., Littlechild J.A., Jacob C. *Chem. Biol.* 2003. V. 10. № 8. P. 677–693.
149. Kemp M., Go Y.-M., Jones D.P. *Free Radic. Biol. Med.* 2008. V. 44. № 6. P. 921–937.
150. Iwakiri Y., Satoh A., Chatterjee S., Toomre D.K., Chalouni C.M., Fulton D., Groszmann R.J., Shah V.H., Sessa W.C. *Proc. Natl. Acad. Sci. USA.* 2006. V. 103. № 52. P. 19777–19782.
151. Yang Y., Loscalzo J. *Proc. Natl. Acad. Sci. USA.* 2005. V. 102. № 1. P. 117–122.
152. Doctor A., Platt R., Sheram M.L., Eischeid A., McMahon T., Maxey T., Doherty J., Axelrod M., Kline J., Gurka M., Gow A., Gaston B. *Proc. Natl. Acad. Sci. USA.* 2005. V. 102. № 16. P. 5709–5714.
153. Torta F., Uselli V., Malgaroli A., Bachi A. *Proteomics*. 2008. V. 8. № 21. P. 4484–4494.
154. Stampler J.S., Lamas S., Fang F.C. *Cell.* 2001. V. 106. № 6. P. 675–683.143.
155. Myllyharju J., Kivirikko K.I. *Trends Biochem. Sci.* 2004. V. 29. № 1. P. 33–43.
156. Koivunen P., Hirsila M., Gunzler V., Kivirikko K.I., Myllyharju J. *J. Biol. Chem.* 2004. V. 279. № 11. P. 9899–9904.
157. Lando D., Peet D.J., Whelan D.A., Gorman J.J., Whitelaw M.L. *Science*. 2002. V. 295. № 5556. P. 858–861.
158. Clifton I.J., Hsueh L.C., Baldwin J.E., Harlos K., Schofield C.J. *Eur. J. Biochem.* 2001. V. 268. № 24. P. 6625–6636.
159. Bruick R.K., McKnight S.L. *Science*. 2001. V. 294. № 5545. P. 1337–1340.
160. Ratcliffe P.J. *Blood Purif.* 2002. V. 20. № 5. P. 445–450.
161. Marxsen J.H., Stengel P., Doege K., Heikinen P., Jokilehto T., Wagner T., Jekmann W., Jaakkola P., Metzgen E. *Biochem. J.* 2004. V. 381. Pt. 3. P. 761–767.
162. Huang L.E., Gu J., Schau M., Bunn F. *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. № 14. P. 7987–7992.
163. Kallio P.J., Wilson W.J., O'Brien S., Makino Y., Poellinger L. *J. Biol. Chem.* 1999. V. 274. № 10. P. 6519–6525.
164. Furie B., Bouchard B.A., Furie B.C. *Blood*. 1999. V. 93. № 6. P. 1798–1808.
165. Bandyopadhyay P.K. *Vitam. Horm.* 2008. V. 78. P. 157–184.
166. Stafford D.W. *J. Thromb. Haemost.* 2005. V. 3. № 8. P. 1873–1878.
167. Wajih N., Hutson S.M., Wallin R. *J. Biol. Chem.* 2007. V. 282. № 4. P. 2626–2635.
168. Ulrich P., Cerami A. *Recent Prog. Horm. Res.* 2001. V. 56. P. 1–21.
169. Acosta J., Hettinga J., Fluckiger R., Krumrei N., Goldfine A., Angarita L., Halperin J. *Proc. Natl. Acad. Sci. USA.* 2000. V. 97. № 10. P. 5450–5455.
170. Thornalley P.J., Langborg A., Minhas H.S. *Biochem. J.* 1999. V. 344. № 1. P. 109–116.
171. Ramirez P., Del Razo L.M., Gutierrez-Ruiz M.C., Gonshebbat M.E. *Carcinogenesis*. 2000. V. 21. № 4. P. 701–706.
172. Frye E.B., Degenhardt T.P., Thorpe S.R., Baynes J.W. *J. Biol. Chem.* 1998. V. 273. № 30. P. 18714–18719.
173. Paul R.G., Avery N.C., Slatter D.A., Sims T.J., Bailey A.J. *Biochem. J.* 1998. V. 330. № 3. P. 1241–1248.
174. Niwa T., Katsuzaki T., Ishizaki Y., Hayase F., Miyazaki T., Uematsu T., Tatemichi N., Takei Y. *FEBS Lett.* 1997. V. 407. № 3. P. 297–302.
175. Hayase F., Nagaraj R.H., Miyata S., Njoroge F.G., Monnier V.M. *J. Biol. Chem.* 1989. V. 264. № 7. P. 3758–3764.
176. Wilker S.C., Chellan P., Arnold B.M., Nagaraj R.H. *Anal. Biochem.* 2001. V. 290. № 2. P. 353–358.
177. Sell D.R., Monnier V.M. *J. Clin. Invest.* 1990. V. 85. P. 380–384.
178. Obayashi H., Nakano K., Shigeta H., Yamaguchi M., Yoshimori K., Fukui M., Fujii M., Kitagawa Y., Nakamura N., Nakazawa Y., Ienaga K., Ohta M., Nishimura M., Fukui I., Kondo M. *Biochem. Biophys. Res. Commun.* 1996. V. 226. № 1. P. 37–41.
179. Tessier F., Obrenovich M., Monnier V.M. *J. Biol. Chem.* 1999. V. 274. № 30. P. 20796–20804.
180. Schmitt A., Schmitt J., Muench G., Gasic-Milencovich J. *Anal. Biochem.* 2005. V. 338. P. 201–215.
181. Tsien R. *Ann. Rev. Biochem.* 1998. V. 67. P. 509–544.
182. Zubova N.N., Bulavina A.Yu., Savitskii A.P. *Biol. Chem. Rev.* 2003. V. 43. P. 163–224.
183. Zimmer M. *Chem. Rev.* 2002. V. 103. № 3. P. 759–781.
184. Wouters F.S., Verwee P.J., Bastiaens P.I. *Trends Cell. Biol.* 2001. V. 11. № 5. P. 203–211.
185. Toomre D., Manstein D.J. *Trends Cell. Biol.* 2001. V. 11. № 7. P. 298–303.
186. Zubova N.N., Savitskii A.P. *Biol. Chem. Rev.* 2005. V. 45. P. 391–454.
187. Kain S.R. *Drug Discov. Today*. 1999. V. 4. № 7. P. 304–312.
188. Taylor D.I., Woo E.S., Giuliano K.A. *Curr. Opin. Biotechnol.* 2001. V. 12. № 1. P. 75–81.
189. Shinbrot E., Spencer C., Natale V., Kain S.R. *Meth. Enzymol.* 2000. V. 327. P. 513–522.
190. Belmont A.S. *Trends Cell. Biol.* 2001. V. 11. № 6. P. 250–257.
191. Matus A. *Trends Cell. Biol.* 1999. V. 9. № 2. P. 43.
192. Matus A. *Trends Cell. Biol.* 2001. V. 11. № 5. P. 183.
193. *Green Fluorescent Protein in Methods in Enzymology* (Conn P.M., ed.) 1999. Academic Press, New York. V. 302. P. 11–449.
194. *Green Fluorescent Protein in Methods in Cell Biology* (Sullivan K.F., Kay S.A., eds.) 1999. Academic Press, New York/ V. 58. P. 1–367.
195. Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zarsaiy A.G., Markelov M.L., Lukyanov S.A. *Nat. Biotechnol.* 1999/ V. 17. P. 969–973.
196. Lukyanov K.A., Fradkov A.F., Gurskaya N.G., Matz M.V., Labas Y.A., Savitsky A.P., Markelov M.L., Zarsaiy A.G., Zhao X., Fang Y., Tan W., Lukyanov S.A. *J. Biol. Chem.* 2000. V. 275. P. 25879–25882.
197. Verkhusha V.V., Lukyanov K.A. *Nat. Biotechnol.* 2004. V. 22. P. 289–296.
198. Yanushevich Y.G., Staroverov D.B., Savitsky A.P., Fradkov A.F., Gurskaya N.G., Bulina M.E., Lukyanov K.A., Lukyanov S.A. *FEBS Lett.* 2002. V. 511. P. 11–14.
199. Jakubowski H. *J. Nutr.* 2000. V. 130. 2S Suppl. P. 377S–381S.
200. Perla-Kajan J., Twardowski T., Jakubowski H. *Amino Acids*. 2007. V. 32. № 4. P. 561–572.
201. Jakubowski H. *J. Nutr.* 2006. V. 136. 6S Suppl. P. 1741S–1749S.
202. Jakubowski H. *Cell. Mol. Life Sci.* 2004. V. 61. № 4. P. 470–487.
203. Jakubowski H. *J. Biol. Chem.* 2002. V. 277. № 34. P. 30425–30428.
204. Jakubowski H. *J. Clin. Chem. Lab. Med.* 2005. V. 41. № 10. P. 1011–1014.
205. Gerasimova Y.V., Knorre D.G., Shakirov M.M., Godovikova T.S. *Bioorg. Med. Chem. Lett.* 2008. V. 18. № 16. P. 5396–5398.
206. Jakubowski H., Zhang L., Bardeguet A., *Aviv A. Circ. Res.* 2000. V. 87. № 1. P. 45–51.
207. Jakubowski H. *FASEB J.* 1999. V. 13. № 15. P. 2277–2283.
208. Glowacki R., Jakubowski H. *J. Biol. Chem.* 2004. V. 279. № 2. P. 10864–10871.
209. Sikora M., Marczak L., Stobiecki M., Twardowski T., Jakubowski H. *FEBS J.* 2007. V. 274. (supplement 1) P. 295.
210. Glushchenko A. V., Jacobsen D. W. *Antioxid. Redox Signal.* 2007. V. 9. P. 1883–1898.
211. Wright H.T. *Crit. Rev. Biochem. Mol. Biol.* 1991. V. 26. № 1. P. 1–52.
212. Robinson N.E., Robinson A.B. *Proc. Natl. Acad. Sci. USA.* 2001. V. 98. № 22. P. 12409–12413.
213. Wakankar A.A., Borchardt R.T. *J. Pharm. Sci.* 2006. V. 95. P. 2321–2336.
214. Powell B.S., Enama J.T., Ribot W.J., Webster W., Little S., Hoover T., Adamovicz J.J., Andrews G.P. *Proteins*. 2007. V. 68. № 2. P. 458–479.
215. Catak S., Monard G., Aviyente V., Ruiz-Lopez M.F. *J. Phys. Chem. A.* 2006. V. 110. № 27. P. 8354–8365.
216. Reissner K.J., Aswad D.W. *Cell. Mol. Life Sci.* 2003. V. 60. № 7. P. 1281–1295.
217. Aswad D.W., Paranandi M.V., Schurter B.T. *J. Pharm. Biomed. Anal.* 2000. V. 21. № 6. P. 1129–1136.
218. Takata T., Oxford J.T., Brandon T.R., Lampi K.J. *Biochemistry*. 2007. V. 46. № 30. P. 8861–8871.
219. Hasegawa M., Morishima-Kawashima M., Takio K., Suzuki M., Titani K., Ihara Y. *J. Biol. Chem.* 1992. V. 267. № 24. P. 17047–17054.
220. Doyle H.A., Gee R.J., Mamula M.J. *J. Immunol.* 2003. V. 171. № 6. P. 2840–2847.
221. Weber D.J., McFadden P.N., Caughey B. *Biochem. Biophys. Res. Commun.* 1998. V. 246. № 3. P. 606–608.
222. Sandmeier E., Hunziker P., Kunz B., Sack R., Christen P. *Biochem. Biophys. Res. Commun.* 1999. V. 261. № 3. P. 578–583.
223. Robinson A.B., McKerrow J.H., Cary P. *Proc. Natl. Acad. Sci. USA.* 1970. V. 66. № 3. P. 753–757.
224. Robinson N.E., Robinson A.B. *Proc. Natl. Acad. Sci. USA.* 2001. V. 98. № 3. P. 944–949.
225. Schmidt G., Sehr P., Wilm M., Selzer J., Mann M., Aktories K. *Nature*. 1997. V. 387. № 6634. P. 725–729.
226. Hoffmann C., Schmidt G. *Rev. Physiol. Biochem. Pharmacol.* 2004. V. 152. P. 49–63.
227. McNichol B.A., Rasmussen S.B., Carvalho H.M., Meysick K.C., O'Brien A.D. *Infect. Immun.* 2007. V. 75. № 11. P. 5095–5104.
228. Jank T., Pack U., Giesemann T., Schmidt G., Aktories K. *J. Biol. Chem.* 2006. V. 281. № 28. P. 19527–19535.
229. Stammaes J., Fleckenstein B., Solld L.M. *Biochim. Biophys. Acta.* 2008. V. 1784. № 11. P. 1804–1811.
230. Caputo L., D'Amato A., Troncone R., Auricchio S., Esposito C. *Amino Acids*. 2004. V. 26. № 4. P. 381–386.
231. Cardamone J.M. *Int. J. Biol. Macromol.* 2008. V. 42. № 5. P. 413–419.
232. Greenberg C.S., Birekbiçler P.J., Rice R.H. *FASEB J.* 1991. V. 5. № 15. P. 3071–3077.
233. Griffin M., Cassadio R., Bergamini C.M. *Biochem. J.* 2002. V. 368. Pt. 2. P. 377–396.

Genetic View on the Phenomenon of Combined Diseases in Man

V.P. Puzyrev*, M.B. Freidin

Research Institute for Medical Genetics, Siberian Branch, Russian Academy of Medical Sciences

*E-mail: valery.puzyrev@medgenetics.ru

ABSTRACT In clinical medicine, the phenomenon of polyopathy, as a particular object of investigation, was first put forth by French clinicians at the end of the 19th century through the “arthritis” doctrine. In the first half of the 20th century, German paediatricians singled out “syntropias,” which are combinations of diseases with common pathophysiological mechanisms, and “dystropias,” which are diseases that rarely co-occur in one individual. In the present paper, syntropy/dystropy is defined as a natural generic nonrandom phenomenon with an evolutionary-genetic basis. The genes involved in the development of syntropy are called “syntropic genes,” whereas the genes that co-participate in pathophysiological mechanisms and prevent the co-occurrence of particular phenotypes are called “dystropic genes.” Prospects for studying the genetic basis of this phenomenon are highlighted. The publicly available database HuGENet can be used in order to identify syntropic genes, as will be shown as examples in an analysis of cardiovascular diseases.

Keywords: syntropy, dystropy, syntropic and dystropic genes, genome, phenome, HuGENet.

INTRODUCTION

Global epidemiological studies of human diseases have yielded plenty of results, among which three observations deserve special attention when considering polyopathy and the phenomenon of polymorbidity; that is, the situation in which an individual carries several diseases at the same time. First, only 30 chronic multifactor diseases account for 65 % of all the diseases human beings suffer from, accounting for morbidity and mortality rates in contemporary populations [1]. The risk of contracting such a multifactor disease in one’s lifetime is estimated at 60 % [2] in Western populations. Secondly, polyopathy is typical of the clinical state of a contemporary patient. In patients over 65 years of age, it is common to observe more than ten related diseases in clinical practice; the co-occurrence of diseases in women is more frequent than in men (this is true for all age groups) [3]. Finally, genetic epidemiology studies of common multifactor diseases point to the importance of inherited factors in their appearance and development. The role of genetic or inherited factors in determining the common phenotype of different diseases can vary, but even with low heritability indices ($h^2 = 20\text{--}30\%$), the genetic factors that affect vulnerability to infectious agents such as viruses, bacteria, helminths and parasites can be determined [4].

COMBINED DISEASES: SYNTROPY VS. DYSTROPY

The term «polyopathy» refers to any manner of combinations of diseases and syndromes in a single patient, including accidental maladies (traumas, iatrogenic illness, etc.). However, special forms of polyopathy do exist and were combined under the term “syntropy” offered by German paediatricians M. Pfaundler and L. von Seht. They defined syntropy as the “mutual disposition, attraction” of two disease conditions, whereas they refer to “mutual repulsion” as dystropy [5]. In the same authors’ opinion, a “syndrome” can also be regarded as syntropy, because it includes selective affinity of the traits

it is made of. Another feature of the affinity of pathological conditions is synchrony, the appearance of at least two diseases simultaneously.

As proof that the syntropy concept is relevant, Pfaundler and von Seht collected and analysed about 30 thousand medical records for children starting from 1906. They offered an index of syntropy (S) showing how much the observed number of combinations of the diseases differs from the number that would be expected at random and provided many examples of syntropias with a high S index. For instance, “congenital heart disease – joint rheumatism” ($S = 58.55$), “psychopathy – enuresis” ($S = 15.31$), “hyperthyrosis – nephropathy” ($S = 4.94$), and “nephritis – exudative erythema” ($S = 4.49$).

Recently, a number of papers have been published whereby a similar idea was tested and proven using contemporary, sophisticated statistical methods [6–8]. They clearly demonstrated a significantly more or significantly less frequent co-occurrence of certain diseases than would be expected at random.

Actually, interest in polyopathy was expressed even earlier, in the 19th century, through the concept of “arthritis” by French clinician Charles Bouchard, who defined arthritis as a specific disposition toward a group of diseases occurring either in isolation or in different combinations in an individual or among many members of his or her family [9]. In the opinion of the author of the concept, these diseases had a common basic effect on metabolism: namely, they decreased it (bradytropy).

Along with syntropic interaction between diseases, antagonistic relationships, so-called “dystrophic” diseases, are described. Examples of dystropic diseases include lung tuberculosis and mitral stenosis, type I diabetes mellitus and peptic ulcer disease, lymphoproliferative and myeloproliferative processes [3], as well as lung tuberculosis and bronchial asthma [9].

There are many well-known syntropic diseases, including cardiovascular diseases [10, 11]; immune-mediated diseases (allergic diseases, autoimmune diseases) [12-14]; endocrine diseases including the combination of diabetes mellitus, autoimmune thyroiditis, and celiac disease [15, 16]; and psychiatric diseases, such as major depressive disorders and bipolar disorder [17], narcotic and addictive states [18, 19].

To assess the genetic contribution to syntropy, it is important to return to its definition: it is the natural generic non-random phenomenon of combination of two or more pathological states in an individual and his or her nearest relatives, with an evolutionary and genetic basis [20]. Syntropias comprise just part of all known polypathias. They include etiologically and pathogenetically linked combinations of diseases, unlike random combinations.

Nonrandom combinations of pathologies in an individual and his or her relatives can highlight the common genes involved in the disposition to separate diseases. Genetic studies of multifactor diseases strengthen our confidence that there are common genetic roots to such combinations (syntropias), notably when such combinations occur more often in families of patients with these groups of diseases compared to the general population.

In this aspect, we postulated that against the background of the huge number of human phenome characteristics one can carve out a fairly legible sector including a considerable number of interrelated pathological traits – syntropias – for which genetics is a worthy subject of investigation. Genes corresponding to such syntropic traits are called syntropic genes [20].

SYNTROPIC GENES (A GENERAL VIEW ON THE PROBLEM)

Epidemiological studies of complex diseases in humans provide good examples of syntropias (cardiovascular diseases continuum, allergic diseases, autoimmune diseases). Yet, for all syntropias it is important to identify the groups of genes that will determine one or the other pathophysiological pathways and can help predict the risk of syntropy among carriers of one or other combinations of those genes. Physical clustering of susceptibility genes in the human genome has been shown for a number of diseases and traits [21-23]. However, syntropic genes do not necessarily belong to a cluster of physically linked genes, but rather represent a set of functionally interacting genes dispersed throughout the human genome, co-regulated and involved in a common biochemical or physiological pathway.

Autoimmune disorders were among the first groups of diseases studied from the point of view of common genetic determinants in their development. There are many common elements in the clinical phenotype of autoimmune disorders, approaches to their therapy, population prevalence, gender ratio (75% of patients with autoimmune diseases are females), and occurrence in families. Becker K.G. *et al.* put forward a hypothesis assuming that, in some cases, clinically different autoimmune diseases can be controlled by a common set of susceptibility genes [21]. They performed a comparison of the linkage results from 23 published autoimmune or immune-mediated disease genome-wide scans in humans (multiple sclerosis, Crohn's disease, psoriasis, asthma, and type I diabetes) and animals (experimental autoimmune en-

cephalomyelitis, rat inflammatory arthritis, rat type I diabetes, murine type I diabetes, *Bordetella pertussis*-induced histamine sensitization, immunity to exogenous antigens, and murine system lupus erythematosus); non-autoimmune disorder genome scans were analyzed as well (type II diabetes, schizophrenia, bipolar disorder, leptin-dependent obesity, and hypertension). It was shown that the majority (about 65%) of human positive linkages for immune-mediated diseases map non-randomly into 18 distinct clusters and overlapping susceptibility loci occur between different human immune diseases. A similar pattern was observed in experimental autoimmune/immune disease models. A number of autoimmune candidate genetic loci from different diseases did not fall into identifiable clusters, and these singleton loci, in the author's opinion, may be independent loci; they may contribute to disease specific susceptibility, tissue or organ tropism, or may be false positives. In the control group of nonimmune-mediated human diseases, linkage with autoimmune/immune clusters was rare.

The need to screen the «immunological genome» for detection of the genetic basis of infectious, inflammatory, and autoimmune diseases was pointed out in the early 1990s [24]. A similar issue was formulated by geneticists in respect to the long-standing clinical observation of the existence of another immunological syntropy including psoriasis, psoriatic arthritis, atopic dermatitis, and asthma. However, in the pre-genomic period, it was established that psoriasis is a clear example of a Th1-mediated disease (cellular immunity) controlled by IFN- γ gene expression, unlike the Th2-mediated diseases (humoral immunity), such as asthma, for which IL-4 gene expression is important. Genomic studies of these diseases confirmed this: asthma, at least partially, results from molecular-genetic mechanisms different from those involved in psoriasis. To date, multiple genome-wide linkage studies and a number of genome-wide association scans for asthma have been performed, and several genomic regions have been repeatedly identified, including those on chromosomes 2q33, 5q23-31, 6p24-21, 11q12-13, 12q24-12, and 13q14-12. Eight new asthma susceptibility genes have been discovered by positional cloning: *ADAM33* (desintegrin and metalloproteinase-33), *DPP10* (dipeptidyl peptidase-10), *PHF11* (plant homeodomain finger protein-11), *GPRA* (G protein-coupled receptor for asthma), *HLA-G* (histocompatibility antigen), *CYFIP2* (cytoplasmic fragile X mental retardation-interacting protein 2), *IRAKM* (interleukin 1 receptor-associated kinase), and *OPN3* (opsin-3) [25, 26]. It has been shown that asthma loci and atopic dermatitis loci identified by genomic scans rarely overlap [12, 27]. At the same time, genome-wide linkage scans have identified multiple loci linked to atopic dermatitis and psoriasis and revealed shared susceptibility loci for these diseases on chromosomes 1q21, 3q21, 17q25 and 20p12 [28]. Hence, in the syntropy including four diseases assumed earlier, asthma genetically significantly differs from the other three disorders.

Certainly, the abovementioned premises are just a general conception of syntropy and its genetic basis (syntropic and dystropic genes). The genetic component of this phenomenon, noticed earlier, should be the subject of contemporary studies based on advances in molecular biology and genetics, as well as functional genomics and bioinformatics. Such concepts

help advance the complex problem of defining the genetic basis of common multifactorial human diseases.

A similar disease network hypothesis was expressed and exploited in a recent study by Rzhetsky A. *et al.* [7]. They analysed 1.5 million patient records for 161 disorders and offered an approach that allowed to estimate the extent of genetic overlaps between the diseases. Based on the results, Rzhetsky A. *et al.* concluded that “disease phenotypes form a highly connected network of strong pairwise correlations” and speculated that this can be immediately applicable to genetic mapping studies involving multiple, apparently disparate phenotypes.

SYNTROPIC GENES FOR CARDIOVASCULAR DISEASE CONTINUUM

Based on the concepts and definitions presented above, we aimed to define syntropic genes for a well-known group of syntropic diseases – cardiovascular disease continuum (CDC), including coronary artery disease (CAD), arterial hypertension (AH), stroke, metabolic syndrome (MS), dislipidemia (DL), obesity, and noninsulin-dependent diabetes-mellitus (NIDDM).

The CDC concept was initially presented in 1991 [10]. This idea considers cardiovascular diseases (CVD) as a sequential line of events determined by multiple, related and unrelated risk factors, progressing over a number of physiological and metabolic pathways and resulting in the development of the final stage of a heart disease. The continuum members (diseases and traits) overlap and interact as a sequence of discrete and tandem states during the progression of CVD [10, 11]. This context allowed us to refer this unity of pathological conditions to a syntropy.

We used the publicly available research tool HuGE Navigator to identify the genes underlying the development of seven CDC diseases (as specified above). The HuGE Navigator provides access to a continuously updated human genome epidemiology database that includes information on population prevalence of genetic variants, gene-disease associations, gene-gene and gene-environment interactions, and the evaluation of genetic tests. It is based on the achievements of the Human Genome Epidemiology Network (HuGENet™), a voluntary, international collaboration focused on assessing the role of human genome variation in health and disease at the population level. Since 2001, HuGENet™ has maintained a database of published, population-based epidemiologic studies of human genes extracted and curated from PubMed [29].

In the HuGENet database, the number of genes studied in relation to the seven CVDs varied between 162 for DL and 466 for AH. HuGE Navigator ranks these genes by score for a given gene estimated as the ratio of the number of studies showing positive results for an association to the total number of published studies.

To stress the strength of the associations between genes and diseases, we only considered genes with a score equal to or greater than 0.01. The maximum scores for different genes associated with the diseases analyzed were 4.1 for DL, 1.60 for CAD, 1.12 for AH, 1.02 for stroke, 1.01 for MS, 0.74 for obesity, and 0.36 for NIDDM.

Twenty-one of the genes found were associated with all of the CDC diseases (Table 1). Certainly, the number of genes

underlying any particular disease included in CDC syntropy is much higher. However, for total CDC syntropy, in accordance with HuGENet data and ranking criteria, only these 21 genes can be attributed to the control of the development and structure of CDC syntropy itself, and these genes can be called syntropic genes of CDC. Two features merit special attention. First, most of these genes are well-characterized and have been studied at length. Secondly, they comprise the inherited basis of the pathophysiological continuum of mechanisms underlying the development of this syntropic disease group, including dysfunction in lipid metabolism, renin-angiotensin-aldosterone system, sympathoadrenal system, inflammation, and endothelial function.

The results of genetic association studies (GAS) are controversial due to sample heterogeneity (ethnic variety, age and gender differences), relatively small sample sizes (mega-studies have appeared only recently), and indistinct clinical criteria in disease group recruiting. Taking this into account, it is generally assumed that GASs should be accompanied by meta-analysis, as well as genome-wide association studies (GWASs). These principles were taken into account in the development of HuGENet, and genes appearing in GWAS and meta-analysis receive higher coefficients.

Table 1 – Syntropic genes for cardiovascular disease continuum

Order number	Gene symbol	Gene product	Chromosomal localisation
1	ABCA1	ATP-binding cassette transporter 1	9q22-q21
2	ACE	Angiotensin I-converting enzyme	17q23
3	ADIPOQ	Adipocyte-specific secretory protein	3q27
4	ADRB2	β ₂ -adrenergic receptor	5q32-q34
5	AGT	Angiotensinogen	1q42-q43
6	AGTR1	Angiotensin receptor 1	3q21-q25
7	APOA1	Apolipoprotein A1	11q23
8	APOE	Apolipoprotein E	19q13.2
9	CETP	Cholesteryl ester transfer protein	16q21
10	ESR1	Estrogen receptor 1	6q25.1
11	GNB3	Beta-3 G-binding protein	12p13
12	IL6	Interleukin-6	7p21
13	LIPC	Hepatic lipase C	15q21-q23
14	LPL	Lipoprotein lipase	8p22
15	LTA	Lymphotoxin-α	6p21.3
16	MTHFR	Methylenetetrahydrofolate reductase	1p36.3
17	NOS3	Endothelial NO-synthase	7q36
18	PPARG	Peroxisome proliferator-activated receptor-γ	3p25
19	SERPINE1	Plasminogen activator inhibitor 1	7q21.3-q22
20	SELE	Selectin E	1q23-q25
21	TNF	Tumor necrosis factor-α	6p21.3

Ninety-one meta-analysis studies carried out for 21 genes and seven diseases were found in HuGENet. The numbers vary for different diseases, and the maximum values found are for the *MTHFR* and *APOE* genes (18 and 13 studies, respectively). No meta-analyses have been published so far for the *SELE*, *ESR1*, and *SERPINE1* genes. Among the diseases considered, CAD, stroke, AH, and obesity were the most common subjects of meta-analysis: 28, 21, 17, and 13 studies, respectively.

At the time of this study, in the HuGENet, there were 13 meta-analyses for the *APOE* gene, and CAD and stroke. Recent analyses [30] included 203 studies for a period from 1970 to 2007, providing ultimate proof of the significant association between genetic variants of the *APOE* gene and CAD and stroke. For carriers of the E4 allele, the risk of CAD development is 20% higher than those without it, and the cholesterol levels in low-density lipoproteins increase in the following direction (in terms of the presence of the *APOE* allele): E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, E4/E4. The odds ratio (OR) for CAD development in carriers of E4 is 1.06 (95% CI 0.99-1.13). In a review of 500 papers [31], a significant and direct association between the *APOE* and strokes in Asians (Chinese, Japanese, Koreans)---but not in Caucasians---was established. Also, the association of strokes with other genes (*ACE* I/D, *MTHFR* 677C/T) was confirmed. Thus, three phenotypes of the CDC syntropy express association with *APOE* and the association is confirmed by meta-analyses.

For the other gene subjected to numerous meta-analyses, *MTHFR*, the association with five CDC phenotypes (CAD, AH, stroke, NIDDM, obesity) and its C677T polymorphism is rather supported [32-36] than not [37]. It is notable that the *MTHFR* polymorphism is an independent risk factor for AH [38].

The association of the *ACE* gene with CVD has been studied for a long period of time, and a large body of data has been accumulated. However, meta-analyses for the gene and their disease groups have been initiated only recently. There are seven such publications in the HuGENet database. In the meta-analysis of 118 studies, the I/D polymorphism of *ACE* was conclusively shown to be associated with CAD and NIDDM [39]; however, there are meta-analyses that cannot confirm this association [36, 40]. Meta-analysis of the other gene in the rennin-angiotensinogen system, *AGT*, showed significant association between its polymorphism M235T and CAD, AH, obesity, and stroke; in all cases, the 235T/T genotype is associated with increased risk of diseases [41, 42].

For the association of the *LPL* gene and CVDs, the results of meta-analyses are quite controversial. There are just seven such analyses, and they deal with separate multiple polymorphisms of the gene. The results of one meta-analysis study identified an association between the Asn291Ser mutation and CAD, NIDDM, and DL (hypertriglyceridemia and low level of cholesterol in high-density lipoproteins) [43]. A meta-analysis in which, for the first time, a gene-gene interaction between *APOE* and *LPL* was shown has been published [44]. According to this study, in co-carriers of *APOE** E4 and *LPL**447X (S447X polymorphism), OR for stroke and myocardial infarction development was 2.2 (p = 0.01).

Three genes critical for the development of inflammation were shown to be important for CDC syntropy: *IL6*, *TNF*, and

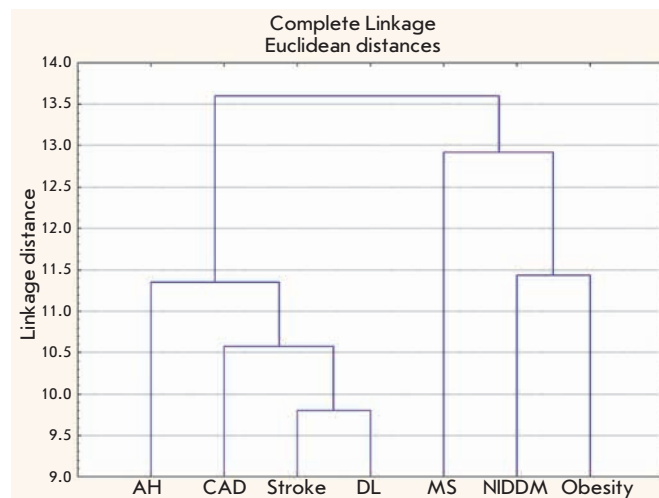


Fig. 1. Tree diagram showing clusters of cardiovascular disease continuum members based on the number of shared/non-shared genes associated with them. MS – metabolic syndrome; NIDDM – non-insulin dependent diabetes-mellitus; AH – arterial hypertension; CAD – coronary artery disease; DL – dislipidemia

LTA. For the *LTA* gene, only one meta-analysis is known (for CAD); however, for the *IL6* and *TNF* genes, five and seven meta-studies are cited in HuGENet, respectively. *IL6* was examined in meta-analyses for association with NIDDM and CAD and confirmed the lack of any association [45-47]. For the *TNF* gene, there are meta-analyses for all diseases involved in the CDC syntropy, except for DL. The analysis of 31 studies of MS and -308G/A polymorphism of *TNF* showed that -308A allele carriers show a 23% higher risk of obesity; they also have significantly increased systolic blood pressure and plasma insulin levels [48]. The same polymorphism shows stable association with CAD and stroke in Asians: genotype -308G/G carriers show a 40% lower risk of stroke than the others [49].

For the other CDC syntropic genes and phenotypes, there are very few meta-analyses, and it is important to study them further using this method.

Based on the information on shared and non-shared genes, we conducted a hierarchical cluster analysis of the CDC syntropy diseases to determine whether they have gene-based relationships. Two tight clusters are seen; one is comprised of AH, CAD, stroke and DL, while the other is composed of MS, obesity, and NIDDM (Fig. 1).

It seems reasonable that AH, CAD, stroke, and DL form a close cluster by shared susceptibility genes. This has been proven by Dzau V. *et al.* [11] in their paper presenting the CDC concept. For these four components of CDC, the overlap of basic pathways (lipid metabolism, rennin-angiotensin-aldosterone and adrenergic systems, oxidative stress, and endothelial dysfunction) in the phenotype development is demonstrated [34, 41].

However, MS, obesity, and NIDDM have their common and specific features in pathogenesis. The common feature being that their leading pathway is insulin metabolism. Nevertheless, MS is distinguished by resistance to hyperinsulin-

ism and insulin, while NIDDM is characterized by the impaired pancreatic beta-cell function and insulin resistance [50, 51]. These features can be attributed to the relations between MS, NIDDM, obesity, and the other CVDs studied, which are reflected in the cluster diagram.

The approach presented here can potentially be applied to any other syntropic disease group. Presumably, it could be interesting to analyse the genetic clustering of the plethora of human diseases for the purpose of building a generic system for their classification.

A similar approach was used recently by Torkamani A. *et al.* [8], who showed a high correlation between the SNPs significant in the GWAS for CAD, AH, and NIDDM, as well as for bipolar disorder (BD) and a number of immune-mediated diseases. In particular, they showed that among the top-ranked (in terms of the statistical significance of their association with diseases) 1 000 SNPs, there are 57 shared by CAD and NIDDM, 81 shared by AH and NIDDM, and 63 shared by AH and CAD. These genetic correlations between the diseases were highly significant. Also, strong correlation between the autoimmunity-related disorders rheumatoid arthritis and insulin-dependent diabetes-mellitus was demonstrated. Surprisingly, a strong genetic correlation of BD and CAD and NIDDM, as well as a strong correlation between AH and Crohn's disease – seemingly unrelated diseases – was observed. This suggests some unexpected links between these diseases and argues for the utility of a genetic correlation-based approach for natural disease categorization.

CONCLUSION

“Phenome” by analogy with the term “genome” is defined as the exact phenotypic representation of a species [52]. It includes the morphological, biochemical, physiological, and ontogenetic characteristics of an organism. Phenomics seeks to define the extent of variability in the phenome but represents a major challenge. The view of a pathological phenotype from the point of view of nonrandom combination of morbid traits (syntropy) does not coincide with the clinical tradition of paying primary attention to a particular diagnosis, a nosology. In the syntropic approach, from the infinite number of traits of the phenome, a sampling of interrelated traits controlled by common genes is assumed. The way to identify such syntropic genes is not significantly different from what is done for the genetic analysis of any non-Mendelian trait. However, substantially larger population sample sizes will be required to achieve confidence in a gene-phenome link. Moreover, unification (standardisation) of a phenotype is critical, if very time-consuming, and dependent on the clinicians and epidemiologists involved in ongoing epidemiological studies in different regions of the world.

In this respect, it is also worth noting the very well-recognized phenomenon of pleiotropy, the multiple phenotypic effects of a single gene. A few recent studies discuss this problem in application to human diseases and put forth ideas similar to the concept of syntropy [53, 54]. Likely, pleiotropy is one of the basic factors of syntropy development. The sum of the pleiotropic effects of genes constitutes the physiological fields of their action, which can be described as the gene's network or, in a more common sense, as the biological network. The overlap of the fields of action of the genes forms a meta-field, which is the basis for the development of a group of diseases bearing relation to these genes. Given the common genetic background based on the interaction between a limited set of genes, these diseases would have a tendency to cluster together, constituting a syntropy. At the same time, these diseases are phenotypically distinguishable, because different parts of the gene meta-field of action will have a conclusive significance for different diseases.

An increase in genetic association studies, both candidate-gene-based and GWAS-based, is forecasted [39]. This prognosis is based on the advances in the availability of mapped SNPs, finalisation of the HapMap Project, microarray-based genotyping technology development, and evolution of statistical and bioinformatical methods. However, looking for genetic markers for complex disease risk is not as straightforward as detecting phenotypic biomarkers of the disease risk currently used in clinical practice [55].

Actually, the OR attributable to most alleles associated with complex diseases both in GASs and GWASs rarely reaches a value of 1.15–1.50; usually, even weaker associations are detected, and their application to clinical practice is estimated to be low [55]. However, it cannot be excluded that the weak effect is a consequence of the genetic and phenotypic heterogeneity of the studied population, when there are individuals showing higher effects of respective genes, too. Taking this into consideration, it is suggested that a well-focused approach to the organization of the study will help detect stronger and more robust associations of alleles with pathological phenotypes, for instance, in young patients, among persons with a definite family aggregation of the diseases, or in patients frequently hospitalized [56]. In our opinion, the analysis of common genes for a chosen syntropy offered an option for such a well-focused study, which is helpful in discovering genes with strong effects and ranking them by their effect in relation to a pathophysiological continuum. ●

*This work was supported in part
by the Russian Foundation for Basic Research,
(grants № 07-04-01613, 07-04-01526, 08-04-01814).*

REFERENCES

1. Czeizel A. The baseline data of the Hungarian Congenital Malformation Register, 1970–1976. *Acta Paediatr Acad Sci Hung*, 1978, V. 19, P. 149–156.
2. Baird P. A., Anderson T. W., Newcombe H. B., Lowry R. B. Genetic studies in children and young adults: population study. *Am J Hum Genet*, 1988, V. 42, P. 677–693.
3. Krylov A. A. On the problem of disease association. *Klin Med*, 2000, V. 78, P. 56–58.
4. Hill A. V. S. Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet*, 2006, V. 40, P. 469–486.
5. Pfaundler M., von Seht L. Weiteres über Syntropie kindlicher Krankheitszustände. *Z Kinderheilkd*, 1921, V. 30, P. 298–313.
6. Williams F. M. K., Cherkas L. F., Spector T. D., MacGregor A. J. A common genetic factor underlines hypertension and other cardiovascular disorders. *BMC Cardiovascular Disorders*, 2004, V. 4, P. 20.
7. Rzhetsky A., Wajngurt D., Park N., Zheng T. Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci USA*, 2008, V. 104, P. 11694–11699.
8. Torkamani A., Topol E. J., Schork N. J. Pathway analysis of seven common diseases as-

REVIEWS

- essed by genome-wide association. *Genomics*, 2008, V. 92, P. 265-272.
9. Bouchard Ch. Lecons sur les maladies par ralentissement de la nutrition. 1890, Librairie F. Savy, Paris.
 10. Dzau V. J., Braundwald E. Resolved and unresolved issues in the prevention and treatment of coronary artery disease: a workshop consensus statement. *Am Heart J*, 1991, V. 121, P. 1244-1263.
 11. Dzau V. J., Antman E. M., Black H. R. *et al.* The cardiovascular disease continuum validated: clinical evidence of improved patient outcomes: part I: Pathophysiology and clinical trial evidence (risk factors through stable coronary artery disease). *Circulation*, 2006, V. 114, P. 2850-2870.
 12. Cookson W. The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol*, 2004, V. 4, P. 978-988.
 13. Gregersen P. K., Behrens T. W. Genetics of autoimmune diseases-disorders of immune homeostasis. *Nat Rev Genet*, 2006, V. 7, P. 917-928.
 14. Wellcome Trust Case Control Consortium & The Australo-Anglo-American Spondylitis Consortium. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet*, 2007, V. 39, P. 1329-1337.
 15. Hanukoglu A., Mizrahi A., Dalal I. *et al.* Extraprostatic autoimmune manifestations in type 1 diabetes patients and their first-degree relatives. *Diabetes Care*, 2003, V. 26, P. 1235-1240.
 16. Doolan A., Donaghue K., Fairchild J., Wong M., Williams A. J. Use of HLA typing in diagnosing celiac disease in patients with type 1 diabetes. *Diabetes Care*, 2005, V. 28, P. 806-809.
 17. Harvey M., Belleau P., Barden N. Gene interactions in depression: pathways out of darkness. *Trends Genet*, 2007, V. 23, P. 547-556.
 18. Decker K. P., Ries R. K. Differential diagnosis and psychopharmacology of dual disorders. *Psychiatr Clin North Am*, 1993, V. 16, P. 703-718.
 19. Li M. D., Burmeister M. New insights into the genetics of addiction. *Nat Rev Genet*, 2009, V. 10, P. 225-231.
 20. Puzyrev V. P. Genomic investigations in genetic cardiology: status and prospects. *Vestn Ross Akad Med Nauk*, 2000, № 7, P. 28-33.
 21. Becker K. G., Simon R. M., Bailey-Wilson J. E. *et al.* Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci USA*, 1998, V. 95, P. 9979-9984.
 22. Niehrs C., Pollet N. Synexpression groups in eukaryotes. *Nature*, 1999, V. 402, P. 483-487.
 23. Brunner H. G., van Driel M. A. From syndrome families to functional genomics. *Nat Rev Genet*, 2004, V. 5, P. 545-551.
 24. Epplen J. T. On genetic components in autoimmunity: a critical review based on evolutionarily oriented rationality. *Hum Genet*, 1992, V. 90, P. 331-341.
 25. Moffatt M. F. Genes in asthma: new genes and new ways. *Curr Opin Allergy Clin Immunol*, 2008, V. 8, P. 411-417.
 26. Postma D. S., Koppelman G. H. Genetics of asthma: where are we and where do we go? *Proc Am Thorac Soc*, 2009, V. 6, P. 283-287.
 27. Bowcock A. M. The genetics of psoriasis and autoimmunity. *Annu Rev Genomics Hum Genet*, 2005, V. 6, P. 93-122.
 28. Bowcock A. M., Cookson W. O. C. M. The genetics of psoriasis, psoriatic arthritis, and atopic dermatitis. *Hum Mol Genet*, 2004, V. 13, Spec. № 1, R43-55.
 29. Yu W., Gwinn M., Clyne M., Yesupriya A., Khoury M. J. A Navigator for Human Genome epidemiology. *Nat Genet*, 2008, V. 40, P. 124-125.
 30. Bennet A. M., Di Angelantonio E., Ye Z. *et al.* Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA*, 2007, V. 298, P. 1300-1311.
 31. Ariyaratnam R., Casas J. P., Whittaker J. *et al.* Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. *PLoS Med*, 2007, V. 4, e131.
 32. Wu A. H., Tsongalis G. J. Correlation of polymorphisms to coagulation and biochemical risk factors for cardiovascular diseases. *Am J Cardiol*, 2001, V. 87, P. 1361-1366.
 33. Klerk M., Verhoef P., Clarke R. *et al.* MTHFR Studies Collaboration Group; MTHFR 677C ->T polymorphisms and risk of coronary heart disease: a meta-analysis. *JAMA*, 2002, V. 288, P. 2023-2031.
 34. Casas J. P., Hingorani A. D., Bautista L. E., Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol*, 2004, V. 61, P. 1652-1661.
 35. Zintzaras E., Chatzoulis D. Z., Karabatsas C. H., Stefanidis I. The relationship between C677T methylenetetrahydrofolate reductase gene polymorphism and retinopathy in type 2 diabetes: a meta-analysis. *J Hum Genet*, 2005, V. 50, P. 267-275.
 36. Banerjee I., Gupta V., Ganesh S. Association of gene polymorphism with genetic susceptibility to stroke in Asian population: a meta-analysis. *J Hum Genet*, 2007, V. 52, P.205-219.
 37. Lewis S. J., Ebrahim S., Smith D. G. Meta-analysis of MTHFR 677C->T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? *BMJ*, 2005, V. 331, P. 1053.
 38. Qian X., Lu Z., Tan M., Liu H., Lu D. A meta-analysis of association between C677T polymorphism the methylenetetrahydrofolate reductase gene and hypertension. *Eur J Hum Genet*, 2007, V. 15, P. 1239-1245.
 39. Zintzaras E., Lau J. Trends in meta-analysis of genetic association studies. *J Hum Genet*, 2008, V. 53, P. 1-9.
 40. Kitsios G., Zintzaras E. Genetic variation associated with ischemic heart failure: a HuGE review and meta-analysis. *Am J Epidemiol*, 2007, V. 166, P. 619-633.
 41. Navarro-Lopez F. Genes and coronary heart disease. *Rev Esp Cardiol*, 2002, V. 55, P. 413-431.
 42. Mondry A., Loh M., Liu P., Nagel M. Polymorphisms of the insertion/deletion ACE and M235T AGT genes and hypertension: surprising new findings and meta-analysis of data. *BMC Nephrol*, 2005, V. 6, P. 1.
 43. Hu Y., Lin W., Huang R., Zhang X. A systematic review and meta-analysis of the relationship between lipoprotein lipase Asn 291 Ser variant and diseases. *J Lipid Res*, 2006, V. 47, P. 1908-1914.
 44. Baum L., Ng H. K., Wong K. S. *et al.* Association of apolipoprotein E exon 4 and lipoprotein lipase S447X polymorphisms with acute ischemic stroke and myocardial infarction. *Clin Chem Lab Med*, 2006, V. 44, P. 274-281.
 45. Huth C., Heid I. M., Vollmert C. *et al.* IL6 gene promoter polymorphisms and type 2 diabetes: joint analysis of individual participants' data from 21 studies. *Diabetes*, 2006, V. 55, P. 2915-2921.
 46. Qi L., van Dam R. M., Meigs J. B. *et al.* Genetic variation in IL6 gene and type 2 diabetes: tagging-SNP haplotype analysis in large-scale case-control study and meta-analysis. *Hum Mol Genet*, 2006, V. 15, P. 1914-1920.
 47. Sie M. P., Sayed-Tabatabaei F. A., Oei H. H. Interleukin 6 -174 g/c promoter polymorphism and risk of coronary heart disease: results from the Rotterdam study and a meta-analysis. *Arterioscler Tromb Vasc Biol*, 2006, V. 26, P. 212-217.
 48. Sookoian S. C., Gonzalez C., Pirola C. J. Meta-analysis on the G-308A tumor necrosis factor alpha gene variant and phenotypes associated with the metabolic syndrome. *Obes Res*, 2005, V. 13, P. 2122-2131.
 49. Pereira T. V., Rudnicki M., Franco R. F., Pereira A. C., Krieger J. E. Effect of the G-308A polymorphism of tumor necrosis factor alpha gene on the risk of ischemic heart disease and ischemic stroke: a meta-analysis. *Am Heart J*, 2007, V. 153, P. 821-830.
 50. Reaven G. M. The insulin resistance syndrome. *Curr Atheroscler Rep*, 2003, V. 5, P. 364-371.
 51. Eckel R. H., Grundy S. M., Zimmet P. Z. The metabolic syndrome. *Lancet*, 2005, V. 365, P. 1415-1428.
 52. Freimer N., Sabatti C. The Human genome project. *Nat Genet*, 2003, V. 34, P. 15-21.
 53. Goh K.-I., Cusick M. E., Valle D. *et al.* The human disease network. *Proc Natl Acad Sci USA*, 2007, V. 104, P. 8685-8690.
 54. Tyler A. L., Asselbergs F. W., Williams S., Moore J. H. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *BioEssays*, 2009, V. 31, P. 220-227.
 55. Cambien F., Tiret L. Genetic of cardiovascular diseases. From single mutation to whole genome. *Circulation*, 2007, V. 116, P. 1714-1724.
 56. Cambien F. About mega-studies, genetics and cardiovascular diseases. In: Prospects and limitations of very large cohort studies. Available via GeneCanvas. http://genecanvas.idf.inserm.fr/readarticle.php?article_id=265. 2006.

Achievements and Peculiarities in Studies of Ancient DNA and DNA from Complicated Forensic Specimens

A.P. Grigorenko^{1,2,3}, S.A. Borinskaya¹, N.K. Yankovsky¹, E.I. Rogaev^{1,2,3*}

¹ Vavliov Institute of General Genetics, Russian Academy of Sciences

² Research Center of Mental Health, Russian Academy of Medical Sciences

³ University of Massachusetts Medical School, Worcester, U.S.A.

* E-mail: rogaev@vigg.ru

ABSTRACT Studies of ancient DNA specimens started 25 years ago. At that time short mitochondrial DNA (mtDNA) fragments were the main targets in ancient DNA studies. The last three years were especially productive in the development of new methods of DNA purification and analysis. Complete mtDNA molecules and relatively large fragments of nuclear DNA are the targets of ancient DNA studies today. Ancient DNA studies allowed us to study organisms that went extinct more than ten thousand years ago, to reconstruct their phenotypic traits and evolution. Ancient DNA analyses can help understand the development of ancient human populations and how they migrated. A new evolutionary hypothesis and reconstruction of the biota history have been re-created from recent ancient DNA data. Some peculiarities and problems specific to the study of ancient DNA were revealed, such as very limited amounts of DNA available for study, the short length of the DNA fragments, breaks and chemical modifications in DNA molecules that result in “postmortem” mutations or complete blockage of DNA replication *in vitro*. The same specific features of DNA analysis were revealed for specimens from complicated forensic cases that result in the lack of experimental data or interpretation problems.

Here, we list the specific features of ancient DNA methodology and describe some achievements in fundamental and applied research of ancient DNA, including our own work in the field.

Keywords: ancient DNA, methods, evolution, DNA identification, forensic examination.

Abbreviations: mtDNA – mitochondrial DNA, cmtDNA – complete mtDNA sequence, STR – short tandem repeats.

INTRODUCTION

Ancient DNA studies allow to empirically verify evolutionary hypotheses and contribute to the complex reconstruction of historical changes in biota. The analysis of DNA from human archeological samples reveals information on the genetic traits of ancient inhabitants of various geographical regions.

The first published reports on the study of ancient DNA appeared 25 years ago. Researchers managed to extract a DNA fragment from a museum sample of dried muscle tissue taken from a quagga – a South-African odd-toed ungulate animal that disappeared in the 19th century. The extracted DNA fragment was cloned in a phage vector and sequenced. Phylogenetic analysis showed that the determined sequence of the mitochondrial DNA (mtDNA) was related to zebra species [1, 2]. The next study described the extraction, cloning, and sequencing of DNA fragment from a ~2,400-thousand-year-old Egyptian mummy [3]. After these, attempts were made to extract DNA from the remains of animals, plants and microorganisms whose ages ranged from several hundreds to more than a million years (see review in [4]). As the data accumulated, it became clear that the age of the remains that could still have analyzable templates, calculated using kinetics of DNA decay, was not greater than 0.1–1.0 Myr, and that the level of DNA preservation depended on the age and type of the biological sample, and also on the conditions in which it was preserved [5, 6]. Reports on the extraction of DNA from specimens older than 1 Myr are most probably erroneous. The most ancient authentic DNA samples have been isolated from

permafrost specimens, such as mammoth, bison and other animal remains, chloroplast DNA from plants, and bacterial DNA [7–12]. Fragments up to 900–1 000 base pairs (bp) have been amplified from these samples. The low temperature and humidity improve DNA preservation, that allows researchers to analyze samples from remains that are tens of thousands of years old (Fig. 1) [13].

Studies of DNA isolated from ancient or historic specimens must deal with a number of methodical problems. These include the exceedingly small quantity and small size of the DNA fragments extracted from ancient samples, as well as the presence of chemical DNA modifications that block DNA replication and cause the appearance of postmortem mutations in the nucleotide sequence. Spontaneous damage of DNA molecules in a living cell is repaired during replication or causes death and elimination. After an organism dies, both the reparation and elimination of cells with damaged DNA come to a halt; that leads to the accumulation of chemical modifications and fragmentation of DNA molecules. Furthermore, the DNA is further destroyed by the organisms of the soil biota. Degradation of ancient DNA makes contamination of the samples by even a single molecule of modern DNA a major cause of false results.

Using the PCR techniques [14–16] has considerably widened the possibilities of ancient DNA analysis, since it allows *in vitro* amplification of a single original DNA molecule. PCR allows selective amplification of target DNA fragments; that is very important in ancient DNA analysis, since 99% of the

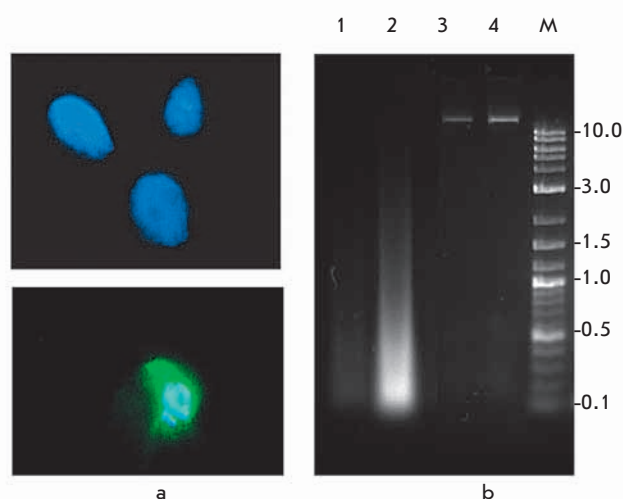


Fig. 1. Unusually well-preserved mammoth *M. primigenius* DNA specimen obtained from permafrost-preserved remains found in 1986 in the valley of the Enmyneem River (Chukotka): a – Fluorescence of nuclei in muscle cells of *M. primigenius* (~33,000 years old) after DAPI staining indicates a relatively high degree of DNA preservation; b – electrophoresis of total genomic DNA extracted from mammoth muscle tissue (lane 1 is 1/10 dilution of the DNA on lane 2), and control DNA isolated from fresh human blood samples (lanes 3 and 4). The right lane is a DNA molecular weight marker (fragments sizes are indicated in kb) [13]

extracted DNA can be a mixture of bacterial or fungal DNA from soil.

The development of extraction and sequencing technologies in ancient DNA study in the past few years has allowed to sequence the complete mitochondrial genomes, to reconstruct nuclear genomic regions, to analyze the genetic population variety of extinct species (moa, mammoth, furry rhinoceros, cave bear, Beringian bison, giant eagle, Neanderthal, etc.), and to study changes in the Pleistocene and Holocene ecosystems (see reviews [4, 17–20]).

Years of research have resulted in the establishment of a number of requirements to ancient DNA studies and criteria upon which the authenticity of resulting data is judged. Contamination of the analyzed samples by modern DNA is still one of the most pressing problems during ancient DNA analysis. One of the well-known examples is the report related to the extraction of a DNA fragment from the dinosaur bone [21] that, as it turned out during further analysis, was a fragment of human nuclear DNA [22], as well as the mentioned above attempt to sequence DNA from an Egyptian mummy [3]. Currently, the sequence obtained in the mummy study is assumed to be the result of modern human DNA contamination [17, 23].

DNA sequences reconstructed from ancient or historic samples can also contain errors because of hydrolytic or oxidative modification of the ancient DNA. For example, sequencing of long regions of the same sample of Neanderthal nuclear DNA was performed by two groups of researchers. The Edward Rubin group from the Joint Genome Institute

of the U.S. Department of Energy published a 65,000 bp sequence, and the Svante Pääbo group from the Max Planck Institute for Evolutionary Anthropology reported the sequencing of 1 million bp [24, 25]. Further data analysis determined that a considerable amount of errors was present in the second group's data. Most of the data was the result of contamination by modern DNA. Furthermore, the “single-run” sequencing approach used by this group did not prevent multiple errors that appeared because of ancient DNA nucleotide modification and could only be excluded by multiple sequencing [26–28]. Errors were also detected in the published sequence of the mtDNA fragment from the Feldhofer cave Neanderthal [29]. Out of the 27 detected alterations, as compared to human mtDNA, 4 turned out to be artefacts [30]. The published nucleotide sequences of other species are also not free of errors (the Pleistocene cave bear [31], mammoth [32, 33], etc.).

Similar problems (exceedingly small amounts of DNA, or DNA damaged by thermal conditions or chemical agents) often come up during the genetic analysis of forensic samples. An analysis of these problems and approaches that allow to solve them are presented in this review.

EXTRACTION OF DNA AND CONTAMINATION PROBLEM

Paleontological and archeological materials and biological samples that are collected at excavation sites or stored in museums yield very small amounts of DNA that is usually highly fragmented. Moreover, this ancient DNA is modified in various ways that prevents amplification or lead to errors in nucleotide sequence reads. Because of the low efficiency of amplification of authentic DNA extracted from ancient and historic samples, contamination of the sample by even a single modern DNA molecule can produce errors. A number of specific measures must be taken in order to prevent contamination and to detect possible contamination. False positive results, caused by in-lab contamination, are one of the major problems in ancient DNA studies. That is why the key step in molecular-genetic analysis of ancient and historic samples is DNA extraction.

Extraction of DNA from ancient samples must be performed in accordance with sample age and quality. It particularly involves the choice of the detergent used for cell lysis. Sodium dodecyl sulfate (SDS) that is used in standard DNA extraction for the purpose of lipid destruction could be substituted by non-ionic detergents for soft lysis (Triton or Twin), or by detergent-free extraction, since lipids have been already destroyed in ancient samples, and the use of SDS lowers the DNA yield. Nevertheless, the use of detergents is recommended for more recent samples. Treating bone material with reagents that include EDTA causes sample decalcification and pH lowering, which can affect DNA binding on the columns used in downstream extraction procedures.

Studies of ancient DNA should be conducted in specially equipped facilities with all possible means for preventing contamination by modern DNA. It includes facilities with altered air pressure: high pressure in the rooms used for ancient DNA work and low pressure in rooms where modern DNA and amplified products are studied. These facilities must be regularly disinfected with chemicals and UV-radiation, to avoid any DNA (target, amplified or contaminant) and cell

REVIEWS

Table 1. Authenticity criteria for ancient DNA

Criteria	Importance for authenticity
Reagents and plastic to be used for work with ancient DNA must be tested for potential contamination (possible presence of amplifiable foreign DNA templates in any reagent). Contamination of the laboratory material by only single DNA molecules can occur, thus the contaminating DNA molecules may yield amplification PCR product in only one out of several tested samples. Thus multiple control tests have to be performed for the laboratory reagents and material.	Prevents contamination through reagents and disposable materials.
All the manipulations for DNA extraction are performed for the control samples using in parallel the same solutions, but with no ancient material added in a course of the procedures. Thus PCR is performed with a double negative control: the regular "PCR control" (PCR mix with no extracted DNA template) and the extraction control (PCR mix with the "empty" extract).	Helps to detect contamination that may happen during extraction or in a process of the PCR mix preparation.
Positive controls are usually not used to avoid the risk of potential cross-contamination of the tested samples.	Prevents contamination.
Whenever possible, several independent extractions of DNA are performed using different samples from the same specimen.	Helps to identify the authentic endogenous DNA if local contaminations by external DNA occurred on the tested specimen itself.
Repeated PCR amplifications of material obtained from the same and from different extractions.	Helps to identify sporadic contamination; facilitates the identification of potential errors due to PCR errors or chemical modifications of ancient DNA templates ("postmortem ancient DNA mutations"); control the phenomenon of drop-out or drop-in alleles (heterozygous and homozygous state) occurring in PCR of very low amount of DNA (low copy number DNA templates)
Cloning of amplification products and sequencing of multiple clones.	Identifies heterogeneity in the amplified products caused by contamination with foreign DNA or incorporation of postmortem DNA mutations due to chemical modification of ancient DNA templates nucleotides.
Quantification of the number of original DNA template molecules taken for each PCR. The minimal amount of DNA templates required can vary noticeably depending on the length and the nucleotide content of the amplified fragment, and also on the sensitivity of the specific pair of primers.	Determines the possibility that nucleotide errors occasionally generated during the initial rounds of PCR will exist in the majority of the molecules in a final PCR product and thereby, the consensus sequence may harbor the errors. Extracts that contain only a few or even a single molecule are very much prone to yield this type of errors. To avoid them several independent amplifications are required to determine the authentic sequence. Amplification using at least 1 000 original DNA molecules in PCR most likely yield the correct consensus sequence in each nucleotide position.
Peculiar "molecular behavior," a reverse correlation between the efficiency of amplification and the length of the amplified fragment.	If the sample does not exhibit more intensive amplification of shorter fragments than that of longer fragments, as compared to modern DNA, this indicates that the source of the amplified DNA is contaminated by modern templates.
Biochemical analysis of the level of preservation of macromolecules.	A high level of biochemical preservation of macromolecules indicates a high probability of preserved DNA molecules being found in the sample. Thus, the test would support the authenticity of the DNA sequencing results.
Careful design of PCR primers for authentic mtDNA sequences to avoid concurrent PCR amplification of nuclear pseudo mitochondrial DNA (numt) sequences along with mtDNA regions.	Nuclear genome has regions of multiple insertions of mtDNA regions diverged during the evolution. Thus the amplification of nuclear pseudogenes instead or along with mitochondrial DNA sequences could be potentially obtained for certain PCR primers designed for amplification of mtDNA genome sequences.
Independent confirmation of results in a different laboratory.	It may help to identify laboratory contamination of samples or reagents, but it does not rule out contamination that can occur in the sample itself (contaminants that were a part of the sample before it arrived at the laboratory, for instance during archeological excavation). The requirement for two independent laboratories involved in DNA extraction and primary analysis was mandatory. Now it has been dropped.

material (aerosols and dust with microorganisms and human and other organism's cells). Ancient DNA work must be conducted in protective clothing, gloves, and masks. A minimum requirement is that DNA extraction procedures performed on ancient DNA and involving single fragmented molecules must be performed in facilities physically separated from the ones for PCR-amplification and downstream amplified DNA procedures that work with millions of molecules. Facilities for ancient DNA handling must not house procedures with amplified fragments, as it is exceedingly difficult to prevent their spreading. DNA from modern organisms must be handled in a separate building, or at least a facility with a separate ventilation system. All these measures help to prevent

contamination but do not affect the contamination of the sample itself that remains as it was before it ever got into the laboratory. In order to decrease contamination, the surface layer of the sample is usually removed.

Contamination is usually a very important issue in ancient human or microbial DNA studies, since both human and bacterial DNA are constantly present in laboratory. Thus, the sequences of contaminant DNA are harder to separate from the authentic DNA than when dealing with exotic or rare species. The guidelines for working with ancient DNA and the criteria of authenticity for the amplified ancient DNA fragments are described in the following reviews [4, 17, 34–36] and are listed in Table 1.

Table 2. Various types of ancient DNA damage (from [4, 17] with modifications)

Type of damage	Cause of damage	Effect on DNA	Possible solution
Nucleobases and deoxyribose degradation	Postmortem destruction by intracellular nucleases, degradation by microorganisms and environmental chemical processes	Apurinization of DNA, strand breaks, decrease of DNA fragment size, decrease of the overall amount of DNA	Amplification of short (<100–200 bp) overlapping fragments
Cross-links that block PCR	Alkylation, Maillard reaction (chemical reaction between a sugar molecule and an amino group of a nucleobase or an amino acid)	Cross-links between DNA strands in a single molecule; cross-links between DNA strands of different molecules; or cross-links between DNA and proteins	Treating the sample with reagents that destroy cross-links
Deamination and other types of oxidative or hydrolytic DNA base modifications	Adenine → hypoxanthine Guanine → xanthine Cytosine → uracil 5-methyl-cytosine → thymine	Insertion during amplification of nucleotides that were not present in the original nonmodified template	Treatment by DNA uracil-N-glycosylase that removes cytosine deamination products. Determination of a consensus sequence based on multiple sequencing of the analyzed regions: Multiple independent PCR, cloning of the original template or PCR products, and sequencing several clones

IDENTIFYING CONTAMINATION

Possible contamination can be identified with a high degree of accuracy during ancient DNA studies if it is *a priori* supposed that the sample could be contaminated, so that the results are viewed with this possibility in mind. In order to identify laboratory contamination, researchers use “empty” extracts that have been processed along with the target sample but without adding tissue sample or DNA. Since contamination templates can be present in very low concentrations and not manifest themselves in each reaction, multiple control reactions are made, usually in proportion 1:5, but with no fewer than 1:1 to the extracted sample. Such “empty” extracts are used in all further analytic procedures in addition to the regular negative controls.

Independent confirmation of results in different laboratories is considered to be one of the strong indicators of authenticity. But even this is not an absolute guarantee [36].

Special attention must be paid to bioinformatic analysis of the obtained nucleotide sequences. Since analysis of ancient samples usually involves mtDNA analysis, it is important to compare the sequence not only with the mtDNA of species closely related to the sample source or with human mtDNA (the most likely source of contamination), but also with the nuclear homologues of mtDNA genes (nuclear mtDNAs, numts), whose similarity to mtDNA is ~98% and more in case of human mtDNA (such as the NT_004350.18 sequence located on Chromosome 1).

CHEMICAL MODIFICATIONS OF ANCIENT DNA AND POSTMORTEM MUTATIONS

Postmortem DNA alterations, and mutations during *in vitro* DNA amplification, are among the central methodological problems in ancient DNA and complex forensic DNA analysis. As opposed to metabolically active tissues that have an active DNA repair system postmortem cells accumulate chemical (hydrolytic or oxidative) DNA modifications and strand damage. Studies show that postmortem DNA damage includes strand breaks, loss of bases and cross-linking between strands that inhibits PCR. Postmortem alterations that modify bases but do not inhibit amplification are espe-

cially important, since they can cause the appearance in the amplification products nucleotide of changes that were not present in the authentic sequence (type I substitutions A → G / T → C and type II substitutions C → T / G → A) (Table 2). The manner how the degraded DNA templates are damaged depends on the samples age, their geographic origin, and the taphonomic conditions (preservation conditions) of the environment where the samples were stored. Postmortem alterations can appear in mutational hot-spots, thus simulating evolutionary mechanisms [37]. The manner and dynamics of accumulation of postmortem DNA alterations are under continuous research [38, 39]. DNA damage limits the size of the DNA fragments found in ancient samples to about 100–500

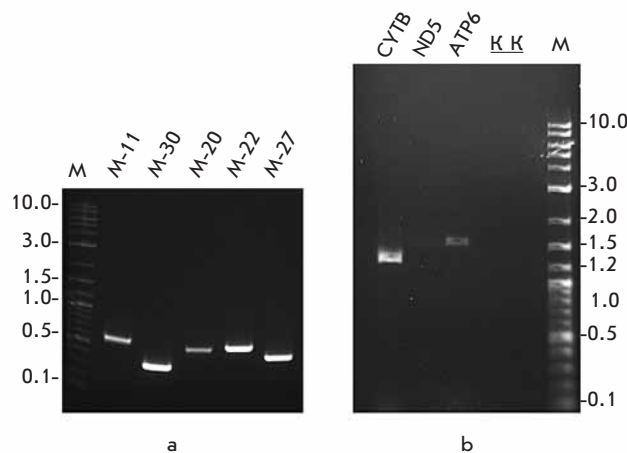


Fig. 2. A typical result of PCR amplification of mammoth mitochondrial genome fragments: a – PCR products of relatively short amplification fragments (300–600 bp); b – Successful amplification of long PCR fragments harboring entire mitochondrial gene sequences (1317 bp for the Cytb gene and 1613 bp for the ATP6 gene) and no amplification products are obtained for larger size PCR fragments (3054 bp for the ND5 gene). M – marker fragment sizes are indicated in kb, K – negative controls [13]

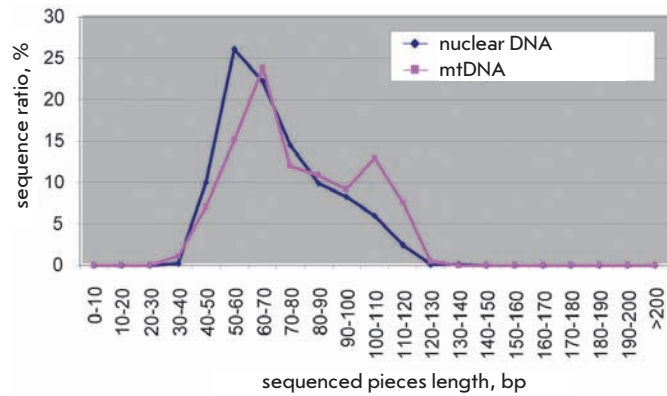


Fig. 3. The size distribution of genomic DNA fragments extracted from the ancient remains sequenced on the 454 platform. Unpublished data obtained in collaboration of E.I. Rogaev et al. with M. Blow and E. Rubin

bp. That is why the primers for ancient DNA PCR are usually chosen for no more than 200–300 bp fragments, although fragments of greater length have been obtained in some cases (Fig. 2).

Most of the published ancient DNA studies have been conducted on mtDNA that is found in hundreds and thousands of copies in cell and can be amplified more successfully than nuclear DNA. There are much fewer studies on nuclear DNA. Amplification, cloning, and sequencing of nuclear DNA from a mammoth *M. primigenius* sample obtained from Chukotka permafrost was performed in order to assess the quality of nuclear DNA preservation (E.I. Rogaev, E. Rubin, unpublished data). Most of the mammoth genome was fragmented into pieces of about 50–100 bp (Fig. 3). It indicates a relatively high quality of nuclear DNA preservation.

Postmortem modifications are randomly located in the preserved DNA fragments. For example, single nucleotide substitutions (~ 6 in 1 000 bp) were found in one study [13] during the cloning and sequencing of PCR-amplified mammoth DNA. This observation was taken into account for correct reconstruction of the complete mitochondrial genome of Chukotka mammoth *M. primigenius* (Fig. 4). The complete genome was obtained as a consensus of multiple overlapping fragments [13]. In order to additionally control the number of postmortem mutations, the overall number of substitutions in all mammoth mtDNA genes was calculated in comparison with elephant *E. maximus* mtDNA. The ratio between nonsynonymous (that cause aminoacid substitutions) and synonymous substitutions was calculated for the same purpose. It was shown that the number of substitutions in the mtDNA genes of the Chukotka mammoth [13] was lower than in the genes of a mammoth sequence obtained from remains found in Yakutia that was published at the same time by German researchers [40]. A comparative analysis showed that this difference was due to the unusually high number of substitutions in a 200–300 nucleotide region of the Yakut mammoth DNA in the locus of *ND1* and *ND2* genes, and that the number of nonsynonymous substitutions was greater than the number of synonymous ones (2:1 for the *ND1* gene and 7:2 for *ND2*). The *ND2* gene of the Chukotka mammoth had only one synonymous substitution, and *ND1* did not differ in any

way from the elephant gene [13]. All undetected postmortem mutations affect the result of phylogenetic reconstruction.

NOVEL DNA-SEQUENCING TECHNOLOGIES

Ancient DNA analysis involves sequencing of a large number of short fragments that have multiple overlapping of the same genomic regions. Low sequencing speed and high cost limit the usage of such research. Novel technologies of massively parallel sequencing of a large amount of DNA samples have appeared in the last 3–4 years, and the cost has dropped by two orders of magnitude. This novel technologies have given researchers sequencing possibilities that were previously available only to large genomic centers. Among the available novel technologies several were used in ancient DNA studies, such as clonal amplification followed by parallel sequencing of dense micropatterns of cloned DNA fragments by repeated enzymatic reaction cycles, with automatic registration of the signal from each cycle and every fragment.

The ordered spatial localization of the PCR amplicons on a chip or immobilization on microne-sized beads that are then placed into ordered cells allows to minimize the reaction mix volumes, thus decreasing the cost of the whole process.

Realization of these strategies involves several steps, and for each of them a specific technical approach has been developed. For example, preparation of DNA fragment libraries using PCR does not keep the ratio of amplification products identical to the ratio of original templates. Some DNA fragments are amplified more efficiently than others that could be lost. The problem can be solved by using emulsion PCR.

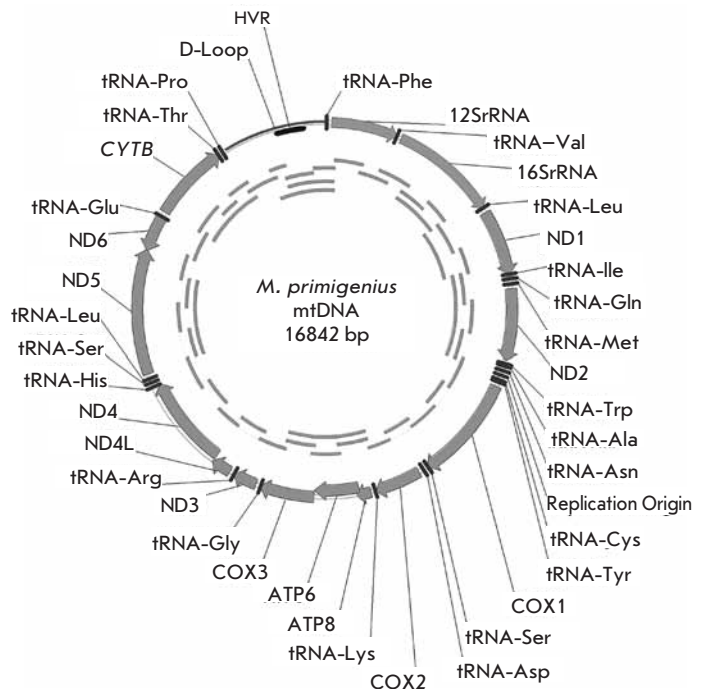


Fig. 4. Mitochondrial genome of the woolly mammoth *M. primigenius*. Determination of the nucleotide sequence of the complete mitochondrial genome was performed in two laboratories. The overlapping PCR-amplification products used for sequencing are shown in the inner circle [13]

The DNA solution is transferred into the mineral oil mix. The ratio is calculated so that each molecule of DNA is encapsulated in a separate lipid droplet that acts as a microreactor for the amplification process. This approach minimizes the loss of certain original templates. There are various technical solutions for fragment library preparation and for the other steps in the process, such as enzymatic reactions, visualization and computerized signal registration, data storage and analysis [41].

Novel sequencing technologies have certain limitations. Massively parallel pyrosequencing, accomplished by using 454 Life Science system (Genome Sequence 20TM DNA sequencing System: GS20, Roche/454 Life Science), provides a 100-fold increase in the sequencing speed, compared to the standard capillary electrophoresis method. Up to 25 million nucleotides are analyzed in a single run. Only a short sequence can be read (usually less than 250–400 bp). It is not much of a limitation for ancient DNA studies, since most of the DNA fragments are precisely of this size.

The Illumina technology, named Solexa (after the name of the company that developed this approach), and SOLiD (ABI corporation) allow the analysis of up to 1 billion nucleotides in a single run but read sequences of only 30–40 nucleotides (last year this number was just 25 nucleotides). The availability of full human genomes and the genomes of commonly used model organisms as reference sequences allow to map the short fragments into a single contig.

Another limitation of these novel platforms is the 10-fold decrease in the accuracy of sequencing, compared to the Sanger method. Nevertheless, these technologies are very promising, and they can be expected to improve in efficiency and quality in the nearest future.

ANALYSIS OF DEGRADED DNA IN FORENSIC EXPERTISE OF HISTORIC SAMPLES

The technological approaches applied for ancient DNA study can also be used for forensic genetic analysis in difficult cases where only microscopic amounts of material are available or the DNA has been severely damaged. Some of these approaches were used in the genetic expertise of the putative remains of the family of the last Russian Emperor Nicholas II Romanov. In the early 1990s, a first grave with human remains was found near Yekaterinburg. During the investigation, it was suggested that the remains belong to the family of the Russian Emperor Nicholas II Romanov, his wife, the Empress Alexandra Fedorovna, their 3 daughters, the court physician, and three servants. They are all thought to have been murdered in 1918 [42–44]. However, the remains of two children of the Romanov family were not identified, and their fate remained unknown. Among other hypotheses, there has been a legend that Alexey and Anastasia, the youngest children of the Romanov family, had survived those turbulent times. In July 2007, a second grave was found not far from the first one. It contained burned bone fragments from two skeletons. Forty-four bone fragments were found in the second grave, all severely damaged by fire and presumably sulfuric acid. Preliminary anthropological analysis of the half-burned bone fragments from the second grave suggested that the bones belonged to a boy 10–14 years of age and a young woman of about 18–23. The least damaged frag-

ments of the femoral bones from both the male and female skeletons were selected for genetic analysis, and they were labeled Samples 146 and 147, respectively. Samples from the first grave were also collected for a more detailed study, and reference samples were taken from living relatives of Nicholas Romanov and Alexandra Fedorovna. Furthermore, swabs of blood stains from a shirt that had belonged to Nicholas II and is stored in the Hermitage museum were also used for analysis. The study included the following steps: preparation of the samples for DNA extraction; DNA extraction; quantification of the extracted total DNA and human-specific DNA; amplification and sequencing of the mitochondrial hyper-variable regions, and later sequencing and reconstruction of the complete mtDNA (cmtDNA) sequence; determination of the STR-profiles of the Y-chromosome; determination of the autosomal STR-profiles; additional sex identification with the use of a special marker designed for degraded DNA analysis [45, 46]; and extraction and analysis of modern DNA from Romanov family relatives and their comparison to historic samples. The steps and methods of DNA identification are described in Table 3 [45].

MITOCHONDRIAL GENOME ANALYSIS

Complete nucleotide sequences of the mitochondrial genome have been determined for the putative remains of Nicholas II and Alexandra Fedorovna from the first grave; and the putative remains of Alexey and his sister, from the second grave. The mitotypes of the putative remains of Nicholas II and Alexandra Fedorovna are from the common European mtDNA haplogroups T2 and H1.

Complete cmtDNA sequences were also determined for the relatives of Queen Victoria (1819–1901) for 2 maternal lineages, the descendants of princess Victoria, sister of Alexandra Fedorovna, and their aunt Beatrice (Fig. 5). Their cmtDNA were absolutely identical with those extracted from the putative remains of Alexandra Fedorovna and the 2 children from the second grave. Searches performed in the available cmtDNA databases (Table 4) showed that not one of the available tens of thousands of sequences identifies with this cmtDNA, which was named “Queen Victoria mitotype.” Thus, the first and second burial sites really do contain the remains

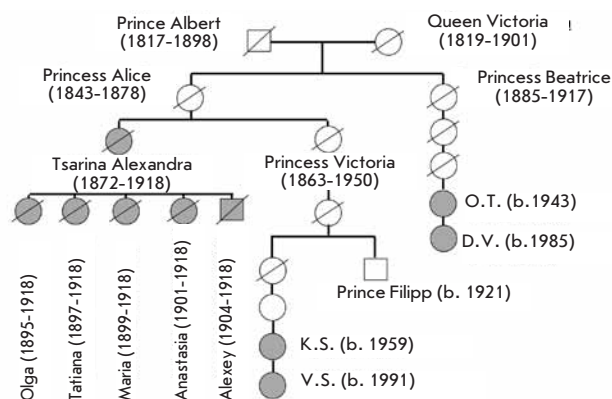


Fig. 5. Maternal lineages of Empress Alexandra Fedorovna. The family members whose mitochondrial DNA was determined are indicated in grey

REVIEWS

Table 3. Methodical approaches specific to analyses of degraded DNA from historical specimens[45]

Stage of analysis	Special procedures	Reagents and methods
Preparation of historical samples	Independent analysis in specialized laboratories in VIGG (Moscow) and University of Massachusetts Medical School (Worcester, USA).	Physical and chemical cleaning of small bone fragments, crushing or drilling to obtain bone powder.
Extraction of DNA from bone specimens	All the experimental procedures were performed in sterile PCR-hoods, in accordance with standards for ancient DNA research, with all the safety precautions, avoiding contamination of the samples by modern DNA.	DNA was extracted from ~170–750 mg of cleaned bone material treated by 0.5 M EDTA and proteinase K and was then purified by a QIAquick PCR purification kit (Qiagen) in accordance with the manufacturer's protocol with minor modifications.
Extraction of DNA from archival blood-stains	The biological material was obtained from 4 different blood stains. At least 3 swabs were taken from each spot. In order to minimize contamination, DNA was extracted only from the 2 nd and 3 rd swabs for each spot.	DNA was extracted with the QIAamp DNA Mini Kit (Qiagen) using the manufacturer's protocol ("DNA Purification from Dried Blood Spots") with our own modifications.
Quantification DNA analysis		The total extracted DNA was quantified by the Quant-iT™ PicoGreen® Assay kit (Invitrogen), human specific DNA was quantified by the Plexor® HY assay kit (Promega) and the 7500 Real-Time PCR System (Applied Biosystems).
Sequencing of HVR1 and HVR2 of mtDNA from historical samples	Potential contamination by foreign DNA was monitored by using negative controls (amplification of "empty" extracts and PCR without addition of the template).	mtDNA fragments were amplified as short overlapping fragments. The PCR products were then extracted from the agarose gel using a QIAquick Gel Extraction kit or a MinElute Gel Extraction kit. For additional studies, the PCR products from some specimens were cloned.
Sequence analysis of the complete mitochondrial genome, extracted from bone specimens.	Since the DNA was highly degraded, short overlapping fragments sized 164–383 bp were obtained, covering whole mitochondrial genome.	Because the amount of DNA was limited, multiplex amplifications were performed using 88 pairs of specially designed primers grouped into 3 sets, and then the products of the first PCR rounds were amplified with individual primer pairs. The secondary PCR products were then sequenced using three different strategies
Analysis of the mtDNA extracted from the blood stains on shirt that belonged to Nicholas II.	Extraction of DNA was performed from different swabs from at least 3 shirt blood stains. Up to 5 or 7 PCR replications were conducted for some of the mtDNA SNPs to identify the potential heterogeneity and contamination by other individuals and selection of the extracts presenting mitotype likely from one individual only.	Since the quality of preservation in the blood stains was unknown, initially a set of primers was developed for the amplification of short (64–109 bp) DNA fragments that would include very rare SNPs identified in the previous analysis of Skeleton №4 (the putative skeleton of Nicholas II).
Extraction and analysis of DNA from modern samples.	For modern DNA analysis all the procedures were performed in separate buildings, physically separated from the ancient DNA laboratories. Informed consent was obtained from all living relatives participating in the study.	DNA obtained from buccal swabs or drops of blood was extracted using standard protocols. PCR was performed using a set of primers for amplifying longer fragments.
Assembly of fragments into a continuous nucleotide sequence (contig).		The sequences were assembled using Seqman software, DNASTAR, and the revised Cambridge reference sequence (rCRS, accession number AC_000021) as a standard.
Sex identification.		Sex was identified using the standard method, amplification of a fragment of the amelogenin gene using several commercial kits: AmpFISTR® MiniFiler™ (Applied Biosystems) and PowerPlex S5 (Promega). Specially designed primers for the amplification of short fragments specific to the X- and Y-chromosomes were also used.
Analysis of nuclear STR markers.	mtDNA or nuclear DNA extracts that consisted of a mix of individual profiles were discarded from further analysis. Each sample from various extracts was amplified in multiple replications. Homozygous loci were considered authentic if multiple independent amplifications confirmed a certain allele for the autosome STR-marker.	The following kits were used in order to obtain autosomal STR profiles from bone samples of the first and second burial sites: AmpFISTR® MiniFiler™ PCR Amplification Kit (Applied Biosystems) and PowerPlex S5 System (Promega), specifically designed for analyzing degraded DNA.
STR-profiles of the Y chromosome.		The AmpFISTR® Yfiler™ (Applied Biosystems) kit was used, according to the manufacturer's protocol with minor modifications for analysis of degraded DNA.
Electrophoresis analysis	In order to increase the signal intensity and lower "noise" in the STR-profiles, the products of multiplex amplification were sometimes purified with Qiagen MiniElute columns.	Electrophoretic analysis was performed using a 96-capillary sequencer 3730xl DNA Analyzer (Applied Biosystems). The results were analyzed with GeneMapper® ID software v3.2 (Applied Biosystems).

of Queen Victoria's granddaughter, great granddaughter, and great grandson.

Determination of the cmtDNA from the putative remains of Nicholas II confirmed the earlier described heteroplasmy (the coexistence of several mtDNA types) at the 16169C/T site. Moreover, nucleotide substitutions were found in the coding region of the mtDNA, including the extremely rare

2850C variant in the 16S rRNA gene (population frequency approximately 0.004). Nucleotide substitutions in the mtDNA from the putative remains of Nicholas II and from the blood stains on the shirt were completely identical [45]. The ratio between the heteroplasmic mtDNA variants was similar both in the remains and in the blood. Nicholas II's brother, George, whose remains were studied previously, also had heteroplas-

REVIEWS

Table 4. Databases used for analyzing the population frequencies of the determined genetic profiles [45, 49]

Query sequence	Number of samples in the database	Name of the database
Complete mitochondrial genome	71,664 2,704	Mitosearch (www.mitosearch.org) contains sequences of the hypervariable region; mtDB, Human Mitochondrial Genome Database (www.genpat.uu.se/mtDB/);
	14,486	EUROS, our own database, compiled from our own and other authors' published data for the USSR and Russian populations and for certain European populations (German, English, etc.).
STR-haplotype of the Y-chromosome	4,163	U.S. Consolidated Y-STR Database has genotypes for 15-17 Y-STR loci (http://www.usystrdatabase.org/);
	1,261	Data for Russians from the Y Chromosome Haplotype Reference Database (YHRD, www.yhrd.org/index.html).
Autosome STRs	> 50,000	For this study we combined the non-overlapping data for European populations from two large databases: ALFRED (http://alfred.med.yale.edu/alfred/) and The Distribution of the Human DNA-PCR Polymorphisms database (http://www.uniduesseldorf.de/WWW/MedFak/Serology/database.html) and also added the recently published data on the Russian population.
Mutation in the IVS3-3 position of the F9 gene		Population SNP database (NCBI, HapMap Project); ~ 928 X-chromosomes of healthy individuals of European descent whose genotypes were assessed for this SNP using PCR-RFLP (Restriction Fragment Length Polymorphism) especially for this study; Hemobase: Hemophilia B mutation registry, Haemophilia B mutation database.

my at the same site [43]. The descendants of Kseniya, sister of Nicholas II, had the homoplasmic 16169 T variant, and the previously performed mtDNA study of Nicholas II's sister's son found the homoplasmic 16169C variant [44]. Not one of the reconstructed complete mtDNA sequences ("Dagmar" type) was identified in available databases (Table 4), which confirms the assumption that the studied remains belong to Nicholas II. The data suggest that the brothers Nicholas II and George Romanov inherited heteroplasmy from their mother Maria Fedorovna (Princess Dagmar), and that her descendants had this heteroplasmy segregated into two distinct mtDNA variants in two generations. The nucleotide substitution at position 16169, on which the heteroplasmy of the "Dagmar type" is based, is located in the noncoding (hypervariable) region of the mtDNA characterized by increased polymorphism compared to the coding regions. The population frequency of heteroplasmy for point substitutions in hypervariable regions averages 6% [47]. It is assumed that rapid segregation of heteroplasmic mtDNA variants in descendants is caused by mechanisms of biological "bottle-neck" (a massive decrease in the number of mtDNA copies) during oocyte maturation in the postnatal folliculogenesis in mammals [48].

NUCLEAR STR-MARKER ANALYSIS

To study the paternal lineage DNA profiles of the putative remains of Emperor Nicholas II and Prince Alexey, the STR-haplotypes of the Y-chromosome were determined. Specialized procedures were developed in order to increase the PCR sensitivity, since the amount of the available DNA was limited, and the molecules were highly fragmented (some of the methods are described in Table 3) [45]. The STR-profiles were determined from multiple independent PCR amplifications using no less than three different DNA extracts. Only the alleles that were identified in no less than 2 amplifications were considered authentic. A full Y-STR profile for the bone specimen of Skeleton №4 and for the museum samples of Nicholas II's blood was obtained using these criteria. Low-copy highly fragmented DNA often loses single STR alleles. Marker DYS385 shows two loci on the Y-chromosome. The high mo-

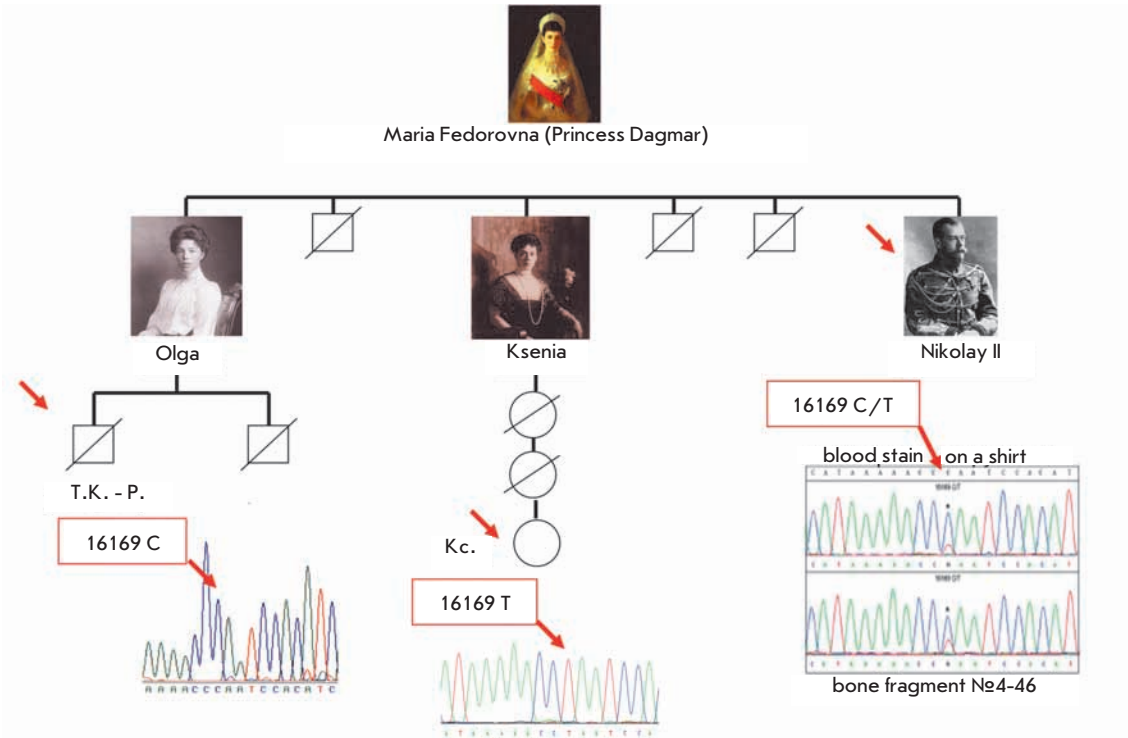
lecular weight allele (DYS385/ 14) was identified only once in the repeated experiments with the DNA extracted from Sample #146, thus this allele for Sample #146 is indicated as not determined (ND). DNA isolated from the archival Nicholas II bloodstain and DNA obtained from Romanov paternal lineage family members were used as reference samples (Fig. 6). Y-chromosome STR-profiles of the studied samples and the reference sequences were completely identical (Fig. 7 and Table 5). This 17-locus Y-STR haplotype is unique. It is not found in large population databases for multi-locus Y-STR (Table 4) and was first encountered in the described study [45].

Further gender and autosomal chromosome genotyping with STR multiplex systems developed especially for degraded DNA demonstrated that the male (Sample № 146) and the female (Sample № 147) from the second grave have

Table 5. STR-haplotype analysis of the Y-chromosome [45]

Markers	№ 4	№ 146	Members of the Romanov family	Archive blood stain from a shirt	Control DNA ABI, 007
DYS456	16	16	16	16	15
DYS389I	13	13	13	13	13
DYS390	24	24	24	24	24
DYS389II	29	29	29	29	29
DYS458	17	17	17	17	17
DYS19	14	14	14	14	15
DYS385	11, 14	11, ND	11, 14	11, 14	11, 14
DYS393	13	13	13	13	13
DYS391	10	10	10	10	11
DYS439	11	11	11	11	12
DYS635	24	24	24	24	24
DYS392	13	13	13	13	13
Y-GATA-H4	12	12	12	12	13
DYS437	15	15	15	15	15
DYS438	12	12	12	12	12
DYS448	19	19	19	19	19

Fig. 6. Heteroplasmy at the 16169C/T locus in the mitochondrial (maternal) lineages of Emperor Nicholas II



autosomal STR profiles nonidentical to any STR profiles determined for Romanov family remains from the first grave but consistent with a biological kinship connection (Fig. 8). These data clearly demonstrate that these newly found remains may belong to Prince Alexey and one of the daughters of the imperial family. The available nuclear DNA analysis data, supported by the anthropological data, prove that the remains from the second grave belong to a young woman (№ 147) and a boy (№ 146), and that samples from the second grave are not from Skeleton № 7 (putative mother, Empress Alexandra Fedorovna) or Skeleton № 4 (putative father, Emperor Nicholas II); however, they are related through the paternal and maternal lineages.

The statistical evaluation (likelihood ratio) based on three identification approaches for determining whether the bones belong to Nicholas II, and not to any other random individual, is on the order of a septillions ($> 10^{26}$) [45]. Taken together our data establish beyond reasonable doubt that the studied remains belong to the last Russian Emperor Nicholas II Romanov, his wife the Empress Alexandra Fedorovna, their 4 daughters (the Grand Duchesses Olga, Tatiana, Maria, and Anastasia), and their son (Prince Alexey).

HEMOPHILIA: SEARCH FOR MUTATIONS IN THE GENES FOR BLOOD CLOTTING FACTORS

There is historical evidence that Prince Alexey suffered from severe bleeding that is characteristic of hemophilia. It is now known that hemophilia is caused by insufficient activity of blood clotting factors. Factor VIII deficit caused by mutations in the *F8* gene is the cause of the most common hemophilia A (about one in 5 000 boys is born with this disease), and Factor IX deficit causes hemophilia B (*F9* gene), which occurs 5 times less often.

A few hundred mutations that cause hemophilia have been described to this day. Both of the blood clotting factor genes are localized on the X-chromosome; that is why males carrying the mutant gene exhibit the disease. Females carrying a single copy of the mutant gene and a copy of the normal one are healthy in most cases, although some of them (10 %) can exhibit decreased efficiency of blood clotting. Females can be assessed for hemophilia mutations when their sons have hemophilia.

This inherited disease was common in the royal families of Europe, the sons, grandsons, and great grandsons of Queen Victoria (Fig. 9, c). The Queen herself did not suffer from this illness, but it seems that she carried the mutant gene. There is no evidence of hemophilia in any of her present living relatives.

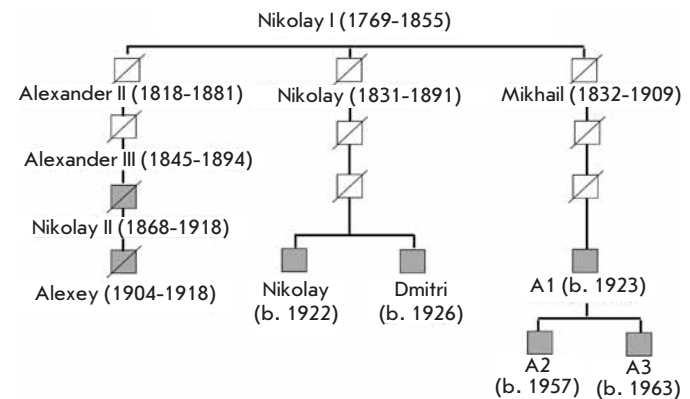


Fig. 7. The paternal lineages of the Romanov family. The family members whose DNA was studied are indicated in grey

REVIEWS

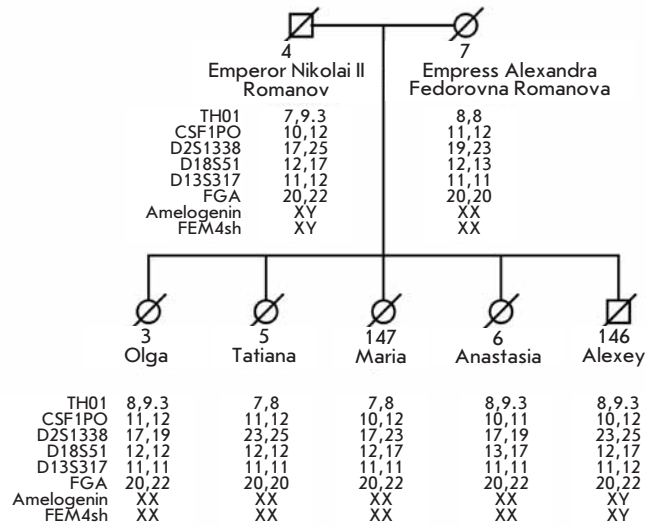


Fig. 8. Analysis of the sex chromosome and autosomal STR-markers in DNA extracted from bone specimens [45]

In order to determine whether Alexandra Fedorovna or Prince Alexey carried any mutations in genes, all the exons and the intron-exon boundaries of *F8* or *F9* genes were amplified using multiplex amplification reaction and then sequenced by massively parallel sequencing. Miniscule amounts of DNA and its high level of degradation required special procedures for the identification of the nucleotide sequences that included the 26 exons of the *F8* gene and the 8 exons of

the *F9* gene (the strategy and methods used in this study are described in Table 6).

The first step was to analyze DNA extracts from Alexandra that showed no contamination based on the results of mtDNA and autosomal STR-marker analysis. Amplification of the *F8* gene and the 8 exons of the *F9* gene was performed in parallel with the amplification of mtDNA that was used as a control and for accurate identification of the sample. Non-synonymous substitutions or insertion-deletion mutations were not found in either the *F8* or *F9* gene. However, we detected a single nucleotide substitution of A =>G at an intron-exon boundary and, 3 nucleotides before the start of the 4th exon of the *F9* gene (referred to as IVS3-3A>G according to standard nomenclature). As expected for a heterozygote carrier, Alexandra Fedorovna carried both mutant and wild-type sequences. Alexey bear only the mutant allele, meaning that he was homozygous for this mutation, while one of his sisters (presumably Anastasia) was a heterozygous carrier of the mutation. The other sisters did not carry the mutant alleles; they were homozygous carriers of the wild-type allele (Fig. 9) [49].

Bioinformatic analysis showed that the IVS3-3A>G mutation activates the cryptic splicing acceptor site, which leads to frame-shift during the translation of the *F9* gene mRNA and the appearance of a premature stop-codon (Fig. 9, a).

The effect of this mutation on RNA splicing was studied by expressing the mutant fragment of the *F9* gene in a cell culture using a specialized recombinant Exontrap vector (MoBiTec). We found that 99.98 % of transcripts were generated by splicing at the mutant site. Less than 1% of the activity of Factors VIII or IX leads to severe manifestations of hemophilia [50]. Population database searches and genotyping

Table 6. Methods used in screening for mutations of the *F8* and *F9* genes [49]

Stage of analysis	Special procedures	Reagents and methods
Multiplex PCR-amplification of the Factor VIII gene, <i>F8</i> (26 exons) and Factor IX, <i>F9</i> gene (8 exons), located on the X-chromosome	Extracts that were shown to be free of contamination (in analysis of mtDNA) were used further for nuclear gene study. Negative controls were used: PCR-amplification of "empty" extracts (all DNA extraction steps were performed, but with no bone specimen material added in the course of the extraction procedures) and PCR amplification without addition of extracted DNA.	~ 210 pairs of primers were designed for multiplex PCR amplification of short overlapping sequences (< 200 bp) that would cover all the exons and the intron-exons boundaries of the <i>F8</i> and <i>F9</i> genes. The primers were grouped into 14 sets for the <i>F8</i> gene and 3 sets for the <i>F9</i> gene, with each set consisting of 7 to 30 pairs of primers. About 100 pg of human DNA (~ 16-17 diploid genomes) were used for the initial multiplex PCR.
Sequencing	Sequencing of the blood clotting factors genes <i>F8</i> and <i>F9</i> was done in parallel with the mitochondrial genome to have a control for contamination and unequivocal identification of the sample.	Individual PCR-fragments were excised and purified from a 2.5 % agarose gel and then sequenced using two strategies. One involved mixing the PCR products in equimolar amounts and using them for massively parallel sequencing (Illumina GA). The other approach involved the direct sequencing of individual PCR-products on a 96-capillary sequencer 3730xl DNA analyzer (Applied Biosystems).
Genotyping of the identified <i>F9</i> gene mutation	8 independent amplifications of the N7 sample (Empress Alexandra) were performed. For other bone samples, 2 to 7 independent amplifications were analyzed for each sample.	The mutation that was initially found during the DNA analysis of Skeleton N7 was verified by sequencing of ultra-short amplicons (63 bp and 83 bp) obtained with specially designed primers. The same primers were used to amplify DNA from bone specimens N146 (Prince Alexey) and N3, N5, N6, N147 (Nicolas II's and Alexandra's daughters).
Analysis of splicing products		The amplified fragment with mutations in <i>F9</i> gene was cloned into the pET01 Exontrap vector (MoBiTec). After verifying the sequence of the recombinant molecule, it was used for cell culture transfection. cDNA derived from transfected cell cultures was amplified, and the PCR product library was sequenced (Illumina GA). The spliced sequences were identified in 812,114 reads. 99.982 % of the transcripts were spliced at the mutant site, and only 0.018 % were spliced at the wild-type site.

REVIEWS

to the extraction of genetic information from organisms that have disappeared tens of thousands of years ago, and to the reconstruction of evolutionary events, which was hitherto unachievable in experimental study. These findings opened new possibilities for precise molecular genetic analysis of severely damaged and decayed DNA, which has already raised the standard of applied procedures in forensic medicine. The results reviewed in this paper could not have been obtained without the development of novel DNA technologies that can

now be incorporated into the everyday routines of fundamental and applied research, making them more reliable, fast and informative, as well as lowering costs. ●

The study was supported by the Russian Federal Agency for Science and Innovation (federal contract 02.512.11.2231) and by the Program "Biodiversity" of Presidium of Russian Academy of Sciences.

REFERENCES

1. Higuchi R., Bowman B., Freiberger M., Ryder O.A., Wilson A.C. // *Nature*. 1984. 312 (5991). P. 282–284.
2. Higuchi R.G., Wrishnik L.A., Oakes E. et al. // *J. Mol. Evol.* 1987. 25 (4). P. 283–287.
3. Pääbo S. // *Nature*. 1985. 314. P. 644–645.
4. Willerslev E., Cooper A. Ancient DNA. // *Proc. Biol. Sci.* 2005. 272 (1558). P. 3–16.
5. Poinar H.N., Höss M., Bada J.L., Pääbo S. // *Science*. 1996. 272 (5263). P. 864–866.
6. Smith C.L., Chamberlain A.T., Riley M.S. et al. // *Nature*. 2001. 410 (6830). P. 771–772.
7. Gilbert M.T., Wilson A.S., Bunce M. et al. // *Curr. Biol.* 2004. 14. P. R463–R464.
8. Shapiro B., Drummond A.J., Rambaut A. et al. // *Science*. 2004. 306 (5701). P. 1561–1565.
9. Willerslev E., Hansen A.J., Brand T.B. et al. // *Science*. 2003. 300 (5620). P. 792–795.
10. Willerslev E., Hansen A.J., Brand T.B. et al. // *Curr. Biol.* 2004. 14 (1). P. R9–R10.
11. Barnes I., Matheus P., Shapiro B., Jensen D., Cooper A. // *Science*. 2002. 295. P. 2267–2270.
12. Lambert D.M., Ritchie P.A., Millar C.D. et al. // *Science*. 2001. 295 (5563). P. 2270–2273.
13. Rogaev E.I., Moliaka Y.K., Malyarchuk B.A. et al. // *PLoS Biol.* 2006. 4 (3). P. e73.
14. Saiki R.K., Scharf S., Faloona F. et al. // *Science*. 1985. 230 (4732). P. 1350–1354.
15. Mullis K., Faloona F., Scharf S. et al. // *Cold Spring Harb Symp Quant Biol.* 1986. 51 Pt 1. P.263–373.
16. Saiki R.K., Gelfand D.H., Stoffel S. et al. // *Science*. 1988. 239 (4839). P. 487–491.
17. Pääbo S., Poinar H., Serre D. et al. // *Annu. Rev. Genet.* 2004. 38. P. 645–679.
18. Ho S.Y., Gilbert M.T. Ancient mitogenomics. // *Mitochondrion*. 2010. 10 (1). P. 1–11.
19. Ramakrishnan U., Hadly E.A. // *Mol Ecol.* 2009. 18 (7). P. 1310–1330.
20. Hofreiter M., Stewart J. // *Curr Biol.* 2009. 19 (14). P. R584–R594.
21. Woodward S.R., Weyand N.J., Bunell M. // *Science*. 1994. 266 (5188). P. 1229–1232.
22. Collura R.V., Stewart C.B. // *Nature*. 1995. 378 (6556). P. 485–489.
23. Cooper A., Wayne R. // *Current Opinion in Biotechnology*. 1998. 9. P. 49–53.
24. Noonan J.P., Coop G., Kudaravalli S. et al. // *Science*. 2006. 314 (5802). P. 1113–1118.
25. Green R.E., Krause J., Ptak S.E. et al. // *Nature*. 2006. 444 (7117). P. 330–336.
26. Wall J.D., Kim S.K. // *PLoS Genet.* 2007. 3 (10). P. 1862–1866.
27. Green R.E., Malaspina A.S., Krause J. et al. // *Cell*. 2008. 134 (3). P. 416–426.
28. Pennisi E. // *Science*. 2009. 323 (5916). P. 866–871.
29. Krings M., Stone A., Schmitz R.W. et al. // *Cell*. 1997. 90. P. 19–30.
30. Schmitz R.W., Serre D., Bonani G. Et al. // *Proc. Natl. Acad. Sci. USA*. 2002. 99 (20). P. 13342–13347.
31. Noonan J.P., Hofreiter M., Smith D. et al. // *Science*. 2005. 309 (5734). P. 597–599.
32. Poinar H.N., Schwarz C., Qi J. et al. // *Science*. 2006. 311 (5759). P.392–394.
33. Schwarz C., Debruyne R., Kuch M. et al. // *Nucleic Acids Res.* 2009. 37 (10). P. 3215–3219.
34. Cooper A., Poinar H.N. // *Science*. 2000. 289 (5482). P. 1139.
35. Binladen J., Wiuf C., Gilbert M.T. et al. Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. // *Genetics*. 2006. 172 (2). P. 733–741.
36. Endicott P., Sanchez J.J., Pichler I. et al. // *BMC Genet.* 2009. 10. P. 29.
37. Pusch C.M., Broghammer M., Nicholson G.J. et al. // *Mol. Biol. Evol.* 2004. 21 (11). P. 2005–2011.
38. Gilbert M.T., Hansen A.J., Willerslev E. et al. // *Am J. Hum. Genet.* 2003. 72 (1). P. 48–61.
39. Lamers R., Hayter S., Matheson C.D. // *J. Mol. Evol.* 2009. 68 (1). P.40–55.
40. Krause J., Dear P.H., Pollack J.L. // *Nature*. 2006.439 (7077). P. 724–727.
41. Shendure J., Ji H. // *Nat. Biotechnol.* 2008.,6 (10). P. 1135–1145.
42. Gill P., Ivanov P.L., Kimpton C. et al. // *Nat. Genet.* 1994. 6 (2). P. 130–135.
43. Ivanov P.L., Wadhams M.J., Roby R.K. et al. // *Nat. Genet.* 1996. 12 (4). P. 417–420.
44. Rogaev E.I. Analysis of mitochondrial DNA of the alleged remains of Nicholas II and his nephew. // In: "Repentance". Government commission materials: Moscow, 1998. P. 171–182.
45. Rogaev E.I., Grigorenko A.P., Moliaka Y.K. et al // *Proc. Natl. Acad. Sci U S A*. 2009a. 106 (13). P. 5258–5263.
46. Grigorenko A.P., Andreeva T.V., Rogaev E.I. // *Medical Genetics*. 2009. 8 (4). P. 45–46.
47. Irwin J.A., Saunier J.L., Niederstatter H. et al. // *J. Mol. Evol.* 2009.68. P. 516–527.
48. Wai T., Teoli D., Shoubridge E.A. // *Nat. Genet.* 2008. 40 (12). P. 1484–1488.
49. Rogaev E.I., Grigorenko A.P., Faskhutdinova G., Kittler E.L., Moliaka Y.K. // *Science*. 2009b. 326 (5954). P. 817.
50. White G.C., Rosendaal F., Aledort L.M. et al. // *Thromb. Haemost.* 2001. 85 (3). P. 560.

Genome Paths: A Way to Personalized and Predictive Medicine

V.S. Baranov

Ott's Institute of Obstetrics and Gynecology, Russian Academy of Medical Sciences
E-mail: baranov@vb2475.spb.edu

ABSTRACT The review is devoted to the impact of human genome research on progress in modern medicine. Basic achievements in genome research have resulted in the deciphering of the human genome and creation of a molecular landmarks map of the human haploid genome (HapMap Project), which has made a tremendous contribution to our understanding of common genetic and multifactorial (complex) disorders. Current genome studies mainly focus on genetic testing and gene association studies of multifactorial (complex) diseases, with the purpose of their efficient diagnostics and prevention. Identification of candidate ("predisposition") genes participating in the functional genetic modules underlying each common disorder and the use of this genetic background to elaborate sophisticated measures to efficiently prevent them constitutes a major goal in personalized molecular medicine. The concept of a genetic pass as an individual DNA databank reflecting inherited human predisposition to different complex and monogenic disorders, with special emphasis on its present state, and the numerous difficulties related to the practical implementation of personalized medicine are outlined. The problems related to the uncertainty of the results of genetic testing could be overcome at least partly by means of new technological achievements in genome research methods, such as genome-wide association studies (GWAS), massive parallel DNA sequencing, and genetic and epigenetic profiling. The basic tasks of genomics today could be determined as the need to properly estimate the clinical value of genetic testing and its applicability in clinical practice. Feasible ways towards the gradual implementation of personal genetic data, in line with routine laboratory tests, for the benefit of clinical practice are discussed.

Keywords: genomics, genetic testing, multifactorial diseases, genetic passport, genome wide association studies.

Abbreviations: GP – genetic polymorphism, FGU – a functional genetic unit; GT – genetic testing, genetic testing, MD – multifactorial diseases; GWAS – genome wide association studies; SNP – single nucleotide polymorphism, PS – polymorphic sites; STR – short tandem repeats; HapMap – haploid genome; VNTR – Variable Number Tandem Repeats; CNV – copy number variation.

INTRODUCTION

The revolutionary achievements in human genetics such as full genome sequencing, the successful completion of the HapMap program (haploid genome), the rapid development of bioinformatics and nanotechnology, advances in the development of efficient methods for genome analysis are signs of a new era: the era of genomics. The 21st century could well be remembered as the century of genetics [1, 2]. Impressive progress in comparative and functional genomics has assured the widespread introduction of this branch in medicine. Thus, it has led to the emergence and rapid development of medical genomics in which such problems of classical medicine as diagnosis, prevention, and treatment are solved at the level of nucleic acids and the products of their expression: RNA and proteins [3, 4, 5]. The preventive direction in molecular medicine gave rise to predictive medicine (PM). Its main features appear to be its individual (the genome of every human individual) and preventive character (analyses of the genome are possible at any stage of ontogeny, long before the onset of a particular disease). The main principles of predictive medicine and genetic testing (GT) as a methodological basis for PM, as well as the concept of a "genetic passport," were formulated by us in 2000 [5, 6, 37].

1. POLYMORPHISM OF GENES AS THE BASIS OF PREDICTIVE MEDICINE

The genomes of all people, except for identical twins, are different. Pronounced population, ethnic, and, most importantly, individual genome features in translatable parts of genes

(exons), as well as in their noncoding sequences (intergenic spaces and introns), are caused by mutations leading to genetic polymorphism (GP). The latter is usually defined as a mendelian trait occurring in the population, at least in two variants, with a frequency of not less than 1% for each one [10]. GPs may be quantitative or qualitative.

Quantitative GP is represented by facultative elements, which account for up to 50% of the entire genome. These are micro- and minisatellite DNA, as well as DNA composed of tandem repeats (STR – Short Tandem Repeats), retrotransposons, and extensive repeats with variability in core nucleotide sequence composition – VNTR (Variable Number Tandem Repeats). Finally, in recent years, thanks to new methods of DNA analysis (CNV – Copy Number Variation and GWAS – Genome Wide Association Studies), the existence of polymorphism in large DNA fragments (1-50 MGB) was demonstrated in the human genome; it is the so-called Copy Number Variation – CNV.

Qualitative GP is mainly single nucleotide polymorphism (SNP). This most frequent GP occurs roughly every 300-400 bp. Thus, the total number of SNP in the full human genome is estimated at about $10-13 \cdot 10^6$. There is a surprising similarity (99.9%) in GP in the genomes of different people. Stable combination of several neighboring alleles – SNP on one strand of DNA (haplotype) – has allowed to use them as specific molecular markers in the program HapMap (haploid map) (see below).

It is assumed that about half of all SNP (5 million) are related to the sense (expressed) part of the genome. It is these

substitutions that correspond to the allelic variants of genes that cause or are associated with various diseases. They play a key role in human GP [5, 7, 8, 9].

It is well known today that polymorphism is typical for almost all human genes. It has been established that it has a clear ethnic and population specificity. Polymorphisms affecting the coding parts of genes often lead to the replacement of amino acids and the appearance of proteins with new functional properties. Replacements or nucleotide repeats in the regulatory (promoter) regions of genes may have a significant impact on gene expression. Inherited gene changes play a crucial role in determining the unique biochemical profile of each individual, his or her hereditary predisposition to various MD.

2. HAPLOID MAP (HAPMAP) PROGRAM.

The key role in the study of GP is played by the international project for the study of the haploid human genome – Haploid Map (HapMap).

The purpose of the project is to obtain a genetic map of the distribution of single-nucleotide substitutions (SNP) in the haploid set of all 23 human chromosomes [11]. The main thrust of the project consists in that the analysis of the distribution of known SNP in individuals for several generations revealed that SNPs that are adjacent or closely located in the DNA of one chromosome are inherited as SNP blocks. This block represents a SNP haplotype – a set of alleles located on the same chromosome (hence the origin of the name of the project HapMap). Thus, each of the mapped SNP acts as an independent molecular marker. By linking these SNP-markers with the manifestations being investigated (disease symptoms), the most likely location of candidate genes' mutations (polymorphisms) associated with a particular MD are identified. Typically, there are 5 or 6 SNPs closely linked with the already known Mendelian traits selected for mapping. Well described SNPs with a frequency of rare alleles of not less than 5% are called marker SNPs (tagSNP). It is assumed that, ultimately, in the process of the project's implementation only about 500 000 tagSNP of the approximately 10 million SNPs present in the genome of every individual will be selected. But this number is more than enough in order to map with SNPs the full human genome and identify new candidate genes linked with various MDs [12].

Thanks to HapMap, which includes the SNPs of not only already known genes, but also those of genes that have yet to be identified, scientists now have at their disposal the powerful universal navigator needed for the in-depth analysis of the genome of each individual for a fast and efficient mapping of genes whose allelic variants predispose one to various MDs (see below).

Francis Collins, the director of the National Institute for the Study of the Human Genome (USA) says: "Already 20 years ago when discussing the program 'Human Genome', I dreamed of a time when the genomic approach will become a tool for the diagnosis, treatment and prevention of severe common diseases that affect patients overcrowding our hospitals and doctors' offices. The success of the HapMap project allows us to make a significant step toward this dream already today."

3. GENES AND HUMAN DISEASES

According to research in the prevalence of diseases in twin pairs and medical genetics data, only about 1.5% of human diseases are directly linked to mutations. These are so-called hereditary diseases. Accuracy in the molecular diagnosis of hereditary diseases is very high and approaches 100%.

All other diseases, including such frequent ones as cardiovascular, cancer, mental, and even infectious diseases, are the result of the combined effect of unfavourable external factors and individual characteristics of the genome, somehow predisposing a particular person to a disease. Hence, the origin of their name: multifactorial (combined or complex) diseases (MD).

Thanks to full human genome sequencing, the development of convenient methods for the mapping of new genes (the total number of genes in humans is estimated at about 22,000), methods for identifying mutations, as well as the problem of diagnosing multiple (especially the most frequent) monogenic human genetic diseases (cystic fibrosis, hemophilia, Duchenne/Baker muscular dystrophy, spinal muscular atrophy, immunodeficiency and others) can be considered as solved [13, 14, 15]. The identification of the genes involved in the genesis of MDs, the so-called "predisposition genes," is much more complicated. According to the current definition, "predisposition genes" are mutant genes (alleles) that are compatible with birth and life but that, under certain unfavorable conditions, can contribute to the development of a MD [14].

Depending on participation in metabolic chains and associations with a MD, susceptibility genes are conditionally divided into several groups, among which are the genes of the detoxification system ("external environment"), the genes of metabolic bypass (genes triggers), cell receptor genes, genes of the inflammation and immune system, and genes associated with a specific MD [Baranov, 2000, Baranov *et al*, 2000]. Unfavorable allelic variants of these genes can cause atherosclerosis, coronary heart disease (CHD), osteoporosis, diabetes, asthma, tumors, etc. The combinations of allelic variants of different genes that provide normal metabolic processes or are involved in the development of a specific multifactorial pathology are called "gene networks"[16]. In each of these networks, the main (central) genes and additional (auxiliary) genes (the so-called genes modifiers) are defined. The concept of genetic networks has further evolved to the study of functional genetic modules (FGM). To this end, a series of studies have compared the MDs and the various genes whose products are involved in the etiology and pathogenesis of these diseases [17, 18, 19]. A network of MDs and the genes common to these diseases was created and named the human disease network (HDN) (Fig. 1) and, vice versa, a network of genes common to various diseases – disease genetic network – DGN (Fig. 2) has also been created. Combining HDN-and DGN-maps has made it possible to create the so-called diseasesome map that reflects the topology of metabolic networks and the genetics of MDs (Fig. 3). As a result of a large-scale study of 1,264 MDs and their associated 1,777 genes [17], the following has been established: 1) each MD is characterized by its own set of genes – a gene network or so-called functional gene modules (FMG), there are central and peripheral genes that can be distinguished by module; 2) most MDs are interconnected through many different genes; 3) 516 MDs show a lot

Circles – diseases; the size of the circle is proportional to the number of the known candidate genes involved in MD.
 Lines – the candidate genes associated with different MDs, the line's thickness is proportional to the number of common genes.
 The number of genes – the scale of genes involved in MDs

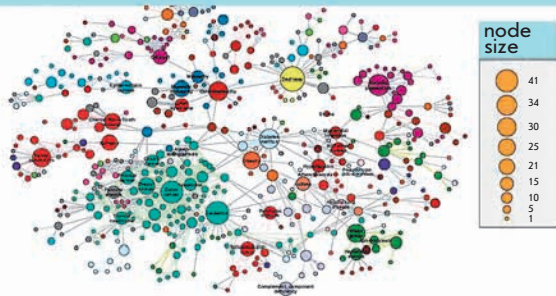


Fig. 1. Gene networks of some common multifactorial (complex) disorders

of genetic linkages; ie, they are associated with many genes (deaf - 41 genes, leukemia - 37, cancer of the colon - 34); 4) the mutation of different genes can lead to the same MD, and mutations (polymorphisms) of one gene may be associated with different MDs; 5) the mutation of the central (essential) genes of FGM are often associated with tumors and causes early death; 6) mutations (polymorphisms) of the peripheral FGM genes play a major role in the phenotypic variability and the development of MDs; 7) the presence of overlapping FGM of MDs shows the pathogenetic proximity of different MDs and argues in favor of syntropy – a combination of pathogenetically related, “family,” of MDs; and 8) genes included in FGM are essentially “syntropy” genes [20] which are functionally similar but not always identical to “predisposition” genes. The coincidence of many MDs in a large number of associated genes was clearly demonstrated when comparing candidate genes associated with various autoimmune diseases (Fig. 4).

About a third of the identified loci are associated with

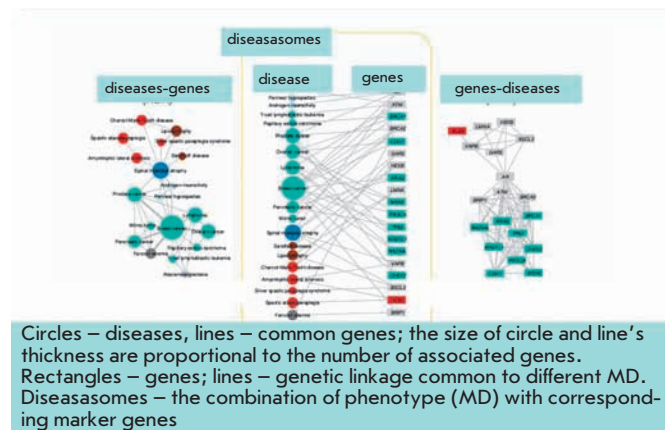
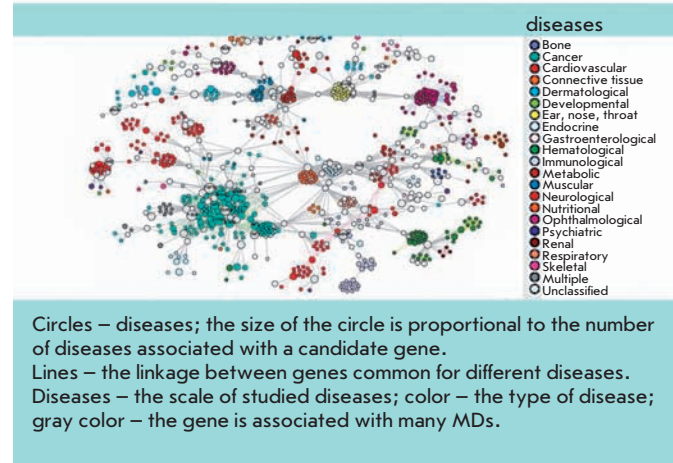


Fig. 3. The topology of metabolic networks and the genetics of complex disorders



Circles – diseases; the size of the circle is proportional to the number of diseases associated with a candidate gene.
 Lines – the linkage between genes common for different diseases.
 Diseases – the scale of studied diseases; color – the type of disease; gray color – the gene is associated with many MDs.

Fig. 2. Functional genetic modules of different MDs

two, three, or more diseases. The presence of more than 5% of common candidate genes associated with the celiac disease, Crohn’s disease, multiple sclerosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, diabetes mellitus type-1, and ulcerative colitis proves the pathogenetic similarity of these autoimmune diseases [21, 22] and allows one to consider them as a single group of syntropy diseases.

Clarification of the FGM of each MD, identification of the central genes and modifier genes in it, the analysis of the association of the alleles of these genes with the disease, development of a set of preventive measures for persons at high risk and, subsequently, for a particular patient on this basis is the main focus of predictive medicine [5].

4. SEARCH STRATEGY FOR “PREDISPOSITION” GENES.

In past decades the search for candidate genes has been performed by means of two approaches: the analysis of association and analysis of linkage.

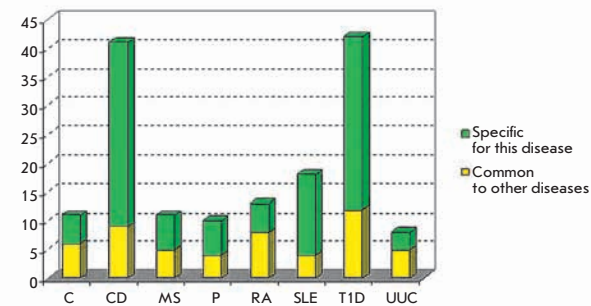


Fig. 4. The number of loci common to autoimmune diseases, and loci that are specific to a particular disease
 Conventional signs: C – celiac disease, CD – Crohn’s disease, MS – multiple sclerosis, P – psoriasis, RA – rheumatoid arthritis, SLE – systemic lupus erythematosus, T1D – type 1 diabetes, UUC – unspecific ulcerative colitis

The basis of the method of analysis of the association is the disequilibrium in the linkage between the mutation and the nearby marker tested [23]. The method involves several steps: 1) selecting the most probable candidate genes on the basis of the particular disease, 2) the selection of functionally important alleles of the corresponding genes, 3) a population analysis of the allele and genotype frequencies of the corresponding genes on the basis of literature data and the Internet, and 4) a comparative analysis of allele frequencies and genotypes of these genes in patients with a clinically confirmed diagnosis and in healthy individuals of the same population, matched by the type of “experience-control.”

Studies are conducted on representative samples of patients and donors (at least 100 people in each group).

Obviously, such a way, being quite lengthy, laborious and costly, at the same time does not guarantee that the identified allelic differences are the main ones in the chain pathogenetic mechanisms involved in the particular disease. It does not rule out the fact that some important genes and polymorphisms of the same, or even more, or another gene network involved in the disease can be missed, and that clinically different forms of the disease under study may have a different pattern of candidate genes.

An alternative strategy of searching for predisposition genes (linkage analysis) is based solely on the positional cloning of locus and does not require a preliminary hypothesis about the pathophysiology of the disease. Initially, the method of genome-wide linkage study (GWLS) was the one available and widely used. The GWLS method is used in families with several sick siblings or in extended pedigrees. It is aimed at detecting in patients blocks of molecular markers that are passed from parents to sick descendants but not to healthy ones. The method allows to localize the gene on the area of 1-10Mb. These extensive areas of chromosomes usually include hundreds of genes; thus, the search for the causative gene in linked locus is a difficult and often impossible task.

Today, a more advanced and widely used method is the method of Genome Wide Association Studies (GWAS). This method was a breakthrough in genetic studies of MDs. It is based on using the HapMap program in conjunction with the technique of high-resolution biochips. As a result of the HapMap Project, the distribution of thousands of polymorphic sites - single-nucleotide substitutions (SNP) was revealed in the human genome and haplotype maps were created - maps of stable combinations of SNP variation within a single-stranded (haploid) DNA sequence [13]. Another important technical achievement was the hybridization of high-density DNA biochips that enabled simultaneous genotyping of thousands of SNP sites in one DNA sample. With knowledge of the exact position of each SNP on the physical map of the haploid genome, it has become not only possible to identify a candidate gene, but also to identify all SNPs associated with MDs [24, 25].

The basis of the GWAS method is in the scanning of hundreds of thousands of markers located on all human chromosomes. Thanks to the haplotypes maps obtained in the HapMap project, modern chip design includes the maximum number of key SNP (tag SNPs) and permits to estimate the frequency of both individual markers and haplotypes for the entire length of the DNA molecule. For example, the wide-

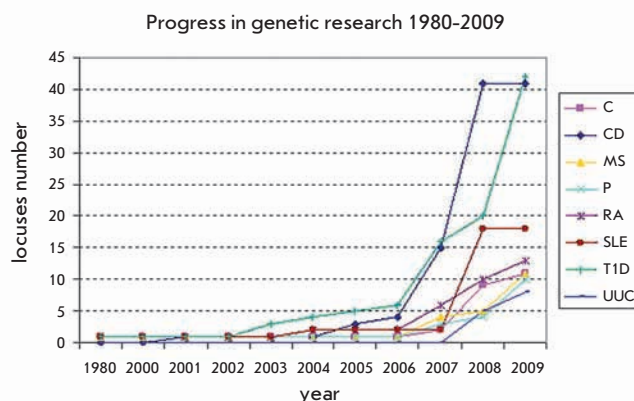


Fig. 5. Progress in searching for genes of MD of immune system. C – celiac disease, CD – Crohn’s disease, MS – multiple sclerosis, P – psoriasis, RA – rheumatoid arthritis, SLE – systemic lupus erythematosus, T1D – type 1 diabetes, UUC – unspecific ulcerative colitis

ly used chip firm Illumina (www.illumina.com) comprising 310,000 tag SNPs (Illumina Hap310K) enables to estimate the frequency of 81% of frequent polymorphisms in the European population. The next development by the same company comprises 550,000 SNP (Illumina Hap550K) and covers more than 90% of frequent polymorphisms [24].

Genome-wide associations screening is conducted on a large number of patients and controls (more than 1,500 – 2,000 people), which ensures highly reliable ($p < 0.000005$) results and includes several stages. At first, hundreds of associations are identified, most of which appear to be false-positive after hundreds of thousands of independent tests. In the next step, associations in an independent cohort of patients and controls are analysed by the same method. Only the results confirmed in a replication cohort are considered to be reliably positive. At present, the scanning of about 300 different associations of MDs has been carried out by using GWAS. The results of these studies are summarized on the website of the National Institute of Health (USA) - [http://www.genome.gov/GWastudies/index.cfm? # 1](http://www.genome.gov/GWastudies/index.cfm?#1). The data include the results of GWAS obtained with a reliability of $p < 1 \cdot 10^{-5}$ and containing not less than 100,000 SNP. They are regularly updated following each publication of new data [25].

GWAS in complex diseases is very popular and has been successfully used in the past few years. Data obtained for genes associated with certain diseases of the immune system are presented in Fig. 5.

Thus, the GWAS method is surely becoming the principal method for the search for candidate genes of all MDs. Unfortunately, this revolutionary technology, to my knowledge, so far remains hardly available in Russia. Given the significant populational differences in genetic polymorphism, the implementation of GWAS technology for the identification of candidate genes of MDs in this country is of utmost necessity.

5. GENETIC PASS

Many diagnostic centers in Russia currently widely use molecular methods for the diagnosis of genetic diseases and de-

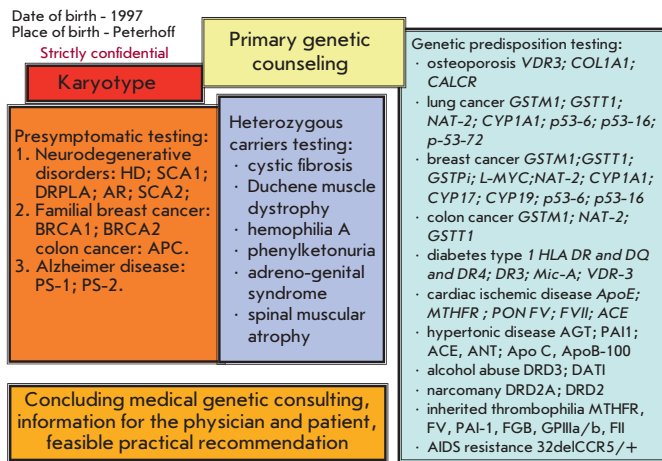


Fig. 6. A version of the “genetic passport” [5,6]

tection of heterozygous carriage of pathological mutations in families at high risk for the preliminary diagnosis of diseases with a late manifestation and for personal identification (genomic fingerprinting). Gradually, genetic testing for predictive medicine is gaining in strength. Obviously, data on both the genome of individuals and entire families and, gradually, individual and family DNA databases are being created. Such an individual DNA database is a “genetic passport (pass).”

Thus, a genetic passport is an individual DNA database reflecting the unique genetic characteristics of an individual, his or her predisposition to some hereditary, multifactorial and other diseases [4, 5, 6, 37].

The information contained in this really unique document should help to avoid the difficult situations in one’s life associated with ignorance of the individual characteristics of one’s genome; ie, the specific characteristics of one’s heredity. These data allow us to more fully fulfill our genetic ability and are of undeniable value for progeny.

The widespread introduction of molecular diagnostics in modern medicine has already made real the idea of a genetic passport. It is already a de facto reality, and the number of GT is increasing rapidly. However, launching the establishment – and especially the practical use – of genetic passports can only be allowed with fairly stringent regulatory requirements. The latter include the following:

1. a well studied genetic network for each MD;
2. reliable clinical and population data confirming the contribution of relevant marker genes in the pathogenesis of MD;
3. representative data for the population of a particular region or the ethnic group demonstrating the association of the tested marker genes with the MD;
4. a well-balanced (adequate) interpretation of GT results in hereditary predisposition;
5. cogent recommendations based on the results of GT (according to the genetic passport);
6. monitoring of the remote results of the patient’s condition after GT and the prescription of recommendations by clinical geneticists; and

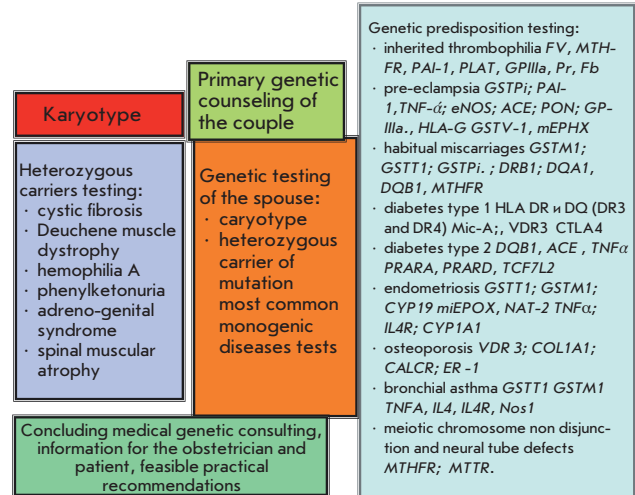


Fig. 7. A version of the “genetic card of reproductive health” [26]

7. privacy, accessibility, and legal protection for patients.

The full version of a genetic map should include not only a study of predisposition genes, but also asymptomatic carrying of mutations of genes of the most common hereditary diseases (hemophilia, cystic fibrosis, phenylketonuria, etc.). Currently, the existing molecular diagnostic laboratories in Russia, including those located in St. Petersburg, allow for a fairly complete set of essential genetic tests. One of the first variants of a genetic passport was proposed by us in 1997 (Fig. 6)

At present, practical applications are found for only some components of a genetic passport (heterozygous carrier testing, genomic fingerprinting, and karyotyping). Rarely, and only in high-risk families, is hereditary predisposition to bronchial asthma, diabetes, or osteoporosis tested.

The genetic map of reproductive health appears to be more advanced as regards its clinical implementation. Such a map was obtained as a result of many years of comprehensive studies of the reproductive function of women conducted at the Ott’s Institute of Obstetrics and Gynecology RAMS (St. Petersburg) [26]. The map is recommended for use in family planning centers, as well as in antenatal and other centers of the institute. It is widely used in outpatient counseling by clinical geneticists and obstetricians in our institute. In addition to karyotype analysis and testing for a carrier of mutations of severe hereditary diseases for spouses who are planning to have a child, the screening of women for an array of genetic diseases complicating pregnancy, fetal development, childbirth and the postpartum period (pre-eclampsia, habitual miscarriage, varicosity, feto-placental failure) has an important prognostic significance (Fig. 7). For gynecologists and endocrinologists, testing of genetic susceptibility to endometriosis, adenomyosis, and postmenopausal osteoporosis is of great interest.

Particular attention is focused on the testing of hereditary forms of thrombophilia, for diagnostics of which a special microbiochip called “Fibrochip” was developed [27]. The clini-

cal trials of the genetic map of reproductive health (GMRH) conducted at the Ott's Institute of Obstetrics and Gynecology, RAMS, focused primarily on individual disorders, such as endometriosis (prognosis and choice of optimal treatment tactics), inherited thrombophilia factors, miscarriage and placental insufficiency, and gestosis (prediction and prevention). The accumulated information on the genetic markers of obstetric pathology that results from prospective testing is a good reason for a wider implementation of such a card (GMRH) in clinical practice.

Based on WHO recommendations, genetic testing should be carried out with the voluntary, informed consent of the patient; ie, after he or she reaches adulthood. Formally, this means that important genetic information could become available relatively late, when its benefits to the subject and his close relatives have largely been lost. However, taking into account the importance of these data for children's health, the harmonious development of their personality, rational nutrition, effective education, athletic training, optimal professional guidance, and the opportunity to prevent the development of several diseases with a late manifestation, introducing genetic passports at an early age would seem to make sense today. A possible genetic map of a child based on the results of studies of his gene associations is shown in Figure 8.

It cannot be excluded that as the ethical and social problems associated with studies of the human genome are solved genetic testing in an earlier than currently recommended age will become common. In any case, in families with a high risk of diabetes type-1, asthma, sudden death due to cardiac conduction and rhythm disorders, metabolic syndrome and obesity, as well as several other nosology, preventive genetic testing at an early age seems reasonable. [37]. Obviously, the testing should only be conducted with parental consent, with recommendations from a pediatrician, and after consultation of the family by a clinical geneticist competent in predictive medicine.

Likewise, information about gene markers, the testing of allelic variants of which allows to evaluate the suitability of a

teenager to a particular sport, accumulates rapidly. Currently, information is available about nearly 150 different genes that control human physical development. That is important for proper fitness and for the selection of potentially promising athletes. The obtained results allowed to start building an individual version of the genetic map of the athlete, which includes the testing of certain genes that determine human physical characteristics.

Despite certain limitations of legal, moral and ethical character, a lack of information on the gene networks of various metabolic processes and multifactorial diseases, the lack of reliable statistical information and the imperfection in the clinical interpretation of the results of genetic testing, the creation of a functioning genetic passport of some type for citizens should be welcomed. Such a medical document can be of substantial assistance during the examination of the health status of an individual, as well as in assessing the potential risk of developing a particular disease by members of a family at high risk of developing an MD [37].

Thus, despite the obvious imperfections of modern predictive medicine, genetic screening of families at high risk for some serious MDs, as well as professional athletes and people exposed to extreme work conditions and individuals simply interested in information about their genome, seems quite real. Obviously, a genetic map of reproductive health is also of great practical significance.

6. THE MAIN PROBLEMS RELATED TO GENETIC TESTING OF MD

Testing for gene associations, i.e. the search for candidate genes coupled with a variety of MDs, is already widespread throughout the world, including at leading laboratories and genetic centers in Russia. According to international reports, thousands of polymorphic sites are tested daily to determine associations with diseases [28]. About 100 major candidate genes, each of which contains several polymorphic sites that affect the function of the gene and its products, have already been identified or are under study for common MDs.

The GWAS method in searching for marker genes of MDs has become more widely used in recent years [29]. Hundreds of candidate genes and anonymous polymorphisms coupled with the most common MDs have already been identified using this method [29]. Comparison of the characteristics of the distribution of allelic variants in patients and healthy people has already led to the definition of the "genomic profile" of a MD that corresponds to the SNP-alleles distribution in the genome typical of a particular disease. Thus, identification of marker genes in itself appears not to be important. The results are written using the numbers of tagSNP that have shown reliable linkage with the MD. For example, 22 variants of the genomic profiles of risk of prostate cancer were identified in a large prospective study of the Icelandic population. The highest linkage (OR = 1.23; P = 6.7 · 10⁻¹²) was found for SNP rs11228565 of locus 11q13. The presence of four variants of the risk SNP rs10934853 (3q21.3), rs16902094, and rs445114 (8q24.21) rs8102476 (19q13.2) increased the probability of prostate cancer in carriers by 2.5 times compared with population SNP variants [30]. Commercial genetic tests, including "individual genomic profiles," have emerged and are being actively promoted by various diagnostic centers in America (Celera, Myriad genetics, Decode, Navigenics,

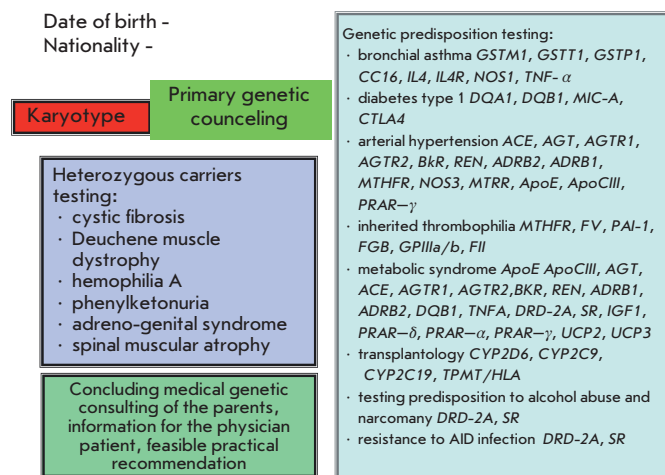


Fig. 8. A version of the "genetic card of a child's health" [37]

23andme) and Western Europe (Sciona, Gendia) [31]. Yet the associations of numerous marker genes and polymorphic loci with a particular disease remain unproven, and more than 5% of already identified associations turn out to be random [28]. In the opinion of many experts in molecular medicine and medical genetics, a quite alarming situation is developing as commercialization and business become the primary motive, instead of science.

The main problems that have appeared in the implementation of GT results of hereditary predisposition to clinical practice are as follows:

1. the completeness of identification of all candidate genes of the gene regulatory network (FGM) of MD;
2. proofs of the reliability of their association with MDs;
3. medical evaluation of GT results; and
4. the clinical significance of predictive GT.

6.1. THE COMPLETENESS OF THE IDENTIFICATION OF CANDIDATE GENES FGM FOR A PARTICULAR MD

The method of analysis of association (see Section 3) at best permits to reveal the most obvious, in terms of the pathogenesis of the disease, candidates genes and conduct a comparative analysis of allele frequencies in cohorts of patients and population controls.

The method of genome-wide linkage analysis allows for a more detailed identification of loci coupled with a specific MD, but the rather extended size of these loci often highly complicates the exact identification of important candidate genes in it.

Today, hope in solving this problem rests on the GWAS method. Using high-density biochips (over 500,000 SNP) allows to overlap the full genome with high reliability. Thus, information on all the SNP linked with MDs can be obtained. However, many such associations appear to be random in fact and can be discriminated only by repeated GWAS analysis conducted on other cohorts of patients with the same MD and other groups of population control. In addition, the association in one or another DNA site does not always coincide with the presence of the corresponding candidate gene in it and does not necessarily mean the identification of the mutations responsible for the predisposition to the MD. In most cases, the identified SNP sites are located in intergenic DNA and, in fact, can be considered only as a molecular marker linked with one or several neighboring genes. However, at the moment the GWAS method, when complemented by the sequencing of linked loci and the analysis of the expression of its constituent genes, is considered as the most effective method for identifying candidate genes comprising FGM of a particular MD. However, the identification of new marker genes associated with the MD is often paired with the revision of the value of the initial individual risk. That was clearly illustrated in the genomic scanning for predisposition to type 2 diabetes [32].

The possibility of testing for hereditary predisposition not only on the basis of the allelic variants of predisposition genes, but also for the genomic profile on the basis of tagSNP [30] significantly bolsters the case for predictive medicine.

6.2. THE RELIABILITY OF THE ASSOCIATION OF CANDIDATE GENES WITH A SPECIFIC MD

Currently, there are already about 1,024 clinical genetic tests of MDs and there are more than 300 genetic tests that are in

the pre-clinical and clinical trials stages [33]. For many MDs, some “major” genes have already been identified. Their involvement in a particular pathology has been confirmed by research on representative groups of patients in many laboratories. Such MDs include, for example, Alzheimer’s disease (AROE4), type-2 diabetes (PPARG, TCF7L2, KCNJ11), senile macular degeneration of the retina (CFH), systemic lupus erythematosus (JRF5), prostate cancer (region JF1H), type-1 diabetes (IL2RA, CD25, PTPN22), autoimmune thyroiditis (CTLA4), hirschsprung disease (RET), Crohn’s disease (NOD2, CARD15), and rheumatoid arthritis (PTPN22) [8, 34]. Nevertheless, the attitude toward GT remains quite skeptical, largely because of the absence of definite statistical evidence of the reliability of GT results.

One important reason for such a discrepancy is the relatively small sample groups of sick and healthy patients. At best, they are limited to hundreds of subjects, whereas a comparative genetic analysis of several thousands of healthy and sick patients is required to obtain statistically reliable data [28].

Populational differences in allelic frequencies for the same candidate gene could be another reason for the variability of gene associations. Non-randomly identified allelic differences demonstrating the significance of a particular association are often averaged and become statistically insignificant when combined with data from different works carried out on groups of patients, and controls of other populations.

The need for large sample groups is dictated by the relatively low rate of relative risk, OR, which shows how much more common is the studied disease in individuals to study the disease with a specific allele or set of different alleles of the marker genes, compared with individuals with other alleles of the same candidate genes [35]. Typically, for known gene markers associated with an MD, the value of OR does not exceed 1.5 and is usually within 1.16-1.2. Thus, the combination of several unfavorable alleles of candidate PTPN22 in diabetes and systemic lupus erythematosus and the NOD2 gene in Crohn’s disease increases the relative risk of the disease (OR) 2-3 times. For most other associations, OR ranges from 1.1 to 1.5. [22]. Therefore, the real contribution of each marker gene in the development of the MD is relatively small. Therefore, for definite proof of association extensive genetic studies are required. It is believed that to prove a 90% probability of any association between a gene and disease with an OR equal to 1.25, the study requires 5,000 patients and a control group of not less than 5,000 people. This is assumed to increase the level of significance ($p < 0.00005$ instead of the usual $p < 0.05$) 10,000 times [28]. The appearance of the GWAS method has radically changed the situation as relates to the problem of predictive validity of GT. Thus, in assessing the risk of myocardial infarction by GT of 85 SNP in corresponding candidate genes, the reliability of associations has increased to $p < 0.0000001$ [36].

Another important factor that proves the strength of identified association is its replicativeness; ie, the reproducibility of GT in the works of other researchers.

The reasons for the variability of genetic polymorphism are very diverse.

- Genetic stratification of the studied population (the presence of subpopulations with different initial frequencies of the analyzed alleles in it) [28];

- Allele frequencies in different populations may vary, and their contribution to the pathogenesis of a specific MD could be different;

- Pathogenetic differences in a MD could be due to the peculiarities of the action of various external factors in different geographical conditions;

- The inaccuracy of the clinical diagnosis could lead to errors in the formation of clinical sample groups;

- Alleles associated with the same MD in different populations could be different.

There are several other factors that substantially hinder the correct evaluation of the observed genotype-phenotype association, even if it is quite statistically significant:

- The identified association cannot relate to an identified gene or SNP-marker but to a gene or locus (alleles) that is closely linked with a yet unknown locus or allele, whose product is involved in the pathogenesis of MD;

- The identified association could, in fact, be not related to the candidate gene in itself but to another gene product which functionally compensates for the effect of the mutant gene (epistatic relationship of genes);

- Not only gene--gene interaction remains very poorly understood, but also the interaction of candidate genes with environmental factors;

- The pathogenesis of any MD can often be the result of disorder in the function of not just one, but different gene networks;

- Along with the typical polygenic forms of a MD, there are also common forms of monogenic (osteoporosis, Parkinson's disease, Alzheimer's disease, various cancers) MDs[35].

Obviously, all these factors significantly complicate the correct identification of marker genes and the assessment of their objective contribution to the pathogenesis of the disease.

Recognizing all these limitations and the very real difficulties in the identification of the marker genes of MDs, it is important to pay attention to the following circumstances:

- All MD genotype--phenotype relationships always have a **stochastic nature** and are not strictly determined; ie, accuracy in the DNA diagnosis of MDs, in contrast to monogenic diseases, never approaches 100%.

- Groups of sick and healthy patients will include thousands of people in an analysis using the GWAS method, which ensures a high reliability of the results.

- The overwhelming majority of marker genes and loci identified with GWAS usually includes associations previously established by other methods.

These statements need further clarification and verification, but they already provide a basis for more extensive pre-clinical and clinical trials of the already identified candidate genes of common MD.

At the current stage in development of predictive medicine in Russia, it seems rational to compare the known candidate genes of MDs identified in the works of local researchers with the corresponding panels of candidate genes identified using GWAS (1), to test the allele frequencies of the main new candidate genes on already existing samples of patient-healthy DNA-banks (2), and to create of a new panel of genes of MDs based on the data obtained (3).

6.3. EVALUATION OF THE RESULTS OF PREDICTIVE GT

The evaluation of genetic testing results should be conducted considering already existing knowledge about the gene networks of specific MDs, the population, gender, and age characteristics of the frequency of polymorphic alleles of the studied genes [37]. Considering the probabilistic nature of forecasts based on the genetic testing of hereditary predisposition to MDs, some assistance in assessing the risk of a hereditary predisposition can already provide a fairly simple method of score evaluation, which is used in some western countries (Harvard School of Public Health) and is already used in some Russian centers conducting genetic testing [38, 39, 40]. The essence of the method is as follows: each genotype variant is estimated on conditional points depending on whether the alleles identified are protective or predispose to the development of pathology. To this end, alleles with altered functional activity of the gene are assigned 1 point, while normal (frequent), wild-type alleles are assigned 0 point. Then, the points corresponding to genotype for every tested candidate gene are recorded, summed up, and divided by the number of tested genes. So, the risk of developing the disease could be conditionally evaluated as medium, low, or high. In the presence of pathogenetically complex MDs, including several different metabolic chains, the calculation is conducted for each gene network separately and the estimates obtained are added up. Some variants of score evaluations, in addition to the calculation of points corresponding to genotypes, also include conditional points of various exogenous factors (harmful effects, habit, medication intake, etc.), anthropometric values, as well as physical activity, sex, weight [26].

Information on the basis of GT results is prepared by a clinical geneticist, together with the specialist who carried out the molecular analysis, and is passed along to the physician and patient. Comparison of this conclusion with the results of clinical, laboratory, and instrumental investigations permits to assess the risk of developing a MD more objectively and to offer the most effective program of prevention and treatment.

The answer may be more objective if it is possible to compare the results with data on the genetic testing of a close relative who has a given MD. But in any case the answer would be essentially probabilistic in nature. A detailed method of scoring the results of GT is given in our user guide devoted to the genetic map of reproductive health [26].

Special computer programs could significantly assist in the correct interpretation of GT results. Such programs are available and are used to compare the genomic profiles of patients with MDs, the control group of patients, which allow us to estimate the risk of hereditary predisposition of a particular person to a MD [29]. A map is created that facilitates the understanding of the results of GT by the medical staff. It helps to find genetic variations corresponding to certain diseases and to monitor their transmission by inheritance. It is believed that such a map will help to reduce the cost of searching for genes of predisposition to a particular MD, as well as to make real development of individual treatment possible [41]. There is no doubt that the establishment of such a computer program to evaluate GT results in Russia would also greatly accelerate the introduction of predictive medicine in clinical practice.

According to foreign reports, GT for assessing genetic susceptibility to cardiovascular diseases, venous thrombosis, hyperlipidemia, and atherosclerosis has already been developed and is widely used [41]. It is important to emphasize that these studies have yet to acquire the status of tests recommended for clinical use, but they are already being used by clinicians. More than 1,000 predictive GT, many of which are at the pre-clinical and clinical trials stages are in the same state. The objects of GT are such genes as APOE4, homozygosity for which increases the risk of Alzheimer's disease 14-fold; gene Filaggrin, mutation in which R501X or 2282del14 increases the risk of atopic eczema and severe asthma 4-fold; the unfavourable alleles of CDKN22a2b (64% increase of risk of infarction by 64%); and 4a/4b alleles of IN, the presence of which doubles the risk of type 2 diabetes. However, the clinical significance of these genetic tests remains unclear; their usefulness to physicians and patients needs more rigorous proof [36].

The first predictive genetic test to calculate the individual dose of the anticoagulant warfarin successfully passed certification and received the formal approval of the U.S. Food and Drug Administration on August 16, 2007. The test includes testing of the genes CYP2C9 and VKORC1, taking into account the age, sex and weight of the patient.

In 2008, the EuroGentest Commission in Europe developed guidelines for GT standardization and prepared the necessary documentation for certification of those genetic tests that, on the basis of the results of clinical trials, can be referred to the category of GTB recommended for clinical use [Nippert *et al.*, 2008].

6.4. RECOMMENDATIONS ON THE RESULTS OF GENETIC TESTING

GT may offer real benefits only if it is accompanied by full consultation with a qualified clinical geneticist, with the provision of relevant recommendations to the physician and patient. GT may have practical value if the following conditions are fulfilled: (1) the results of GT are based on the analysis of genes whose association with the corresponding disease is shown in the population of the particular region; (2) the subject is a family member at high risk, where there is already a patient with this pathology; and (3) the GT data are subjected to adequate statistical analysis. Effective use of such information is largely determined by the level of genetic knowledge of physicians, their ability to apply the data to the diagnosis, prevention and treatment of diseases, as well as the willingness of the patient to follow the recommendations of physicians based on the results of genetic testing. [39]. But even under these conditions, GT results concerning hereditary predisposition should be interpreted very cautiously. When possible, GT should be complemented with an appropriate biochemical analysis allowing to evaluate the functional activity of the studied genes. It should be required that more objective information be obtained by testing the genes controlling only a single metabolic process; ie, the genes belonging to the same gene network. Thus, already today on the basis of genetic testing the functional conditions of the following systems can be assessed objectively: the detoxification system, blood coagulation, lipid or carbohydrate metabolism, the renin-angiotensin system,

and others. Assessing the outcome and prognosis of hereditary predisposition to MDs caused by damage to several gene networks is much more difficult.

The main difficulties in the widespread implementation of predictive medicine in clinical practice are related to the lack of objective data to prove the usefulness of preliminary (no symptom) testing of hereditary predisposition to MDs for the patient.

According to the Gene Dossier recently developed by the United Kingdom Genetic Testing Network (www.ukgtn.nhs.uk), certification of each new GT should include (1) information on the analytical precision of the used molecular genetic methods; (2) the clinical reliability of the GT, i.e. its ability to diagnose or predict the presence or absence of a certain phenotype; (3) the clinical usefulness of GT, its compliance with ethical, legal and social norms, which means that GT should be aimed at specific populations and serve to solve a particular problem (www.labtestonline.org.ru).

The British fund Wellcome Trust, which initially funded the Human Genome Project program, in 2008 began supporting a project aimed at improving and strengthening the evidence base of genetic testing, as well as the development of a handbook for the coordination and integration of GT into clinical practice with sufficient explanations of their usefulness to physicians and patients. At the same time, the estimation of the clinical usefulness of GT is equivalent to phase III of clinical trials, but it remains unclear whether GT should be paid for by governments or the firms working on and advertising GT [43].

The ultimate goal of the project is to move predictive medicine from the field of research of genetic polymorphism features and identification of genes - markers of MDs to the level of evidence-based medicine.

The results of every GT should include not only information about the features of the allelic variants of the genes of a metabolic chain, but also recommendations for the patient and physician [44]. A "Genetic" reorientation of the health care industry is already happening in the developed countries of Western Europe and America. In the near future, it will reach Russia. A significant improvement in genetic knowledge, especially in the field of predictive medicine, for doctors in all specialties is very important for the effective implementation of the achievements of medical genetics and genomics in the Russian health care system.

CONCLUSION

Thanks to impressive advances in genomics, the emergence of new, highly effective methods of molecular analysis, the search for marker genes associated with MDs has rapidly developed. As a result, (1) thousands of new marker genes, the allelic variants of which predispose to the development of pathological processes, have been identified; (2) the genetic panels of the most frequent chronic diseases has been established; and (3) marker genes defining the severity of the disease and predisposition to some complications are being identified.

The unprecedented scale of the genotyping of representatives of different races, nationalities, and ethnic groups has required hard joint work from clinicians and molecular biologists. Their work has led to numerous DNA databanks con-

taining information on all known mutations and DNA alterations associated with chronic diseases (4).

Extensive information on the genotyping of frequent chronic diseases among inhabitants of Russia has also been collected at many Russian scientific centers such as the Institute of Medical Genetics of RAMS (Tomsk); Medical Genetics Research Center (Moscow); Institute of Biochemistry and Genetics, Academy of Sciences (Ufa); N.I. Vavilov Institute of General Genetics, RAS (Moscow); Institute of Molecular Genetics, Academy of Sciences (Moscow); and the Institute of Molecular Medicine, RAMS, etc.). Only in our laboratory at the Ott's Institute of Obstetrics and Gynecology, RAMS (St. Petersburg), has the the frequency of allelic variants of 80 marker genes in 5,000 patients with different parts of MDs and the same number of people from the control group been studied.

As an individual DNA database, the genetic passport (GP) already exists and is gradually becoming more complex with the identification of new genetic markers and the increase in the number of gene networks and panels of predisposition genes determining hereditary susceptibility to MDs. If at its early stages GP was a fairly simple map including test results for approximately 100 marker genes corresponding to gene panels of 15-20 frequent chronic MDs, since the emergence of the GWAS technology the number of candidate genes has grown rapidly. Its characteristic genetic profile corresponding to the distribution in the genome of more than 30-500, 00 SNP in the SNP map could be determined for each disease with the help of this technology. Comparison of the genetic profiles of sick and healthy patients with that of a particular subject provides a high degree of reliability in testing inherited predisposition to the relevant disease.

It is obvious that GT results have to be treated with great caution. The clinical usefulness of such testing, even with GWAS technology, still needs to be proved. Of particular concern is the lack of information on how and what environmental factors provoke the onset of MDs in a particular person. To this end, the scientific community has already been tasked with quantifying the genetic and exogenous risk factors and their combinations in the pathogenesis of MDs [45]. Despite the serious complexity, the implementation of predictive medicine in clinical practice is scientifically founded and strategically unavoidable.

In conclusion, let me note that the emergence of new, highly efficient methods of DNA sequencing has made it possible to sequence the full genome of an individual. Massive parallel sequencing is particularly promising in this respect [46]. Recently, it was reported that each American can now get his or her whole genome for \$ 50,000; ie, 20 times less than in 2007. It should be noted, however, that the completely sequenced genome of an individual is unlikely to replace the genetic passport in the near future. A genetic passport is much more convenient and practical in everyday use for a clinical geneticist and for a physician using the GT data. The

full genome sequence will certainly be important for a more in-depth analysis of the unique features of the individual genome; ie, it can play the role of a universal genetic handbook for everyone, while GP will provide information on the status of the predisposition genes of several common MDs. It is assumed that within the next 2-3 years, everyone will be able to obtain a complete set of his genome for \$1,000, and that the cost of a genetic passport with the commentary of a specialist will cost about \$300. Hence, in addition to conventional major medical tests and an anthropometric survey, the personal medical card of each person will also include the results of GT, the number of which will invariably increase. However, according to the opinion of academician V.P. Pusyrev, GT will not replace but will only supplement the results of other laboratory studies [20].

Anticipating such further development, a lot of work on the integration of genomics research in the national (public) health policy and practice has already been carried out in the countries of Western Europe and North America. Knowledge of the genome sequence has to be integrated into the health doctrine of each country, with emphasis on predictive medicine. To do this, however, there is still the need to figure out what clinical value "high genetic sensitivity" represents and how to quantify genetic and exogenous risks in accordance with the principles of evidence-based medicine [45].

The advances of genomics in society and in medicine can be accelerated, but it cannot be halted. This fully applies to predictive medicine. The introduction of the technology of genome-wide screening for the identification of candidate genes of MDs, the comparison of the individual profiles of the allelic variants of candidate genes of a subject with those of sick patients with a particular MD and of obviously healthy people supported by the outcomes of prospective genetic testing will open a wide avenue to a new and promising era of predictive medicine for humanity. The main task of modern genomics is to evaluate the significance of the results of GT to determine the conditions of their implementation in medical practice. Possible solutions to this problem in Russia include the following: (1) a comparison of the obtained GT results of MDs of local populations with international data of their genome-wide screening, (2) the establishment of representative (at least 1,000 samples) DNA banks for each MD, (3) testing of new candidate genes on already existing domestic DNA collections, and (4) organizing centers for the implementation of genome-wide association screening GWAS. ●

The authors express their deep appreciation to the researcher of the Department of Biomedical Genetics, University Medical Center of Utrecht (Netherlands) A.P. Zhernakova for her valuable advice on the manuscript and the possibility to use her data concerning GWAS of a number of autoimmune diseases.

REFERENCES

1. Collins F.S., McKusick V.A. // J. Am. Med. Ass. 2001.V. 285. № 5. P. 540-544.
2. Collins F.S. Shattuck. // New Engl. J. Med. 1999. V. 341. №1. P. 28-37.
3. Bochkov N.P. // Vestnik Russian Academy of Med.Sci 2001.- № 4 P. 4-13.

4. Baranov V.S., Kiselev L.L. Human Genome and molecular medicine In : Genomics Input into Medicine // V.I.Ivanov, L.L.Kiselev (eds) IKC Akademkniga ,2005, P. 4-13
5. Baranov V.S., Baranova E.V., Ivaschenko T.E., Aseev M.V. // Human Genome and Predisposition Genes. Introduction into Predictive Medicine. Intermedica, SPb, 2000, 63 p.

REVIEWS

6. Baranov V.S., Aseev M.V., Baranova E.V. // Priroda 1999,N3, P. 17 -27
7. Puzyrev V.P. // Med.Genetics 2003, V,2,N12. P.498-508
8. Puzyrev V.P., Fraydin M.B., Kucher A. N. Genetic Populations Diversity and Human Diseases. Tomsk Pechatnaya Manufactura 2007,320p
9. Puzyrev V.P., Stepanov V.A. Pathological Anatomy of Human Genome. Novosibirsk,Nauka 1997, 223 p.
10. Fogel F., Motulsky A. Human Genetics Moscow, "Mir" 1989, v.1. 308 p
11. The Intern.HapMap Project // Nature 2003. V. 426. P. 789-796.
12. Gabriel S.B., Schaffner S.F., Nguyen H. et al. // Science. 2002. V. 296. № 5576. P. 2225-2229.
13. Gorbunova V.N., Baranov V.S. Introduction into molecular diagnostics and gene therapy of inherited disorders / Spb, "Spetsialnaya Literatura" 1997, 287p
14. Baranov V.S. Vestnik Russian Academy of Med.Sci 2000.- № 10.- P. 27-37.
15. Gorbunova V.N. Molecular background of medical genetics. SPb. "Intermedica", 1999, 213 p.
16. Kolchanov N.A., Podkolodnaya O.A., Ignatyeva E.V. et al // Vestnik of Vavilov's Society for Genetisists & Selectionists 2005, V.9, N2, P.179-199
17. Kwang-II Goh, Cusick M.E., Valle D. et al. // Proc. Natl. Acad. Sci. USA 2007. V. 104. № 21. P. 8685-8690.
18. Lee D.S., Park J., Kay K.A. et al. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 29. P. 9880-9885.
19. Hidalgo C.A., Blumm N., Barabasi A-L., Christakis N. A. // A dynamic network approach to the study of human phenotype. PLoS Computational Biology www.ploscombiol.org 2009. V. 5. is41-11. e1000353.
20. Puzyrev V.P. //Med.Genetics 2008, V.8, N9, P.3-9
21. Zhernakova A., van Diemen C.C. and Wijmenga C. // Nat Rev Genet. 2009. 10, 43-55.
22. Zhernakova A. and Wijmenga C. HLA and Non-HLA Genes in Celiac Disease. In Fasano, A. (ed.) // Frontiers in Celiac Disease. Karger, Basel. 2008. V. 12. P. 32-45.
23. Aulchenko Y.S., Aksenovitch T. I.// Vestnik of Vavilov's Society of Genetisists & Selectionists
24. Xavier R.J., Rioux J. D. // Nat. Rev. Immunol. 2008. V. 8. P. 631-43.
25. Hindorf L.A., Sethupathy P., Junkins H.A., et al. // Proc. Natl. Acad. Sci. U S A. 2009. V. 106. P. 9362-7.
26. Baranov V.S., Aylamazian E.K. (eds) Evaluation of inherited predisposition to some common ostetritian disorders . Genetic Chart of Reproductive Health . Practicalical Recommendations . SPb, 67 p/
27. Glotov A.S., Nasedkina T.V., Ivaschenko T. E. // Molecular Biology 2005,V.39,N.3, P.403-412
28. Colhoun H.M., McKeigue, Smith G. D. // Lancet. 2003. V. 361. P. 865-871.
29. Seng K.Ch., Seng Ch. K. // Europ. J. Hum. Genet. 2008. V. 16. P. 554-564.
30. Gundmundsson J., Sulem P., Gudbjartsson D.F., Blondal Th., Gylfason A. et al. // Nature Genetics. 2009. V. 41. P. 1122-1126.
31. Bouchie A.A. // The Scientist. 2006. V. 20. № 4. P. 78.
32. Mihaescu R., van Hock M., Sindjbrand E. M., Uitterlinden A. G. et al. // Genet. Med. 2009. V. 11. № 8. P. 588-594.
33. Berry T.A., Wiliams J.K. // Genetic Testing. 2007. V. 11. № 2. P. 111-117.
34. Diego C.D., Alcantara M., Valle J. et al. // Genet. Testing. 2006. V. 10. № 3. P. 178-185.
35. Ginter E.K. Medical Genetics . Moscow, "Medicina" 2003, 447 p.
36. Meltzer D., Hogarth St., Liddel K., Ling T., Sanderson S., Zimmern R.L. // Brit. Med. J. 2008. V. 336. P. 590-593.
37. Baranov V.S., (ed) Genetic Pass – Background of Personal & Predictive Medicine" SPb., "N-L" 2009. 527 p.
38. Minaycheva L.I., Stepanov V.A., Puzyrev V.P., Nazarenko L.P., Spiridonova M.G., Makeeva O.A. Problems in implications of genomic medicine achievements in clinical practice / In Molecular-biological technologies in medical practice. Novosibirsk, Alfa-Vista, 2004 P.115 -120 //
39. Pogrebenkova V.V., Makeeva O.A. Analysis of genetic testing of multifactorial disorders predisposition application by population and physicians of Tomsk city In Human Genetics in Pathology. Tomsk, "Pechatnaya Literature, 2007, P.103-105
40. Glotov A.S., Ivaschenko T.E., Obrazova G.I., Nasedkina T.V., Baranov V.S.// Molecular Biology 2007,V.41, N1, P.18-25
41. Humphries S.E., Ridker P.M., Talmud P.J. // ArteriosclerThromb.Vasc.Biol. 2004.
42. Nippert I., Krisoffersson U., Schmidtke J. et al. // Europ. J. Human Genetics. 2008. V.16. Suppl. 2. P. 421.
43. Furness P., Zimmern R.L., Wright C., Adams M. The evaluation of diagnostic laboratory tests and complex biomarkers. Summary of a Diagnostic Summit 14-15 January 2008. www.phgfoundation.org March 2008.
44. Baranova E.V. DNA – making self-acquaintance or how to extend the Youth Moscow, ACT, SPB, Astrel 2006, 222p.
45. Brand A., Brand H., Baumen T.C. // Europ. J. Human Gene. 2008. V.16. P. 5-13.
46. Tucker T., Marta V., Friedman J. M. // Am. J. Hum.Genet. 2009. V. 85. P. 142-154.

D-Arabinose Methabolism: Characterization of Bifunctional Arabinokinase/Pyrophosphorylase of *Leishmania major*

N.M. Novozhilova*, N.V. Bovin

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, RAS

*E-mail: nnovogilov@gmail.com

ABSTRACT In this work we describe an unusual enzyme from *Leishmania major* (Arabinokinase/Pyrophosphorylase) that catalyzes the synthesis of GDP-D-arabinopyranose (GDP-D-Arap) via a D-arabinose-1-phosphate intermediate in the presence of ATP and GTP. Our data indicate GDP-D-Arap transport *in vivo* by the LPG2 multispecific nucleotide sugar transporter into the *Leishmania* Golgi apparatus, in which it can be used by glycosyltransferases as a donor substrate for glycosylation.

Keywords: D-arabinopyranose, *Leishmania*, bifunctional enzyme, kinase, pyriphosphorylase, lipophosphoglycan.

Abbreviations: D-Arap – D-arabinopyranose; L-Fuc – L-fucose; K_m – the Michaelis-Menten constant; V_{max} – the maximum reaction velocity.

INTRODUCTION

It is known that both bacteria (*Bacteriodes*) and plants (*Arabidopsis*) can synthesize GDP-L-Fuc from L-fucose (L-Fuc) via intermediate L-fucose-1-phosphate using a bifunctional enzyme L-fucokinase/GDP-L-fucose pyrophosphorylase [3, 4]. Since D-Arap and L-Fuc are structurally similar, it makes sense to assume that biosynthesis of GDP-D-Arap in *Leishmania* can occur through a mechanism similar to that of GDP-L-Fuc biosynthesis in other species. To check this hypothesis, the *L.major* genome was evaluated for open reading frames homologous to fucokinase and GDP-L-fucose pyrophosphorylase gene sequences. As a result, two near-identical genes (*lmjF16.0440* and *lmjF16.0480*) were found as possessing high homology with the *Bacteriodes fragilis fkp* and *Arabidopsis thaliana at1g01220* genes, both encoding L-fucokinase/GDP-L-fucose pyrophosphorylase [3, 4].

The open reading frames *lmjF16.0440* and *lmjF16.0480* correspond to putative polypeptides composed of 1,187 aminoacid residues with a calculated molecular mass of 126.5 kDa. The only difference (three aminoacid residues) between these two proteins is the segment 196–199, namely PheGlnAsnHis in LmjF16.0480 and LeuGlnAspTyr in LmjF16.0440. The N-terminal sequences of these proteins contain a highly conserved segment Val100–Lys117 that closely resembles the conserved pyrophosphorylase motif Lys(X)₂GlyXThrXMet(X)₄Lys [5]. The C-termini also contain a conserved segment Gly950–Ile958 homologous to the GlyXGly(X)₂Gly(Ser)₂Gly motif forming the ATP-binding pocket of many kinases [6]. The presence of these conserved sequences has allowed to attribute kinase and pyrophosphorylase activities to the proteins' C- and N-ends, respectively.

RESULTS

The *lmjF16.0440* and *lmjF16.0480* genes were cloned in the pET16b plasmid vector for further examination of the activities of the putative enzymes. The recombinant LmjF16.0440 and LmjF16.0480 were expressed in *E.coli* BL21 (DE3)-RIPL cells and tested for enzyme activity (data are shown on Fig. 1). LmjF16.0480 catalyzed the formation of D-Arap-phosphate and GDP-D-Arap, whereas LmjF16.0440 only synthesized D-Arap-phosphate. Reaction products were not observed when the reaction mixture contained only nonspecific proteins extracted from the *E.coli* cells transformed with an empty pET16b vector. This information allows to conclude that the product of the *lmjF16.0480* gene really acts as a bifunctional arabinokinase/pyrophosphorylase; that is, it exhibits both D-arabinokinase and GDP-D-Arap-pyrophosphorylase activity, whereas the recombinant LmjF16.0440 protein only exhibits D-arabinokinase activity.

The necessity for the presence of ATP for kinase activity and GTP for pyrophosphorylase activity in the proteins stud-

Table 1. Kinetic Parameters for Recombinant LmjF16.0440 and LmjF16.0480 Proteins

Enzyme	Substrate	K_m (mM)	V_{max} ($\mu\text{M/L}/\text{min}$)
LmjF16.0440	D-Arabinose	1.23	0.35
	D-Arabinose-1-phosphate	-	-
LmjF16.0480	D-Arabinose	2.90	0.50
	D-Arabinose-1-phosphate	67.0	160.0

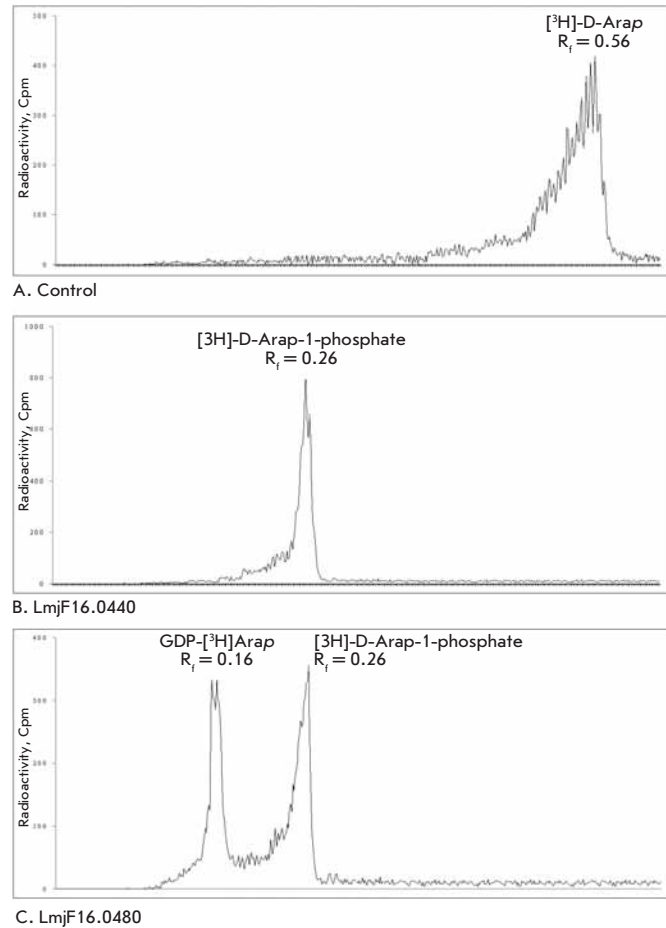


Fig. 1. Chromatographic analysis of enzymatic reaction products. Cultures of *E. coli* BL21(DE3) were transformed with pET16b-LmjF16.0440(His)₆ and pET16b-LmjF16.0480(His)₆ plasmids and induced with IPTG. Following lysis of the induced cells, the proteins were solubilized with 0.1% Tween 20 and immobilized to a column of TALON resin via N-terminal His-tag. Enzymatic activities of the TALON-bound recombinant LmjF16.0440 and LmjF16.0480 then were tested *in vitro*. Reaction mixtures contained 5 mM ATP, 5 mM GTP, 5 mM MgSO₄ and 0.15 μCi of [³H]D-Arap as substrate. Assay tubes were incubated overnight at 37°C, and the reaction products were analysed by TLC on glass-backed Silica Gel 60 TLC sheets (1-butanol/acidic acid/water, 2:1:1). [³H]D-Arap-containing products were located using a Bio Scan System 200 radiochromatogram scanner. Control assay containing only nonspecific proteins from *E. coli* cells failed to produce any product (A). Reaction containing the LmjF16.0440 protein produced only D-Arap-phosphate (B), whereas the LmjF16.0480 protein produced [³H]arabinose 1-phosphate and GDP-D-[³H]Arap (C)

ied was confirmed when carrying out the enzymatic reactions in the presence of only one triphosphate. Formation of the reaction product was not observed in the absence of both ATP and GTP. When only ATP was present, the reaction was not completed and led to the formation of arabinose-phosphate only. When only GTP was present in the reaction mixture, a small amount of GDP-D-Arap was produced, apparently due to the enzyme's capability to utilize GTP instead of ATP. However, in this case the reaction proceeded slowly with a low yield of the reaction products as compared with the reac-

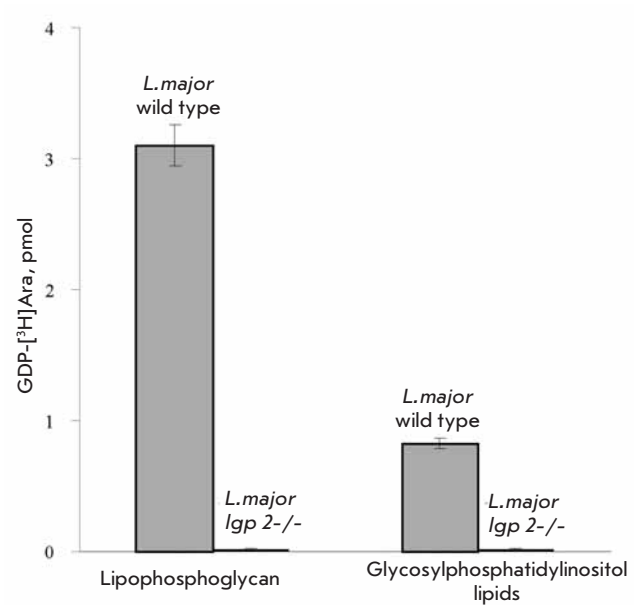


Fig. 2. Identification of the protein responsible for GDP-D-Ara transport into the Goldgi. Cultures of *L. major* wild-type and knock-out mutants *lpg2-/-* (*L. major lpg2-/-*) were metabolically labeled with [³H]D-Arap, their surface glycoconjugates (lipophosphoglycan and glycosylphosphatidylinositol lipids) were extracted as described [8], and the radioactivity in all fractions was measured by scintillation counting. [³H]D-Arap incorporation of the cell surface phosphoglycans was abolished in the *L. major lpg2-/-* knock-out mutant, lacking the multispecific nucleotide sugar transporter LPG2

tion in the presence of both ATP and GTP. Thus, the entire reaction requires both triphosphates in the reaction mixture: ATP – for kinase activity and GTP – for pyrophosphorylase activity of the bifunctional enzymes under study.

The kinetic parameters, namely, the Michaelis constant (K_m) and maximum reaction velocity (V_{max}), were determined for recombinant LmjF16.0440 and LmjF16.0480. These parameters are shown in Table 1 for each reaction catalyzed by each of the enzymes. Both enzymes demonstrate similar features as kinases. However, an extremely high $K_m = 67$ mM was determined for pyrophosphorylation catalyzed by LmjF16.0480. Such low affinity to arabinose-phosphate could suggest the association of both catalytic centers of the enzyme, so that the D-Arap-1-phosphate from the kinase center moves straight to the pyrophosphorylase one. So, a considerable amount of exogenous arabinose-phosphate is necessary for its appearance in the pyrophosphatase active center, which is reflected in the high K_m value.

The human genome contains genes that are partially homologous to the bifunctional enzyme from *Leishmania*. In particular, the locus NP 659496 encodes a protein composed of 1,185 aminoacid residues with 44% homology with the kinase C-end, whereas its N-end has no significant resemblance with pyrophosphorylase domains. The gene was cloned, and its recombinant product was tested for GDP-D-Arap synthesizing capability. This test has shown no activity. The absence of pyrophosphorylase activity for the human protein is not surprising considering the low homology of the human

protein N-end with the pyrophosphorylase domains. The D-arabinose-activating capability was also lacking, despite the close resemblance between the kinase C-end and both LmjF16.0440 and LmjF16.0480.

CONCLUSION

In most cases, activated carbohydrates are synthesized in cytoplasm and subsequently transported into the Golgi lumen, in which they are utilized by corresponding glycosyltransferases as substrate donors in glycosylation reactions. Earlier, the LPG2 protein found in the *Leishmania* Golgi apparatus membrane was classified as a multispecific transporter that can transport not only GDP-Man, but also GDP-D-Arap and GDP-L-Fuc [7]. To make sure that LPG2 is the only transporter responsible for GDP-D-Arap transport into the Golgi apparatus, we grew the cells of wild-type *L. major* and its *lpg2* knockout mutant (*L. major lpg2*^{-/-}) in the presence of [³H]-arabinose and determined [³H]Ara incorporated into the cell surface glycoconjugates (fig. 2). One can see from the presented data that the *lpg2* knockout cells do not incorporate [³H]-arabinose. The absence of arabinose in lipophosphoglycan of *L. major lpg2*^{-/-} is easy to explain, because the loss of LPG2 leads to cancellation of GDP-Man transport into the Golgi apparatus with the corresponding termination of the synthesis of the carbohydrate moiety of the molecule [9].

Thus, even in the presence of GDP-Arap transported into the Golgi lumen by another protein, arabinosyltransferases had no acceptor site for arabinose transfer to lipophosphoglycan. There is no telling this about glycosylinositolphospholipids. Glycosylation of these molecules occurs through another biosynthetic pathway with dolichol-phosphate mannose as a donor [10], so the absence of arabinose residues in glycosylinositolphospholipid molecules of *L. major lpg2*^{-/-} can be explained by cancellation of GDP-Arap transport into the Golgi apparatus in the cells devoid of *lpg2* gene.

Thus, our data very likely suggest GDP-D-Arap transport from cytoplasm (in which it is synthesized) into the Golgi lumen by the multi-functional transporter protein LPG2. We have also classified the *L. major lmjF16.0480* gene product as a bifunctional arabinokinase/pyrophosphorylase that can synthesize GDP-D-Arap from D-Arap via the intermediate D-arabino-1-phosphate in the presence of both ATP and GTP. ●

The study was supported by the Program of Presidium of Russian Academy of Sciences “Molecular and Cell Biology.”

*The authors are grateful to Prof. H. Guo (University of Washington, USA) for assistance in identification and cloning of *Leishmania major* and human genes and to Prof. S. Turco (University of Kentucky, USA) for fruitful discussions.*

REFERENCES

1. McConville M.J., Turco S., Ferguson M., Sacks D. // *EMBO J.* 1992. V. 11. P. 3593–3600.
2. Schneider P., McConville M., Ferguson M. // *J Biol Chem.* 1994. V. 269. P. 18332–18337.
3. Coyne M., Reinap B., Lee M., Comstock L. // *Science.* 2005. V. 307. P. 1778–1781.
4. Kotake T., Hojo S., Tajima N., Matsuoka K., Koyama T. *et al* // *JBC.* 2008. V. 283. P. 8125–8135.
5. Peneff C., Ferrari P., Charrier V., Taburet Y., Monnier C. *et al* // *EMBO J.* 2001. V. 20. P. 6191–6202.
6. Hanks S., Quinn A., Hunter T. // *Science.* 1988. V. 241. P. 42–52.
7. Hong K., Ma D., Beverley S., Turco S. // *Biochemistry.* 2000. V. 39. P. 2013–2022.
8. McConville M.J., Bacic A. // *JBC.* 1989. V. 264. P. 757–766.
9. Späth G., Lye L., Segawa H., Sacks D., Turco S. *et al* // *Science.* 2003. V. 301. P. 1241–1243.
10. Ilgoutz S.C., Zawadzki J.L., Ralton J.E., McConville M.J. // *EMBO J.* 1999. V. 18. P. 2746–2755.

Protein Tyrosine Kinase Panel As a Tool for Anticancer Drug Design

T.V. Rakitina^{1,2,*}, O.V. Yudkina^{3,4}, E.V. Smirnova¹, A.V. Lipkin^{3,4}

¹ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences

² A.N. Bakh Institute of Biochemistry, Russian Academy of Sciences

³ Institute of Crystallography, Russian Academy of Sciences

⁴ Russian Research Center Kurchatov Institute

*E-mail: taniarakitina@yahoo.com

ABSTRACT The discovery of the pharmaceutical potential of small molecule inhibitors of oncogenic protein tyrosine kinases is one of the directions in target therapy in oncology. Presently, investigations aiming at developing new therapeutically important inhibitors have to be based on a combination of computational and experimental approaches including biochemical, cell-based or *in silico* screening and the study of the three-dimensional structure of the kinase active center, in complex with an inhibitor, using crystallography and X-ray analysis or molecular modeling. This work is an example of a combination of inhibitor experimental search with the computational analysis of the potential mechanism of the inhibitors' action, which allowed to propose the 2-hydroxyphenol group as a scaffold for the design of new tyrosine kinase inhibitors.

Keywords: protein tyrosine kinases; small molecule inhibitors; screening; 2-hydroxyphenol group.

INTRODUCTION

The discovery of the role of oncogenic protein tyrosine kinases (PTKs) capable of noncontrolled activation in the development of cancer due to genetic alterations and the therapeutic potential of their inhibition has led to the emergence of a new era in oncology characterized by the appearance of selective (targeted to the specific protein) drugs in clinical practice [1]. Application of the small molecule inhibitors that prevent the binding of ATP to the catalytic domains of PTKs, thus interfering with the activities of the kinases, was considered as the most promising strategy in the inhibition of oncogenic PTKs [2]. The first ATP-competitive drug successfully used in human therapy was imatinib (gleevec), which acts against the Kit and PDGF receptors and inhibits nonreceptor fusion kinase Bcr-Abl [3]. Recently, a number of inhibitors have been approved for clinical use and more of them are at different stages of evaluation. However, the search for novel classes of chemical compounds acting against PTKs continues [4]. The ever-increasing degree of activity shown by scientists in the field of protein kinase inhibitor development could be attributed to the mentioned role of the majority of PTKs in oncogenesis [2], as well as the phenomena of resistance by some mutant PTKs to known inhibitors [5]. Furthermore, the problem of the low selectivity of ATP-competitive small molecule inhibitors [6] and, on the other hand, the therapeutic advantage of parallel inactivation of several oncogenic key points [7] deserve notice.

Modern research aiming at developing new therapeutically important inhibitors has to be based on a combination of computational and experimental approaches including biochemical, cell-based, or *in silico* screening and the study of the three-dimensional structure of the kinase active center, in complex with an inhibitor, using crystallography and X-ray analysis or molecular modeling [8, 9]. Obviously, computational methods require information on the three-dimensional structure of the active center of a target protein or its homo-

logues, even as all modeling predictions have to be validated experimentally. Thus, the search for novel active compounds and the assessment of the known inhibitor's molecular specificity require the generation of various recombinant PTK panels [6]. The conventional approach to obtaining functionally active PTKs is the baculoviral expression system [10].

The aim of the present study was to generate a panel of functionally active protein tyrosine kinases and to search for their inhibitors in a small-molecule collection. Analysis of the screening results using molecular modeling allowed to propose the 2-hydroxyphenol group as a potential scaffold for the design of novel tyrosine kinase inhibitors.

MATERIALS AND METHODS

MATERIALS

To obtain functionally active protein tyrosine kinases, baculovirus expression system «Bac-to-Bac» (Invitrogen, USA) was used.

For the search for the protein tyrosine kinase inhibitor, a collection of small organic molecules containing more than ten thousand individual compounds with molecular weights ranging from 150 to 600 was provided by Chembridge Corp. Moscow (<http://chembridge.com/datasheets/KINASet.pdf>). All compounds were dissolved in DMSO to a concentration of 1 mM, and the aliquots were stored at -20°C.

METHODS

Generation of baculoviruses. cDNAs corresponding to the 16 PTKs were cloned into the T-easy vector (Promega, USA) using the conventional approach of reverse transcription – PCR with the total RNA as a template. The full list of the PTKs and the summary of cloning are presented in Table 1. Once the cloned inserts were verified by sequencing, cDNA fragments were recloned into the vector pFastBacHT-B downstream of and in frame with 6xHis-tag. The obtained plasmids were

RESEARCH ARTICLES

used to transform *E. coli* cells *DH10Bac* (Invitrogen, USA), wherein the recombination of the target genes with the baculoviral genome occurred. Recombinant bacmides isolated from *E. coli* were introduced into insect cells *Spodoptera frugiperda* (Sf9) plated on 6-well plates and cultivated at 27°C. Virus particles assembled in the transfected cells and then induced cell lysis and accumulated in the growth media. Infected insect cells were expressing and accumulating recombinant target 6xHis-PTKs. All manipulations with baculoviruses were conducted according to the manufacturer's instructions for the expression system Bac-to-Bac (Invitrogen, USA).

Purification of 6xHis-PTKs. Cell pellets (~10⁹ cells or 10 g of a biomass) from 1 L of the infected cell culture were frozen at -70°C and then lysed in 50 ml of buffer A (20 mM Tris-HCl pH 8.5, 500 mM NaCl, and 0.1 % Triton X-100) supplemented with 10 u of RQ1 DNAase and a protease inhibitor cocktail. The lysate was cleared by centrifugation (15,000 g, 1 h, +4°C) and incubated with 2 ml of Ni²⁺-sepharose at +4°C for an

hour. After the binding, the resin was washed with buffer A containing 30 mM imidasol until the protein detected in the wash fractions was absent. Then the 6xHis-PTK was eluted with buffer A containing 350 mM imidasol. For the exchange of the elution buffer to the storage buffer (50 mM Tris-HCl pH 8.5, 100 mM NaCl, and 0.05 % Triton X-100), PD10 columns were used according to the manufacturer's recommendations. All isolated proteins were stored at -20 °C in 50% glycerol and 2 mM DTT.

Measurement of 6xHis-PTK activities. The reaction mixture for each kinase assay contained 10 μM ATP, 10 μg of universal protein tyrosine kinase substrate poly(Glu₄-Tyr) (Sigma, USA), and one of the sixteen PTKs in amounts of 30, 60 or 90 ng in the 1xkinase buffer (50 mM Tris-HCl, pH 7.5, 5 mM MnCl₂, 5 mM MgCl₂, 0.01% Tween-20, and 2 mM DTT). To control the starting ATP level, a reaction mixture lacking kinase was prepared and introduced in each set of assays. Kinase assays were performed in a total volume of 30 μL, in

Table 1. Summary of the PTK panel generation

Protein Tyrosine Kinase (PTK)	Acc. No AA	Direct & Reverse primers ¹ (5'-3')
Abl - Abelson Murine Leukemia Viral Oncogene Homolog 1	NP_005148 2-601	GGATCCTTGGAGATCTGCCTGAAGCTG ACTCGAGCCGAACAAGTTGGTCTTTTG
Alk - Anaplastic Lymphoma Kinase Receptor	NP_004295 1092-1406	CATGGATCCCTACAACCCCAACTAC GCTCGAGTTATTCCACAAGTGGACCAT
Blk - B-Lymphocyte Kinase	NP_001706 2-505	GGATCCGGGCTGGTAAGTAGCA CTCGAGGGCTGCAGCTCGTACTG
CSF1R - Colony Stimulating Factor 1 Receptor	NP_005202 545-972	GGATCCAAGTACCAGGTCCGCT CTCGAGCAGAACTGATAGTTGTTG
Csk - C-Terminal c-Src Kinase	NP_004374 2-450	TGGATCCTCAGCAATACAGGCCGCC ACTCGAGAGGTGCAGCTCGTGGGTT
Eph A2 - Ephrin Receptor A2	NP_004422 562-976	AGATCTAGGAGGAAGAACCAGC CTCGAGATGGGGATCCCCACAG
FGFR1 - Fibroblast Growth Factor Receptor 1	NP_056934 398-820	GGATCCAAGAGTGGTACCAAGAAGAGT TTCTCGAGCGCGTTTGAGTCCGCCATT
FGFR2 - Fibroblast Growth Factor Receptor 2	NP_075259 402-822	GGATCCAAGAACACGACCAAGAAGC CTCGAGGTTTAAACACTGCCGTTTATG
IGFR1 - Insulin-like Growth Factor 1 Receptor	NP_000866 974-1294	GGATCCAGAAAGAGAAATAACAGCAGG GCTCGAGTTAATCCAGCTCCTCCGGCTC
InsR - Insulin Receptor	NP_000199 982-1382	GGATCCAGGCAGCCAGATGGGCCGCTG CTCGAGGAAGGATTGGACCGAGGCAAG
Kit - Stem Cell Factor Receptor	NP_000213 545-976	TGGATCCTACAAATATTTACAGAAACCC TTCTCGAGACATCGTCGTGCACAAGCAG
Lyn - Yamaguchi Sarcoma Viral Related Oncogene Homolog 2	NP_002341 full length	GGGATCCGGATGTATAAAATCAAAAAGG GGAATTCTCGAGGGCTGCTGCTGGTATT
PDGFRα - Platelet-Derived Growth Factor Receptor-α	NP_006197 552-1089	GGATCCAAGCCAGTTACGAGATCCGAT GTCGACAGGAAGCTATCTCTGCTTCCG
Pyk2 - Focal Adhesion Kinase 2	NP_004094 353-762	TGGATCCCGGCTGCAGGGTGAGCACCA TTCTCGAGTTAACGGGAGATGGATACTC
Syk - Spleen Tyrosine Kinase	NP_003168 full length	GGATCCGCCAGCAGCGGCATGGCTGAC CTCGAGTTCACCACGTATAGTAGTA
Yes - Yamaguchi Sarcoma Viral Oncogene Homolog 1	NP_005424 11-542	CGGGATCCCCAGCCATTAATAACAGAC TCGTGACAAATTTCTCCTGGCTGGTA

¹ - sites used for subcloning from T-easy vector into pFastBacHT-B are in bold.

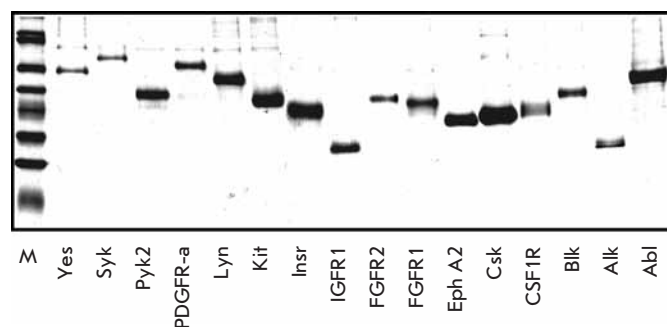


Fig. 1. Analysis of purified samples of recombinant 6xHis-PTKs. The 10 μ L of protein stocks were loaded onto 10 % SDS-PAAG and the gel was stained with Coomassie G-250

384-well plates at 30°C for 15 min and developed with 10 μ L of the Kinase-Glo reagent. The luminescence was detected using the Fusion Universal Microplate Analyzer (PerkinElmer USA). The activities of the recombinant PTKs were assessed in two independent experiments performed in triplicates.

Screening of the small-molecule collections. The screening was performed by assessment of the chemical compound's potential to inhibit tyrosine kinase activities using the luminescent kinase assay. All kinase reaction components were freshly diluted in the 1xkinase buffer. Assays were performed in 384 well microplates manually or using MultiPROBE II (Packard, USA). Control samples containing ATP and kinase with DMSO instead of the inhibitors (0% inhibition) and ATP without kinase (100% inhibition) were included in each assay plate. Kinase assays were set up and performed as follows: 1) Add 10 μ L of 30 μ M inhibitor (to a final concentration of 10 μ M) or 3% DMSO to a well, 2) Add 10 μ L of the appropriate kinase dilution, 3) incubate 20 min at 20°C, 4) add 10 μ L of 30 μ M ATP with 10 μ g of poly(Glu₄-Tyr), 5) incubate 90 min at 30°C, 6) add 10 μ L of the Kinase-Glo reagent and measure the luminescence.

RESULTS AND DISCUSSION

To generate the protein tyrosine kinase panel for the screening of the chemical collection, 16 PTKs from five families of both receptor and cytoplasmic PTKs were selected (Table 1). For the receptor and certain cytoplasmic PTKs, protein fragments containing kinase domains were used, instead of full-length proteins, as they are believed to be a good functional model for the search and study of kinase inhibitors [11]. Expression of the recombinant proteins was carried out in the baculovirus expression system, which is believed to be optimal for obtaining functionally active PTKs [10]. N-terminal 6xHis-tags were added to the kinases that allowed one-step purification of the His-tagged PTKs from infected insect cells using metal chelate affinity chromatography on Ni²⁺-sepharose. Protein purification was conducted following instructions appropriate for this type of chromatography and according to the results of the analytical experiments on the optimization of purification conditions. The final protocol is presented in the Material and Methods section.

Purified proteins were analyzed by SDS-PAAGE in de-

naturing conditions, followed by Coomassie G-250 staining (Fig. 1). All proteins showed good correlation of their electrophoretic mobilities with calculated molecular weights, and the purities of isolated kinases were at least 70% (Table 2). The identities of all sixteen recombinant 6xHis-PTKs were further confirmed by MALDI-TOF mass spectrometry. The amounts of purified proteins were determined using a Bradford protein assay, and the yields of purified kinases from 1 L of infected insect cells were 1 to 20 mg (Table 2).

The activities of the isolated PTKs were assessed by direct one-step measurement of the amount of ATP in the kinase assay using a luminescent Kinase-Glo reagent, the method having good compatibility with high throughput screening [12]. Relative luminescence units measured in the kinase assay or control reaction were converted to ATP molarities in the reaction mixture using the standard curve from the ATP titration experiment, then the amounts of ATP (pmoles)

Table 2. A summary of the 6xHis-PTKs in typical expression and purification experiment

6xHis-PTK	MW ¹ (kDa)	Purity ² (%)	Yield ³ (mg)	Specific activity ⁴ , (nmoles/min × mg)	Kinase amount used in one assay ⁵ , ng/assay
Abl	72.5	85	6	118	50
Alk	38.8	75	1	60	90
Blk	63.8	80	3.5	91	60
CSF1R	54.0	80	5	98	60
Csk	56.7	90	20	153	40
Eph A2	53.4	95	10	110	50
FGFR1	52.5	70	1.5	223	20
FGFR2	53.8	70	1.5	245	20
IGFR1	39.8	80	2	147	40
Insr	51.4	85	6.5	131	40
Kit	55.0	75	3	80	70
Lyn	63.4	75	1.5	332	20
PDGFR-a	67.4	70	2.5	84	70
Pyk2	52.0	80	3.5	86	70
Syk	78	70	1,9	131	50
Yes	65.7	70	1,5	273	20

¹ - Calculated molecular weights of the recombinant 6xHis-PTKs.

² - The purities of the 6xHis-PTKs were determined by Coomassie G-250 staining of SDS-PAAG.

³ - The yields of purified 6xHis-PTKs from 1 L culture of insect cells (10 g) were calculated on the basis of the protein concentration determined by Bradford assay.

⁴ - Specific activities were determined by luminescent kinase assay at a final protein concentration of 1.5 μ g/ml (Fig. 2).

⁵ - The amounts of 6xHis-PTKs hydrolyzing at maximum 80% of the initial amount of ATP in 90 min of incubation time were calculated on a basis of specific activities and verified experimentally.

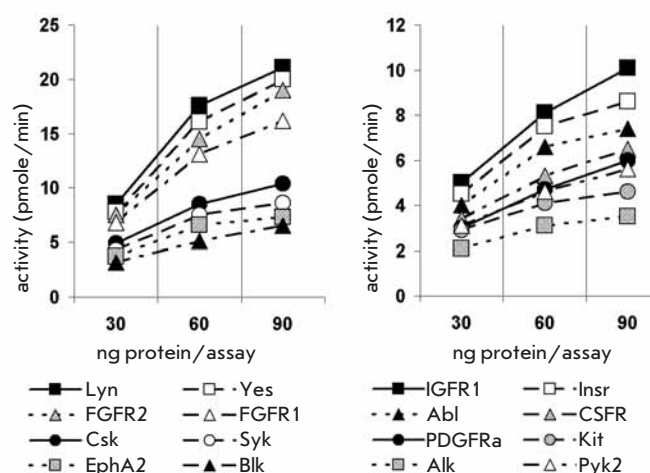


Fig. 2. Activities of 6xHis-PTKs. The activities were measured by luminescent kinase assay and plotted as pmoles of phosphate transferred from ATP to the poly (Gly₄-Tyr) substrate per min versus increasing amounts of the kinases. The averages of the results of two independent experiments performed in triplicates are presented

hydrolyzed by each kinase in a minute were calculated and plotted versus the amount of the kinase (Fig. 2). The specific activities of the recombinant enzymes were determined at a final protein concentration of 1.5 µg/ml (45 ng per assay) as nmoles of the phosphate transferred from ATP to the substrate per min per mg of kinase in standard conditions. The specific activities of the PTKs and the amounts to be added to the kinase reactions for the screening of the chemical collection were calculated on the basis of kinase specific activities and summarized in Table 2.

The protocol of the screening experiment is presented in the Material and Methods section. Potential inhibitors were

tested simultaneously against each of the 16 target kinases in a single working concentration of 10 µM. The screening protocol was designed taking into account the fact that the Kinase-Glo reagent provides the linearity of the luminescent signal to the ATP amount in an ATP concentration range of 1 to 100 µM. Based on the ATP starting concentration of 10 µM, the amount of kinase in each assay was set so that the enzyme could not convert more than 80% of ATP. The parameter was chosen on the basis of the specific activity of the PTK and verified in an analytical experiment. After the development of the kinase reaction and measurement of the luminescence, the relative luciferase units (assay outcome data) were normalized to the controls and presented as percentages of inhibition. The validity of the screening results was confirmed by calculating Z' factor values for each plate using the method of Zhang *et al.* [13]. The Z' values for most plates were >0.5, and the average Z' value of the entire screen was 0.59±0.1. Compounds capable of inhibiting kinase activity by at least 50% at 10 µM concentration were chosen for further validation in two independent experiments.

Six small molecule inhibitors of the PTKs comprising 2-hydroxyphenyl group were found in the screening experiment (Fig. 3). Database search using the Chemical Abstracts Service (<http://www.cas.org/expertise/cascontent/registry/index.html>) and PubChem BioAssay (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcassay>) showed that compounds I-VI have not been yet described as kinase inhibitors. These compounds were active against 4 to 9 panel PTKs. FGFR1, Abl, Blk, FGFR2 and PDGFR-*a*, as well as Lyn, Eph A2 and Csk, were the most frequent targets, while IGFR1, Pyk2, and Yes were rarely inhibited. Since all compounds comprised the 2-hydroxyphenyl group, this led us to suggest that this group is precisely the one responsible for the inhibitors' binding with the kinase domains. It is known that typical kinase inhibitors have pharmacophore (a minimal core structure that is responsible for the biological activity) in which the neighbor aromatic amino group and aromatic nitrogen or carbonyl oxygen form a couple of correlated hydrogen bonds

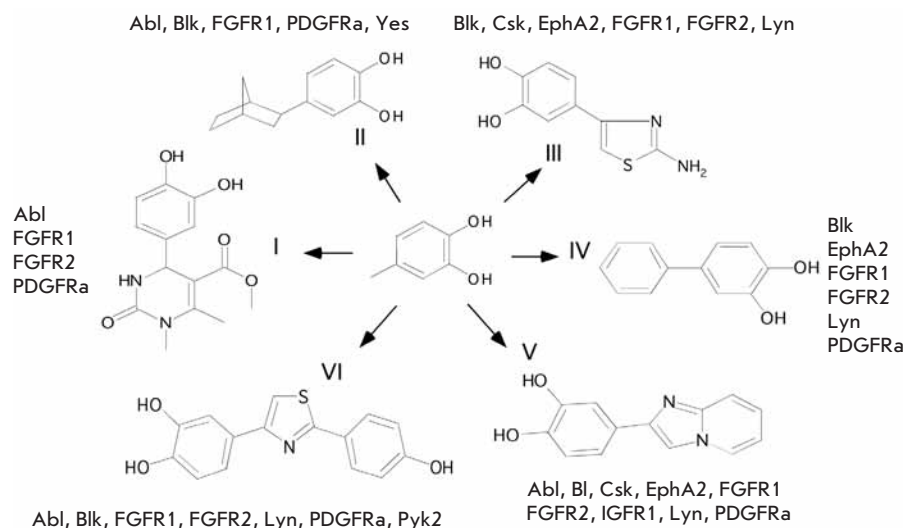


Fig. 3. The small molecule tyrosine kinase inhibitors that were found by screening of the chemical collection. The 2-hydroxyphenol group is in the center. The PTKs inhibited by each compound at least by 50% are shown

with the kinase domain hinge region, mimicking the hydrogen bonds formed between this region and the ATP molecule [14]. Molecular docking of the compounds I-VI to Csk, FGFR1, and FGFR2 having a lot of PDB structures available for modeling, which was performed using the Lead Finder software [15], revealed that the 2-hydroxyphenol moiety was also capable of forming a pair of hydrogen bonds with the kinase hinge regions and, thus, that the 2-hydroxyphenol group could be considered as a novel pharmacophore for tyrosine kinase inhibitors. Analysis of the Protein Data Bank (<http://www.pdb.org/>) for analogous patterns of kinase inhibition revealed that similar hydrogen bonds were formed in complexes of phosphoinositide 3-kinase with the flavonoids quercetin and myricetin. It should be noted that the molecular weights of the compounds I-VI are significantly lower than the MW of known PTK inhibitors, which probably explains their ability to dock to the ATP-binding pockets of various kinases. At the same time, the low MW of the compounds specifically provide the possibility of their further modification and combinations with the fragments of known inhibitors to increase selectivity and efficiency towards the selected PTK. In the future, we are planning to verify the

modeling results by the crystallization and X-ray analysis of the PTKs, in complex with the inhibitors.

CONCLUSIONS

We have reported here on the generation of a protein panel of the functionally active protein tyrosine kinases and the screening of the chemical collection of small molecules that allowed to identify six earlier unknown inhibitors comprising the 2-hydroxyphenol group. Molecular docking performed using the Lead Finder software revealed that the 2-hydroxyphenol scaffold could serve as a basis for the design of novel PTK inhibitors. ●

This work was supported by the Russian Federal Agency for Science and Innovation (federal contract 02.512.12.2051) and the Program of Presidium of Russian Academy of Sciences "Molecular and Cell Biology." We thank Chembridge Corp. Moscow for the chemical collection, Dr. R.X. Ziganshin (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry) for performing the experiments on MALDI-TOF mass spectrometry and MolTech Ltd, Moscow, for the molecular modeling.

REFERENCES

- Cohen P. // *Nat. Rev. Drug Discov.* 2002. V.1. P. 309–315.
- Gunby R.H., Sala E., Tartari C.J., Puttini M., Gambacorti-Passerini C., Mologni L. // *Anti-Cancer Agents in Med. Chem.* 2007. V.7. P. 594–611.
- Cohen M.H., Johnson J.R., Pazdur R. // *Clin. Cancer Res.* 2005. V. 11 (1). P. 12–9.
- Jianming Z., Priscilla L., Gray N.S. // *Nat. Rev. Cancer.* 2009. V. 9. P. 28–39.
- Hofmann W.-K., Jones L.C., Lemp N.A., de Vos S., Gschaidmeier H., Hoelzer D., Ottmann O.G., Koeffler H.P. // *Blood.* 2002. V. 99. P. 1860–1862.
- Bain J., Plater L., Elliott M., Shpiro N., Hastie J., McLauchlan H., Klevernic I., Arthur S.C., Alessi D., Cohen P. // *Biochem. J.* 2007. V. 408. P. 297–315.
- Faivre S., Djellou S., Raymond E. // *Semin. Oncol.* 2006. V. 33 (4). P. 407–420.
- Drews J. // *Science.* 2000. V. 287. P. 1960–1964.
- Anderson A. // *Chem. & Biol.* 2003. V. 10. P. 787–797.
- Chambers S.P., Austen D.A., Fulghum J.R., Kim W.M. // *Protein Expr. Purif.* 2004. V. 36. P. 40–47.
- Mologni L., Sala E., Riva B., Cesaro L., Cazzaniga S., Redaelli S., Marin O., Pasquato N., Donella-Deana A., Gambacorti-Passerini C. // *Protein Expr. Purif.* 2005. V. 41. P. 177–185.
- Koresawa M., Okabe T. // *Assay and Drug Dev. Technol.* 2004. V. 2. P. 153–160.
- Zhang J.H., Chung T.D., Oldenburg K.R. // *J. Biomol. Screen.* 1999, V.4, P. 67–73.
- Fabbro D., Ruetz S., Buchdunger E., Cowan-Jacob S.W., Fendrich G., Liebetanz J., Mestan J., O'Reilly T., Traxler P., Chaudhuri B., Fretz H., Zimmermann J., Meyer T., Caravatti G., Furet P., Manley P.W. // *Pharmacol Ther.* 2002. V. 93 (2–3). P. 79–98.
- Stroganov O.V., Novikov F.N., Stroylov V.S., Kulkov V., Chilov G.G. // *J. Chem. Inf. Model.* 2008. V. 48. P. 2371–2385.
- Walker E.H., Pacold M.E., Perisic O., Stephens L., Hawkins P.T., Whymann M.P., Williams R.L. // *Mol. Cell.* 2000. V. 6. P. 909–919.

Atomic Resolution Crystal Structure of NAD⁺-Dependent Formate Dehydrogenase from Bacterium *Moraxella sp. C-1*

I.G. Shabalin¹, K.M. Polyakov^{2,1}, V.I. Tishkov¹, V.O. Popov^{1*}

¹ A.N. Bach Institute of Biochemistry RAS

² V.A. Engelhardt Institute of Molecular Biology RAS

*E-mail: vpopov@inbi.ras.ru

ABSTRACT The crystal structure of the ternary complex of NAD⁺-dependent formate dehydrogenase from the methylotrophic bacterium *Moraxella sp. C-1* with the cofactor (NAD⁺) and the inhibitor (azide ion) was established at 1.1 Å resolution. The complex mimics the structure of the transition state of the enzymatic reaction. The structure was refined with anisotropic displacement parameters for non-hydrogen atoms to a R factor of 13.4%. Most of the nitrogen, oxygen, and carbon atoms were distinguished based on the analysis of the temperature factors and electron density peaks, with the result that side-chain rotamers of histidine residues and most of asparagine and glutamine residues were unambiguously determined. A comparative analysis of the structure of the ternary complex determined at the atomic resolution and the structure of this complex at 1.95 Å resolution was performed. In the atomic resolution structure, the covalent bonds in the nicotinamide group are somewhat changed in agreement with the results of quantum mechanical calculations, providing evidence that the cofactor acquires a bipolar form in the transition state of the enzymatic reaction.

Keywords: formate dehydrogenase, X-ray diffraction study, atomic resolution, enzymatic catalysis.

INTRODUCTION

Formate dehydrogenases catalyze the oxidation of formate into carbon dioxide and can be divided into several groups based on the quaternary structure, as well as on the presence of prosthetic groups and cofactors. The structurally simplest formate dehydrogenases are NAD⁺-dependent formate dehydrogenases (FDH, EC 1.2.1.2), which oxidize formate coupled with reduction of the coenzyme NAD⁺ into NADH [1]:



Formate dehydrogenases belong to a large superfamily of D-isomer specific 2-hydroxyacid dehydrogenases [2]. Formate dehydrogenases of this type contain no metal ions or prosthetic groups in the active sites and have a high specificity towards both NAD⁺ and formate. FDHs from different organisms (bacteria, yeast, plants) function as dimers consisting of two identical subunits with a molecular weight from 35 to 50 kDa. The molecular mechanism of FDH is characterized by the direct transfer of a hydride ion from the substrate to the C4 atom of the nicotinamide ring of NAD⁺ without additional proton transfer steps, which usually occurs in reactions catalyzed by related NAD⁺-dependent dehydrogenases. Hence, the FDH-catalyzed reaction is a convenient model for studying the mechanism of hydride ion transfer in the active site of NAD⁺-dependent hydrogenases by methods of quantum mechanics and molecular dynamics [3–5].

The knowledge of the three-dimensional structure of the enzyme under study with accurate atomic coordinates is of great importance for the investigation of the molecu-

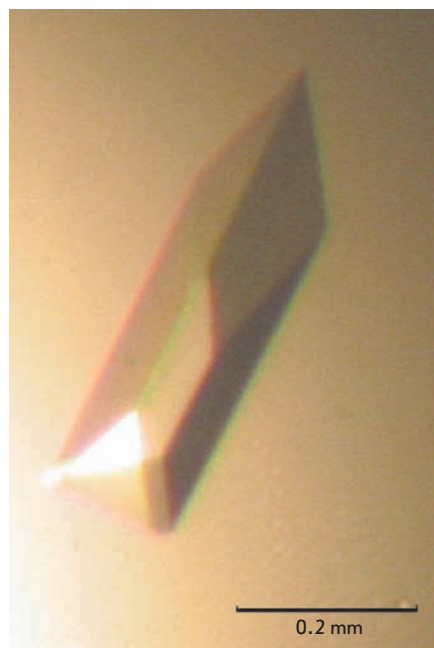
lar mechanisms of the enzyme. Based on X-ray diffraction data at atomic resolution, the protein structure can be refined with anisotropic displacement parameters for individual non-hydrogen atoms, which substantially increases the reliability of structural information and allows the determination of atomic coordinates with 0.02 Å accuracy [6]. At this resolution, it is possible to reveal the fine-structural organization of the active site, which cannot be done at lower resolution.

Currently, 937 protein structures with atomic resolution are available in the Protein Data Bank (PDB, <http://www.rcsb.org>). This is approximately 1.6 % of the total number of structures in the PDB. Atomic resolution structures were solved only for three NAD⁺-dependent dehydrogenases, such as horse liver alcohol dehydrogenase (five structures) [7–9], R-specific alcohol dehydrogenase from *Lactobacillus brevis* (three structures) [10], and lactate dehydrogenase from *Plasmodium falciparum* (one structure) [11]. The available three-dimensional structures of FDH were determined at resolutions no higher than 1.8 Å for the holo-form [12] and 1.55 Å for the apo-form [13]. In this work, the three-dimensional structure of the ternary complex of formate dehydrogenase from the methylotrophic bacterium *Moraxella sp. C-1* (MorFDH) with NAD⁺ and an inhibitor (azide ion), which mimics the transition state of the enzymatic reaction, was established at atomic resolution (1.1 Å).

MATERIALS AND METHODS

Crystals of the MorFDH-NAD⁺-azide ternary complex were grown by the hanging drop vapor diffusion method. Recombinant full-size MorFDH was expressed and purified accord-

Fig. 1. Photograph of crystals of the ternary MorFDH-NAD⁺-azide complex



ing to the procedure described in [14]. The purity of the enzyme estimated by polyacrylamide gel electrophoresis was at least 97%. The crystallization was performed using a solution of MorFDH at a concentration of 10.5 mg/ml in 0.1 M potassium phosphate buffer, pH 7.0, containing 5 mM NAD⁺ and 5 mM sodium azide. The reservoir solution consisted of 0.1 M Bis-Tris buffer, pH 6.5, and 2.0 M ammonium sulfate. The hanging drops (4 μ l) were prepared by mixing equal volumes of the protein and reservoir solutions on siliconized glass cover slides. Wells of plastic Linbro plates (Hampton Research) were filled with 500 μ l of the reservoir solution, and the cover slides were inverted and sealed over the wells. The plates were stored in a temperature-controlled cabinet at 20 °C. The crystals grew within two months to an average size of 0.6 \times 0.3 \times 0.2 mm (Fig. 1).

X-ray diffraction data were collected at a wavelength of 0.8166 Å using a MAR165 CCD detector at the EMBL X11 beamline at the DORIS storage ring of the DESY synchrotron (Hamburg). To decrease radiation damage in the protein crystal and increase the resolution due to a decrease in the thermal displacements of the atoms, the crystal was cooled with nitrogen (100 K) using a Cryojet cryocooler during data collection. Before freezing in a stream of nitrogen, the crystal was soaked in a cryosolution (0.1 M Bis-Tris buffer, pH 6.5, 2.3 M ammonium sulfate, and 30 % (v/v) glycerol) for 30 s. Two X-ray diffraction data sets were collected from the same crystal. The first data set was collected in the resolution range 20–1.5 Å with a short exposure time; the second data set, in the resolution range 20–1.1 Å with a longer exposure time. This made it possible to avoid overexposure of low-resolution reflections, which is often the case when the exposure time is high, and to measure weak high-resolution reflections with high accuracy. The diffraction data sets were processed with DENZO and SCALEPACK program packages [15]. The crys-

tals belong to the space group C2 with the unit-cell parameters $a = 79.2$ Å, $b = 66.2$ Å, $c = 74.2$ Å, and $\beta = 103.4^\circ$.

Since these crystals are isomorphous with the crystals that were used for the structure determination of the MorFDH-NAD⁺-azide ternary complex at 1.95 Å resolution [14], the latter structure was used as the starting model. The structure refinement was performed using the REFMAC program [16]. The contribution of hydrogen atoms to the X-ray scattering was included in the refinement. The coordinates of the hydrogen atoms were calculated based on the known stereochemistry of amino acid residues and the coordinates of the corresponding covalently bonded atoms in each refinement cycle. The temperature factors for all non-hydrogen atoms were refined anisotropically. The progress of the refinement was monitored, the atomic model of the structure was visually inspected, and water molecules were located using the COOT program [17]. The corrections were made based on the analysis of difference Fourier maps with the coefficients $(2|F_o| - |F_c|)$ and $(|F_o| - |F_c|)$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes. The quality of the protein model was analyzed using the PROCHECK program [18]. The errors in the atomic coordinates were calculated using the SFCHECK program [19]. The diffraction data collection and atomic model statistics are summarized in Table 1.

Table 1. X-ray diffraction data and atomic model statistics. Values in parentheses are for the highest-resolution shell (1.11 - 1.10 Å)

Resolution, Å	20 - 1.1
No. of observed reflections	586345 (9155)
No. of unique reflections	145564 (3662)
Mosaicity, °	0.7
Redundancy	4.0 (2.5)
Completeness, %	96.5 (73.8)
R_{merge} , %	5.1 (52.9)
$\langle I \rangle / \langle \sigma(I) \rangle$	44 (1.9)
B factor from Wilson plot, Å ²	11.5
$R_{\text{work}}/R_{\text{free}}$, %	13.4/15.9
R.m.s. deviation of covalent bonds from ideal, Å	0.015
R.m.s. deviation of covalent angle from ideal, °	1.7
Minimal coordinate error, Å	0.014
Diffraction precision index (DPI), Å	0.028
No. of non-H atoms:	
protein	3106
water	437
ligands	84
Average B factors, Å ² :	
all atoms	20.7
main-chain atoms	18.1
side-chain atoms	20.1
water molecules	31.8
ligand atoms	24.9
No. of residues in Ramachandran plot*:	
in most favored regions	306
in additional allowed regions	33
in disallowed regions	1 [°]

*Except for Gly and Pro residues

[°]Ala198 in the FDH structure is outside the allowed region. The unusual conformation of this residue is discussed in detail in [12].

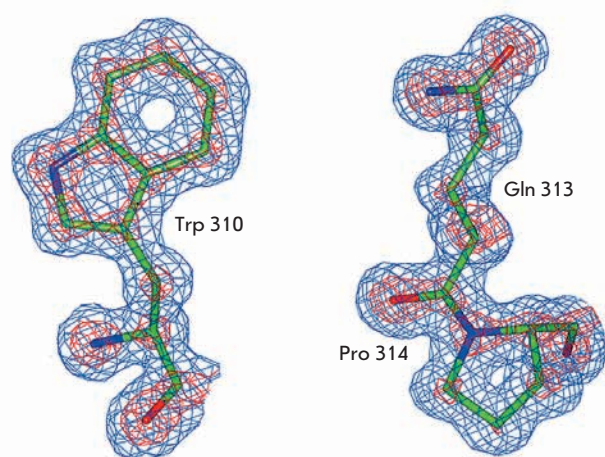


Fig. 2. Electron density map with $2|F_o| - |F_c|$ coefficients for residues Trp310, Gln313, and Pro314. The map is contoured at 1.0σ (blue) and 3.5σ (red) levels

RESULTS AND DISCUSSION

The great advantage of atomic resolution is the high ratio of observations to parameters, due to which the temperature factors of non-hydrogen atoms can be refined anisotropically. In the isotropic approximation, three positional parameters and the isotropic temperature factor of each non-hydrogen atom are refined, whereas six anisotropic thermal parameters are used instead of one isotropic temperature factor for each atom (three of these parameters provide the orientations of the principal axes of the ellipsoid; the other three parameters represent the magnitudes of displacement along the ellipsoid axes) in the anisotropic refinement. This approach allows a much more accurate interpretation of the diffraction data. The inclusion of the anisotropic parameters into the crystallographic refinement of the MorFDH-NAD⁺-azide structure lowered the R factor by 5.3 % and the R_{free} by 5.2 %. The mean weighted error of the atomic coordinates characterized by the diffraction precision index (DPI) also decreased from 0.036 to 0.028 Å. For the previously published structure of the MorFDH-NAD⁺-azide complex at 1.95 Å resolution, this value was 0.141 Å [14]. The high quality of the electron density maps for the atomic resolution structure is demonstrated in Fig. 2.

The polypeptide chain fold of the solved structure of the ternary MorFDH-NAD⁺-azide complex is depicted in Fig. 3. The asymmetric unit cell contains one subunit of the enzyme. The MorFDH molecule consists of two subunits related by a crystallographic twofold rotation axis (Fig. 3). In the crystal structure, 399 amino acid residues, 437 water molecules, one NAD⁺ molecule, four azide ions, and six glycerol molecules were located. The NAD⁺ molecule and one azide ion are bound in the active site of the enzyme, whereas three azide ions and all glycerol molecules are located on the surface of the protein globule. Like in the structure solved at 1.95 Å resolution, the C-terminal residues 392-399 were located in the atomic resolution structure. These eight residues were not visible in electron density maps calculated for the previously solved structure of the holo-form of FDH from the bacterium

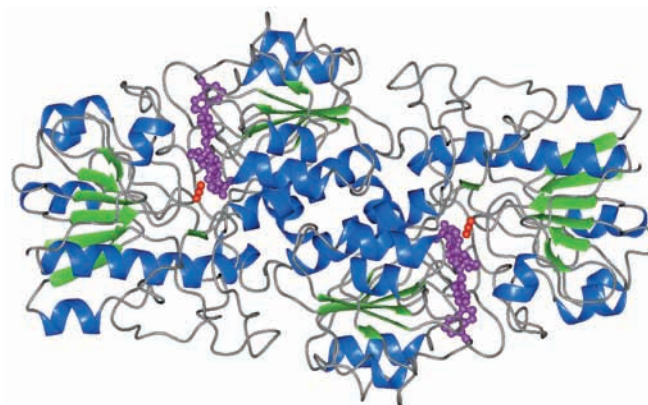


Fig. 3. Ribbon representation of the ternary MorFDH-NAD⁺-azide complex. The subunits of the MorFDH dimer are related to each other by the crystallographic twofold rotation axis perpendicular to the plane of the figure. The NAD⁺ molecule is depicted in purple, the azide ion is depicted in red

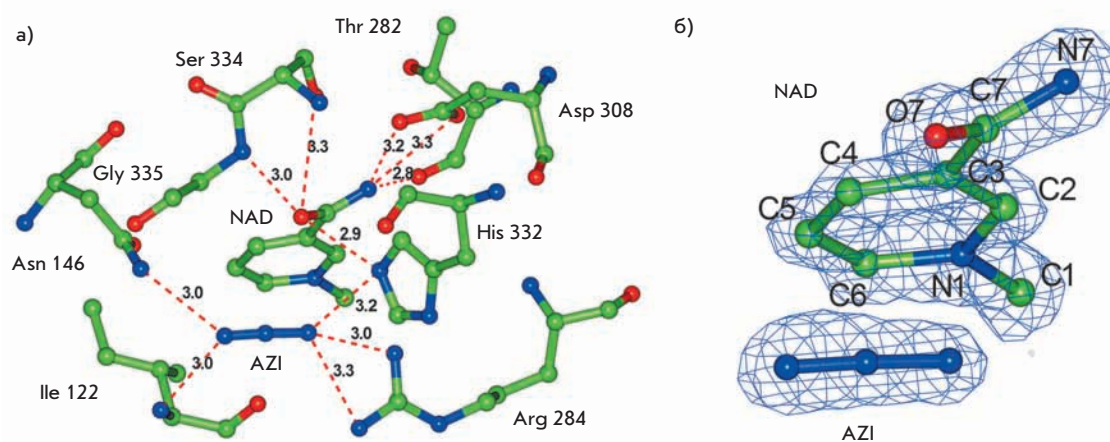
Pseudomonas sp. 101 [12]. Nevertheless, the last two residues were not located in the atomic resolution structure as well, which is probably due to the disorder of these residues.

The rms deviation of the coordinates of all C α atoms of the MorFDH-NAD⁺-azide structures solved at 1.1 Å and 1.95 Å resolution is 0.30 Å, which indicates that they are identical. The maximum deviation is 1.1 Å (for the residue Ser18). The atomic coordinates for 18 amino acid side chains located on the surface of the protein globule differ by more than 1 Å, with the maximum deviation being 9 Å (for the residue Arg26). These differences may be associated with the properties of the crystals grown in slightly different crystallization conditions or by the erroneous location of the side chains in the structure at lower resolution.

A considerable difference of the structure at atomic resolution is that the number of water molecules located in this structure is 2.4 times larger than that in the structure solved at 1.95 Å resolution (437 and 181 molecules, respectively). It should be noted that 169 water molecules have virtually identical positions in both structures. In fact, only conserved highly ordered water molecules, which are located in the cavities of the protein globule and in the first solvent shell, were located in the structure at 1.95 Å resolution. The atomic resolution structure allowed the much more accurate and reliable determination of the solvent structure in the crystal structure of the MorFDH-NAD⁺-azide complex.

Due to the high accuracy of diffraction data, the atomic resolution allows the more precise determination of the alternative conformations of residues. In the structure of the MorFDH-NAD⁺-azide complex at 1.95 Å resolution, only five residues were found to adopt two different conformations. A considerable improvement of the quality of the electron density maps for the atomic resolution structure allowed the identification of ten such residues. The side chains of five lysine residues (Lys40, Lys61, Lys231, Lys383, and Lys395) and one glutamine residue (Glu397) were located only partially, which is evidence that these residues exist in several

Fig. 4. Structure of the active site of MorFDH with the bound NAD⁺ molecule and azide ion **a)** Binding of the nicotinamide moiety of the NAD⁺ molecule and the azide ion by the active site residues (hydrogen bonds are depicted as dotted lines). **b)** Electron density map with $2|F_o| - |F_c|$ coefficients countered at the 2.0 σ level for NAD⁺ and azide and the atomic numbering scheme for the nicotinamide moiety of NAD⁺



conformations. All residues adopting alternative conformations and with missing side chain atoms in the model are on the surface of the protein globule, which accounts for their high conformational flexibility.

A considerable advantage of atomic resolution diffraction data is that carbon, oxygen, and nitrogen atoms can be reliably differentiated, which allows the identification of rotational isomers of asparagine, glutamine, and histidine side chains. The rotational isomers of these residues are usually determined by analyzing the hydrogen bonds with adjacent atoms, and in some case definite conclusions cannot be drawn. The X-ray diffraction data at atomic resolution allows one to determine the isomers of asparagine and glutamine residues by analyzing the temperature factors of the side chain nitrogen and oxygen atoms. If the carboxamide group of asparagine or glutamine in the model is rotated by 180° with respect to the actual conformation in the protein crystal, the B factor of the nitrogen atom will be considerably lower, whereas the B factor of the oxygen atom will be considerably higher than the average temperature factors of covalently bonded atoms. A similar situation (higher B factors of the nitrogen atoms Nδ and Nε and lower B factors of the carbon atoms Cδ and Cε) is observed also for incorrectly rotated histidine side chains. In addition, the diffraction data at atomic resolution allows one to differentiate between carbon, oxygen, and nitrogen atoms based on electron density maps, because atoms with more electrons have higher electron density peaks (Fig. 2). To check for rotational isomers of asparagines, glutamine, and histidine, the analysis of electron density maps was accompanied by a separate refinement cycle using a modified model, in which all side chains of these residues were rotated by 180°. The rotational isomers of most residues in the structure of the MorFDH-NAD⁺-azide complex at 1.1 Å resolution were confirmed, and the side chains of the Gln66 and Asn135 residues were found to be rotated compared to their conformation in the 1.95 Å resolution structure. It was not always possible to unambiguously choose the isomer based on the analysis of the temperature factors, which is an indication that both rotational isomers are present in the structure.

The structure of the active site of MorFDH containing the bound NAD⁺ molecule and azide ion is depicted in Fig. 4. A detailed description of the structure of the FDH active site and the role of individual amino acid residues in the binding of the substrates and the catalysis can be found in the studies [12, 14]. The kinetic isotope effects study showed that the hydride ion transfer is the rate-limiting step of the enzymatic reaction, the transition state being the late one i.e., it is structurally more similar to the reaction products [20, 21]. The linear azide ion is isoelectronic with the reaction product CO₂, which is also linear. Moreover, azide ion is characterized by an extremely high binding constant to the holo-FDH, which is about five orders of magnitude higher than that of the formate ion [20]. Hence, the ternary MorFDH-NAD⁺-azide complex is considered as a stable analog of the transition state of the enzymatic reaction [20, 22].

The X-ray diffraction data at atomic resolution confirm that the carboxamide group of the cofactor in the active site of MorFDH is fixed in the *trans* conformation (the O7 atom is directed to the C4 atom) by hydrogen bonds with active site residues. According to quantum-mechanical calculations in the gas phase, the *trans* conformation is 2 kcal/mol less favorable than the *cis* conformation [23]. The analysis of the refined structure showed that the length of the C7-O7 double bond in the carboxamide group of the cofactor is 1.26 Å, which is 0.03 Å longer than the standard C-O double bond length in the carboxamide group (1.23 Å) [24]. At the same time, the length of the C3-C7 single bond is 1.47 Å, which is 0.03 Å smaller than the standard value (1.50 Å). Since the rms deviation for covalent bonds in the structure is 0.015 Å, these changes in the bond lengths between the atoms with relatively low temperature factors (O7 – 15.6 Å², C7 – 13.7 Å², C3 – 14.1 Å²) inside the protein molecule may be an indication of a change in the bond orders. These previously unknown characteristics of the MorFDH-NAD⁺-azide ternary complex may reflect important details of the structural organization of the transition state in the FDH-catalyzed reaction.

Previously, it has been shown by calculation methods, using combined molecular dynamics and quantum mechanics

studies, that the cofactor molecule adopting the energetically activated *trans* conformation in the transition state of the FDH-catalyzed reaction may take on the properties of the so-called bipolar form [4]. A similar hypothesis was also suggested on the basis of the kinetic isotope effects analysis [21]. The bipolar form is characterized by a higher negative charge on the O7 atom of the carboxamide group, the shorter C3-C7 bond, and the longer C7-O7 bond compared to those in the free cofactor molecule. These changes promote the increase in the partial positive charge on the C4 atom of the coenzyme, thus increasing its electrophilicity and facilitating the rate-limiting hydride ion transfer and the enzymatic reaction.

Due to the high accuracy of the atomic resolution diffraction data, the influence of standard stereochemistry on the atomic coordinates of the nicotinamide group of the cofactor can be weakened in the course of the crystallographic structure refinement. After a separate refinement cycle using the REFMAC program with weaker bond length constraints for the nicotinamide group (by increasing the rms bond length deviation from 0.02 Å to 0.20 Å), the C7-O7 bond length was 1.29 Å, which is 0.06 Å smaller than the standard value, and the C3-C7 bond length was 1.43 Å, which is 0.07 Å longer

than the standard value. These differences are only 1.3 times larger than the error in the bond lengths estimated as the sum of two mean-weighted errors in the atomic coordinates (Table 1). Nevertheless, our experimental data reveal a tendency toward a change in the covalent bond lengths, which corresponds to the fact that the nicotinamide group of the cofactor adopts the bipolar form in the transition state of the enzymatic reaction.

To sum up, the atomic resolution structure of FDH was determined for the first time, which provides a deeper insight into the details of the molecular mechanism of this enzyme and the hydride ion transfer in the active sites of NAD⁺-dependent dehydrogenases. ●

This work was supported by the Russian Federal Agency for Science and Innovation (federal contract 02.512.12.2002) and the Russian Foundation for Basic Research (grant № 08-04-00830-a). The authors thank the EMBL Hamburg Outstation for the allotted time on the synchrotron radiation source and Alexander Popov personally for assistance with diffraction data collection.

REFERENCES

1. Tishkov V.I., Popov V.O. // *Biochemistry Mosc.* 2004. V. 69. № 11. P. 1252.
2. Vinals C., Depiereux E., Feytmans E. // *Biochem. Biophys. Res. Commun.* 1993. V. 192. № 1. P. 182.
3. Bandaria J.N., Dutta S., Hill S. E., Kohen A., Cheatum C.M. // *J. Am. Chem. Soc.* 2008. V. 130. № 1. P. 22.
4. Castillo R., Oliva M., Marti S., Moliner V. // *J. Phys. Chem. B.* 2008. V. 112. № 32. P. 10012.
5. Torres R.A., Schitt B., Bruice T.C. // *J. Am. Chem. Soc.* 1999. V. 121. № 36. P. 8164.
6. Dauter Z., Lamzin V.S., Wilson K.S. // *Curr. Opin. Struct. Biol.* 1997. V. 7. № 5. P. 681.
7. Rubach J.K., Plapp B.V. // *Biochemistry.* 2003. V. 42. № 10. P. 2907.
8. Meijers R., Morris R.J., Adolph H.W., Merli A., Lamzin V.S., et al. // *J. Biol. Chem.* 2001. V. 276. № 12. P. 9316.
9. Meijers R., Adolph H.W., Dauter Z., Wilson K. S., Lamzin V.S., et al. // *Biochemistry.* 2007. V. 46. № 18. P. 5446.
10. Schlieben N.H., Niefind K., Muller J., Riebel B., Hummel W., et al. // *J. Mol. Biol.* 2005. V. 349. № 4. P. 801.
11. Cameron A., Read J., Tranter R., Winter V. J., Sessions R.B., et al. // *J. Biol. Chem.* 2004. V. 279. № 30. P. 31429.
12. Lamzin V.S., Dauter Z., Popov V.O., Harutyunyan E.H., Wilson K.S. // *J. Mol. Biol.* 1994. V. 236. № 3. P. 759.
13. Schirwitz K., Schmidt A., Lamzin V.S. // *Protein Sci.* 2007. V.16. № 6. P. 1146.
14. Shabalina I.G., Filippova E.V., Polyakov K.M., Sadykhov E.G., Safonova T.N., et al. // *Acta Crystallogr. D Biol. Crystallogr.* 2009. V. 65. № 12. P. 1315.
15. Otwinowski Z., Minor W. // *Methods in enzymology.* 1997. V. 276. P. 307.
16. Murshudov G.N., Vagin A.A., Dodson E.J. // *Acta Crystallogr. D Biol. Crystallogr.* 1997. V. 53. № 3. P. 240.
17. Emsley P., Cowtan K. // *Acta Crystallogr. D Biol. Crystallogr.* 2004. V. 60. № 12. P. 2126.
18. Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J. M. // *J. Appl. Cryst.* 1993. V. 26. № 2. P. 283.
19. Vaguine A.A., Richelle J., Wodak S.J. // *Acta Crystallogr. D Biol. Crystallogr.* 1999. V. 55. № 1. P. 191.
20. Blanchard J.S., Cleland W.W. // *Biochemistry.* 1980. V. 19. № 15. P. 3543.
21. Rotberg N.S., Cleland W.W. // *Biochemistry.* 1991. V. 30. № 16. P. 4068.
22. Popov V.O., Lamzin V.S. // *Biochem. J.* 1994. V. 301. № 3. P. 625.
23. Kahn K., Bruice T.C. // *J. Am. Chem. Soc.* 2001. V. 123. № 48. P. 11960.
24. Engh R.A., Huber R. // *Acta Crystallogr. A Found. Crystallogr.* 1991. V. 47. № 4. P. 392.

Mutation of Residue β F71 of *Escherichia coli* Penicillin Acylase Results in Enhanced Enantioselectivity and Improved Catalytic Properties

I.V. Shapovalova¹, W.B.L. Alkema², O.V. Jamskova¹, E. de Vries², D.T. Guranda¹, D.B. Janssen², V.K. Švedas^{1,*}

¹Belozersky Institute of Physicochemical Biology and Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Russia

²Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Netherlands

*E-mail: vytas@belozersky.msu.ru

ABSTRACT Residue phenylalanine 71 of the β -chain of penicillin acylase from *E. coli* is involved in substrate binding and chiral discrimination of its enantiomers. Different amino acid residues have been introduced at position β F71, and the mutants were studied with respect to their enantioselectivity and substrate specificity. Some mutants demonstrated remarkably improved catalytic activity. Moreover, mutation of β F71 residue allowed to enhance penicillin acylase enantioselectivity. The catalytic activity to the specific substrates was improved up to 36 times, most notably for K, R, and L mutants. Increased activity to a D-phenylglycine derivative – a valuable specificity improvement for biocatalytic synthesis of new penicillins and cephalosporins – was shown for β F71R and β F71L mutants. The synthetic capacity of penicillin acylase with 6-aminopenicillanic acid as an external nucleophile was especially sensitive to mutation of the β 71 residue in contrast to the synthesis with 7-aminodeacetoxycephalosporanic acid.

Keywords: penicillin acylase; β F71 mutants; enantioselectivity; improved catalysis;

Abbreviations: PA – penicillin acylase; Ntn – N-terminal nucleophile; WT – wild type.

INTRODUCTION

Penicillin acylase (PA) from *Escherichia coli* is probably the most studied enzyme among the group of Ntn-hydrolases. The enzymes of this superfamily are characterized by their unique catalytic mechanism in which the N-terminal amino acid residue (serine, cysteine or threonine) acts as a nucleophile to form an acylenzyme intermediate. PAs, as well as other Ntn-hydrolases, are activated by processing of a precursor protein to form the mature enzyme that has an $\alpha\beta\beta\alpha$ -fold topology of helices and strands around the active site [1]. Despite the high interest in the structure and biosynthesis of PAs [2-13], as well as in the new synthetic applications [14-20], little work has been done on the substrate specificity and especially the enantioselectivity of these enzymes [21, 22]. The reason for this is that the substrate specificity studies with phenylacetylated compounds are seriously complicated by a very strong competitive inhibition by the reaction product phenylacetic acid [23, 24]. PAs seem to possess a much wider substrate range than previously assumed (especially concerning the leaving group), and acylases of different origins may have quite different catalytic activities and enantioselectivities [24]. The substrate specificity and catalytic properties of PAs are of very high practical interest as this family of en-

zymes plays a decisive role in the biocatalytic preparation of semisynthetic β -lactam antibiotics [25].

A breakthrough in the elucidation of the catalytic mechanism of PA from *E. coli* was obtained when the X-ray crystallographic structure of the native enzyme [2, 10] and an enzyme-substrate complex was solved [12, 13]. This provided the first information about the residues involved in binding of the leaving group of the substrate. Afterwards, molecular modeling has helped to reveal the intimate details of the substrate binding in the penicillin acylase active center, especially concerning the binding pattern of the leaving group [26]. Additional contacts with residues β G385, β S386, and β N388 have been found, which were missing in X-ray structures. Based on structural information and molecular modeling, the key amino acid residues that control interactions between the enzyme and substrate have been identified. Such insight is of crucial importance for the rational redesign of basic biocatalytic properties, including substrate specificity, enantioselectivity, and the catalytic activity. Amino acid residue F71 of the β -chain of PA was demonstrated to be one of the principal amino acid residues interacting with the substrate's leaving group. Such a residue is an attractive target for mutagenesis aimed at modifying the catalytic performance of PA for the

biocatalytic modification of β -lactam antibiotics, chiral resolution of amino compounds, etc. In this paper, we report our results on the modification of the major catalytic properties of PA from *E. coli*: enhancement of the enantioselectivity and improvement of the catalytic activity. These changes were obtained by mutating a single amino acid residue (β F71) in the enzyme's active center.

MATERIALS AND METHODS

Mutagenesis of residue β F71 was achieved using three PCR reactions. The first reaction contained the BST_{FW} primer, 5'-CAGGGAAGAACC GGAAACTATTG-3', as the forward primer and the F71_{RV} reverse primer carrying the mutation in the codon for β F71 at the underlined position, 5'-AAAAATATCGACATCGTCGCCCAACCTGCCGT-3'. The second reaction contained F71_{FW} 5'-GGCGACGATGTCGATATTTTT-3' as the forward primer and NHE_{RV} 5'-CACTCCTGCCAATTTTTGGCCTTC-3' as the reverse primer. Products from both reactions were isolated from gel and used together in a third reaction without wild-type template, to which also the primers NHE_{RV} and BST_{FW} were added. The resulting products, carrying the mutations on position β F71, were cut with BstX1 and NheI and ligated into pEC, which was digested with the same enzymes. Ligation mixtures were used to transform competent *E. coli* cells as described [27]. PA genes from the resulting transformants were sequenced to confirm the mutations and the absence of second-site mutations that might have been introduced by PCR. Expression and purification of mutants, as well as WT PA, was done as described [12].

The concentration of active PA was determined by titration with phenylmethylsulfonylfluoride as described earlier [28]. Kinetic studies were performed by the initial rate analysis of corresponding reactions according to [24, 30]. Enantioselectivity (E) of the WT PA and its mutants was characterized as a ratio of the second order rate constants determined for the hydrolysis of the individual enantiomers $E = (k_{\text{cat}}/K_{\text{m}})^L / (k_{\text{cat}}/K_{\text{m}})^D$ as described earlier [21, 24].

Molecular dynamics simulations based on the available X-ray structure 1H2G of β F71 [29] were done with the Gromacs

package (www.gromacs.org) as described earlier [26]. Molecular docking was performed using Lead-Finder software from MolTech Ltd., Russia (www.moltech.ru).

RESULTS AND DISCUSSION

Amino acid residue F71 of the β -chain slightly moves upon binding of penicillin G in the active centre of PA and stays in close proximity to the β -lactam group, displaying van der Waals interactions [12]. Molecular modeling shows that residue β F71 plays an important role in the binding of other substrates, for example N-(3-carboxy-4-nitrophenyl)phenylacetamide (Fig.1a) and N-(2-hydroxy-4-nitrophenyl)phenylacetamide (Fig.2a). Moreover, mutation of this amino acid residue (Figs.1b and 2b) enhances the substrate's interaction with the oxyanion hole and its orientation. It therefore was expected that mutation of residue β 71 might influence the catalytic activity of PA and change enantioselectivity, as well as enzyme specificity to the leaving group. Consequently, β F71 mutants containing different functionalities in the side chain were prepared and studied with respect to their enantioselectivity and catalytic properties towards substrates of different chemical structures. All mutants were catalytically active, demonstrating that mutation of the β F71 residue did not influence the processing and correct folding of the mutant enzymes. Phenylmethylsulfonylfluoride, the well known irreversible inhibitor and active site titration reagent for penicillin acylases [24, 28], was used to determine the active site concentration of the mutant enzymes. Active site titration of WT PA and each mutant allowed to determine the absolute catalytic activity and accurately compare their catalytic properties. Kinetic studies included measuring the catalytic activity in the hydrolysis of phenylacetic acid derivatives with different, but structurally related leaving groups and α -amino substituent in the acyl group. The enantioselectivity of the mutants in the hydrolysis of N-phenylacetylated amino acids was used as a quantitative measure of chiral discrimination in the leaving group binding subsite. The most remarkable changes induced by the mutation of the single β F71 residue are discussed below.

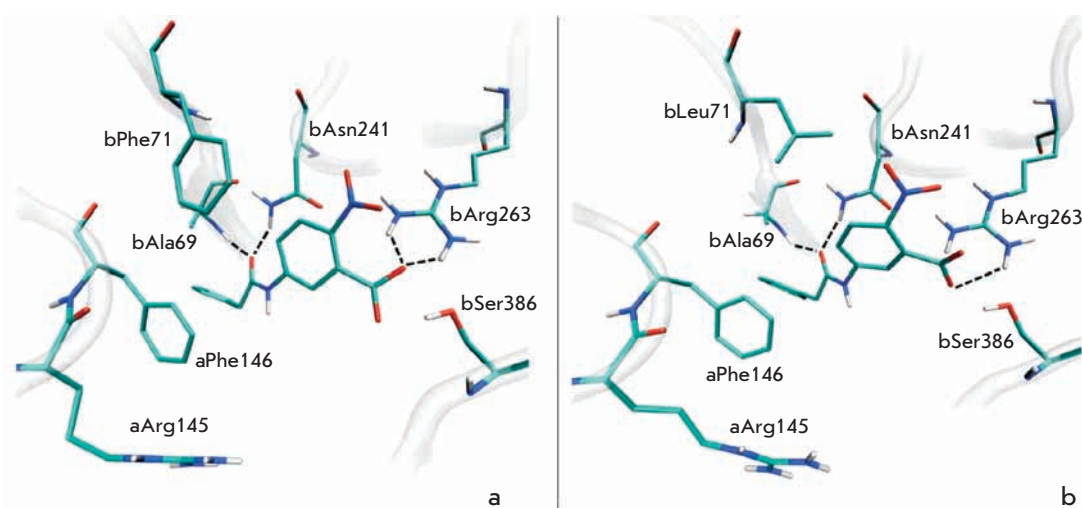


Fig.1. Molecular modeling of enzyme-substrate complexes for wild-type penicillin acylase (a) and its β F71L mutant (b) with N-(3-carboxy-4-nitrophenyl)phenylacetamide. Interactions of the substrate with oxyanion hole residues (β A69 and β N241) and residue β R263 are indicated by dotted lines

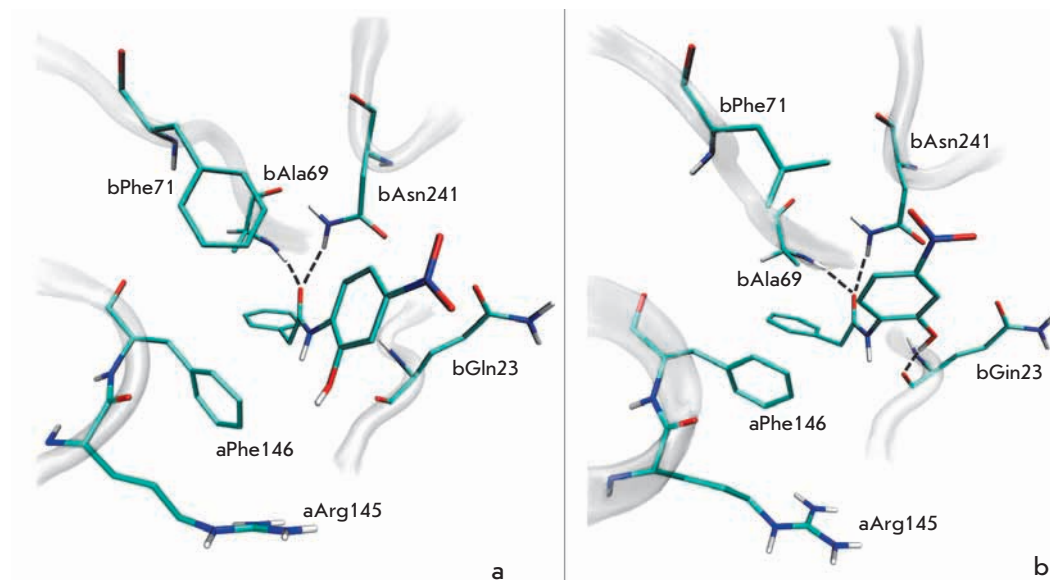


Fig.2. Molecular modeling of enzyme-substrate complexes for wild-type penicillin acylase (a) and its β F71L mutant (b) with N-(2-hydroxy-4-nitrophenyl)phenylacetamide. Interactions of the substrate with oxyanion hole residues (β A69 and β N241) and residue β Q23 are indicated by dotted lines

Fig.3. Structure of penicillin acylase substrates used to test catalytic activity of wild-type enzyme and its mutants

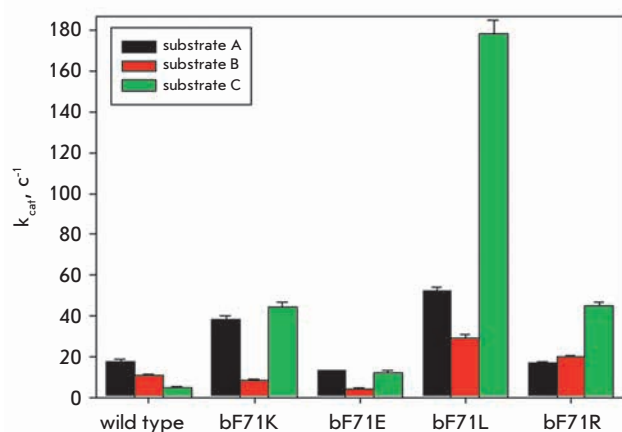
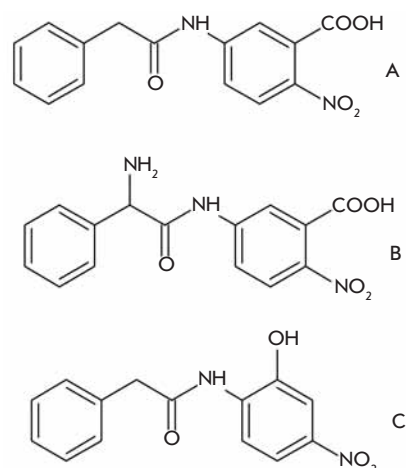


Fig.4. Catalytic activity of wild-type penicillin acylase and its β F71 mutants expressed as the value of the catalytic constants of the enzymatic hydrolysis of substrates presented in Fig.3

CATALYTIC EFFICIENCY OF MUTANTS

To compare the catalytic efficiency of wild-type penicillin acylase and its mutants, we measured the kinetics of the hydrolysis of chromogenic phenylacetic acid and D-phenylglycine derivatives (Fig.3). It can be seen that PA catalytic activity has changed considerably by mutating residue β F71. As expected, mutation influences the specificity to the leaving group (Fig.4). Most tested mutants demonstrated improved catalytic properties, but that was not the case with all tested substrates. The K, R, and L mutants showed a very large increase in catalytic activity only with some selected substrates, whereas the catalytic properties of the E mutants were improved to a lesser extent. β F71E mutant also had a decreased affinity to the substrates studied. Out of all tested compounds, the highest k_{cat} was found for the β F71L mutant with an improvement of more than 36 times compared to the WT PA.

In addition to the improved catalytic activity, the K, R, and L mutants also had substantially increased affinity to different substrates (data not shown), demonstrating that in general the phenylalanine at position 71 is not an optimal amino acid residue, neither for the catalytic activity nor for the affinity of PA to its substrates. It was observed that mutation of β F71 changes the properties of the acyl group binding subsite and improves PA specificity to D-phenylglycine derivatives, which are key acyl donors in enzymatic synthesis of the most important semisynthetic penicillins and cephalosporins, such as ampicillin and cephalixin. The search for new PAs that are more specific to D-phenylglycine derivatives has a long-lasting history [31-33], and there was no expectation that the specificity of PA from *E. coli* can be improved for this group of compounds. However, the β F71L mutant had increased specificity (k_{cat}/K_m value improved by a factor of 4.4) due to the increased catalytic activity and the improved affinity to the substrate. In fact, the nature of the "optimal" residue on position β 71 depends on the structure of the substrate converted. Our results clearly indicate that mutation of residue

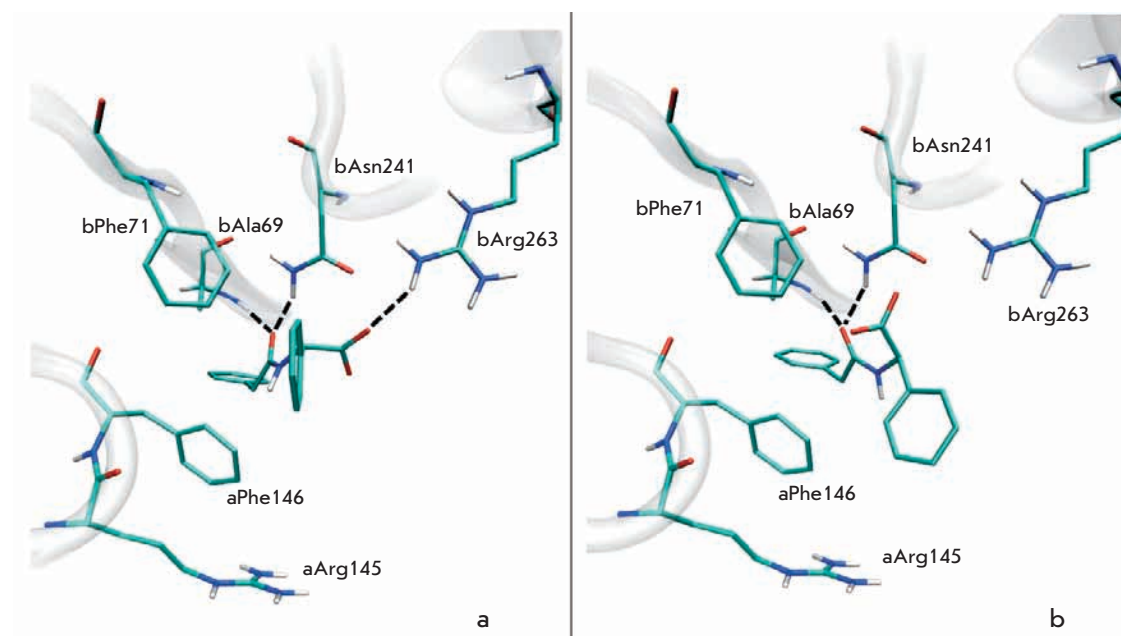


Fig. 5. Molecular modeling of enzyme-substrate complexes for wild-type penicillin acylase with N-phenylacetyl-L-phenylglycine (a) and N-phenylacetyl-D-phenylglycine (b). Interactions of the L-form and the D-form of the substrate with oxyanion hole residues (β A69 and β N241) are indicated by dotted lines. Interaction of the substrate with residue β R263 is observed only with the L-form of the substrate

β F71 may be used to design modified PAs with improved catalytic performance.

NUCLEOPHILE REACTIVITY STUDIES

The most straightforward quantitative way of studying the effect of mutations on the binding of β -lactam nuclei in the course of penicillin acylase-catalyzed synthesis of new β -lactam antibiotics would be to measure the corresponding binding constants. However, because of the complex kinetics of PA-catalyzed acyl transfer reactions [34] there is no proper method to determine the nucleophile binding constant to that particular subsite of the enzyme. Theoretical analysis has shown that an adequate way of characterizing the interaction of the nucleophile with the enzyme is to measure the ratio of initial synthesis and hydrolysis rates, the $(S/H)_{ini}$ ratio, as a function of the nucleophile concentration [35]. Kinetic experiments were, therefore, performed with two different side chain donors at a fixed concentration of the β -lactam antibiotic nucleus (6-APA or 7-ADCA). The results show that the nucleophile reactivity of 6-APA is much more sensitive to the amino acid located at position β 71 compared to that of 7-ADCA (Table 1). Obviously, mutation of β F71, especially substitution of the phenylalanine for a charged amino acid residue, leads to a substantial change of the β -lactam binding, and, as a result, the specificity of the leaving group binding subsite is modified.

ENANTIOSELECTIVITY

Since residue β F71 interacts with the leaving group of the substrate, mutating it should influence PA reactivity towards the stereochemical isomers of chiral substrates. Therefore, the enzyme's enantioselectivity value E (a ratio of k_{cat}/K_m values for conversion of L- and D-enantiomers of the substrate) in the hydrolysis of N-phenylacetylated amino acids was chosen as a quantitative parameter to describe the effect

of mutations. Although the WT PA exhibits already a very good chiral discrimination of N-phenylacetyl-phenylglycine enantiomers due to the orientation of the L-form of substrate in the active site by β R263 residue (Fig.5), it was still possible to improve it (Fig.6). The mutant that was superior in its catalytic activity, β F71L, displays a markedly decreased enantioselectivity by almost one order of magnitude, while the β F71E and β F71K mutants reveal a three and six-fold increase, respectively. The positive effect of the mutation (*i.e.* the higher enantioselectivity) originated in the decreased affinity for the slow-reacting D-enantiomer, while negative effects are mainly caused by the altered reactivity of both enantiomers (data not shown). Enantioselectivity, as well as

Table 1. Initial synthesis/hydrolysis rate ratios $(S/H)_{ini}$ for the ampicillin, amoxicillin and cephalixin synthesis catalyzed by wild-type penicillin acylase and its β F71 mutants using D-phenylglycine amide (D-PGA) and D-p-hydroxyphenylglycine amide (D-HPGA) as acyl donors

Enzyme	D-PGA/7-ADCA	D-PGA/6-APA	D-HPGA/6-APA
wild type	9 ± 2	1.9 ± 0.1	1.2 ± 0.04
bF71Y	7.5 ± 0.4	0.74 ± 0.01	0.61 ± 0.03
bF71L	4.9 ± 0.4	0.94 ± 0.04	0.86 ± 0.03
bF71W	2.9 ± 0.2	0.77 ± 0.05	0.62 ± 0.03
bF71R	1.8 ± 0.03	0.76 ± 0.01	0.45 ± 0.01
bF71K	1.2 ± 0.01	0.51 ± 0.01	0.46 ± 0.02

Experimental conditions: pH 7.0, 25°C, 0.05 M phosphate; D-PGA and D-HPGA concentration 0.015 M, 6-APA and 7-ADCA – 0.025 M.

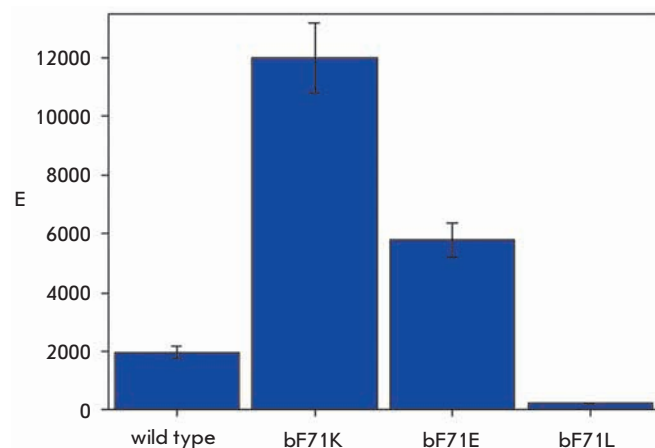


Fig. 6. Enantioselectivity (E) of penicillin acylase mutants in the hydrolysis of N-phenylacetyl-phenylglycine, expressed as the ratio $(k_{cat}/K_m)_L / (k_{cat}/K_m)_D$

catalytic activity studies, shows that PA catalytic properties can be finely tuned by mutating the β F71 residue for the specific substrate of interest.

CONCLUSIONS

Residue β F71 of PA from *E. coli* strongly influences the catalytic activity and chiral discrimination of the substrate's enantiomers. Depending on the structure of the substrate, β F71 mutations may deteriorate or improve the catalytic rate, enantioselectivity, or affinity. This indicates that a universal improvement of all major catalytic properties for biocatalytic application is hardly possible, and that the enzyme should be fine-tuned for its specific substrate of interest. The availability of structure/function insight concerning specific residues and molecular modeling can facilitate the design of a toolbox of dedicated PAs for different applications. ●

This work was supported by the Russian Federal Agency for Science and Innovation (federal contract 02.527.11.0001), the Seventh Framework Programme (FP7) of the European Commission and the Dutch Ministry of Economics.

REFERENCES

- Oinonen C., Rouvinen J. // *Protein Sci.*, 2000, V.9, P.2329-2337.
- Duggleby H.J., Tolley S.P., Hill C.P., Dodson E.J., Dodson G., Moody P.C.E. // *Nature*, 1995, V.373, P.264-268.
- Brannigan J.A., Dodson G., Duggleby H.J., Moody P.C., Smith J.L., Tomchik D. R., Murzin A.G. // *Nature*, 1995, V.378, P.416-419.
- Ispov M.N., Obmolova G., Butterworth S., Badet-Denisot M.A., Badet B., Polikarpov I., Littlechild J.A., Teplyakov A. // *Structure*, 1996, V.4, P.801-810.
- Groll M., Ditzel L., Lowe J., Stock D., Bochtler M., Bartunik H.D., Huber R. // *Nature*, 1997, V.386, P.463-471.
- Guo H.C., Xu Q., Buckley D., Guan C. // *J. Biol. Chem.*, 1998, V.273, P.20205-20212.
- McDonough M.A., Klei H.E., Kelly J.A. // *Protein Sci.*, 1999, V.8, P.1971-1981.
- Suresh C.G., Pundle A.V., Rao K.N., SivaRaman H., Brannigan J.A., McVey C.E., Verma C.S., Dauter Z., Dodson E.J., Dodson G.G. // *Nature Struct. Biol.*, 1999, V.6, P.414-416.
- Schumacher G., Sizmman D., Haug H., Buckel P., Bock A. // *Nucleic Acid Res.*, 1986, V.14, P.5713-5727.
- Done S.H., Brannigan J.A., Moody P.C.E., Hubbard R.E. // *J. Mol. Biol.*, 1998, V.284, P.463-475.
- Hewitt L., Kasche V., Lummer K., Lewis R.J., Murshudov G. N., Verma C. S., Dodson G.G., Wilson K.S. // *J. Mol. Biol.*, 2000, V.302, P.887-898.
- Alkema W.B.L., Hensgens C.M.H., Kroezinga E.H., De Vries E., Floris R., Van der Laan J.-M., Dijkstra B.W., Janssen D.B. // *Protein Eng.*, 2000, V.13, P.857-863.
- McVey C.E., Walsh M.A., Dodson G.G., Wilson K.S., Brannigan J.A. // *J. Mol. Biol.*, 2001, V.313, P.139-150.
- Baldaro E., D'Arrigo P., Pedrocchi-Fantoni G., Rosell C.M., Servi S., Tagliani A., Terreni M. // *Tetrahedron: Asymmetry*, 1993, V.4, P.1031-1034.
- Waldmann H., Sebastian D. // *Chem. Rev.*, 1994, V.94, P.911-937.
- Soloshonok V.A., Soloshonok I.V., Kukhar V.P., Jbvedas V.K. // *J. Org. Chem.*, 1998, V.63, P.1878-1884.
- Topgi R.S., Ng J.S., Landis B., Wang P., Behling J.R. // *Bioorg. Med. Chem.*, 1999, P.2221-2229.
- Basso A., De Martin L., Ebert C., Gardossi L., Linda P. // *J. Mol. Cat. B: Enzymatic* 2001, V.16, P.73-80.
- Guranda D.T., Van Langen L.M., Van Rantwijk F., Sheldon R.A., Jbvedas V.K. // *Tetrahedron: Asymmetry*, 2001, V.12, P.1645-1650.
- Guranda D. T., Khimiuk A.I., Van Langen L.M., Van Rantwijk F., Sheldon R. A., Svedas V.K. // *Tetrahedron: Asymmetry*, 2004, V.15, P.2901-2906.
- Jbvedas V.K., Savchenko M.V., Beltser A.I., Guranda D.F. // *Annals N.Y. Acad. Sci.*, 1996, V.799, P.659-669.
- Galunsky B., Lummer K., Kasche V. // *Monatsh. Chem.*, 2000, V.131, P.623-632.
- Berezin I.V., Klyosov A.A., Nys P.S., Savitskaya E.M., Jbvedas V.K. // *Antibiotiki (Rus.)*, 1974, V.19, P.880-887.
- Svedas V.K., Guranda D.T., Van Langen L.M., Van Rantwijk F., Sheldon R.A. // *FEBS Lett.*, 1997, V.417, P.414-418.
- Bruggink A., Roos E.C., De Vroom E. // *Org. Process Res. Dev.*, 1998, V.2, P.128-133.
- Chilov G.G., Stroganov O.V., Svedas V.K. // *Biochemistry (Moscow)*, 2008, V.73, №1, P.56-64.
- Sambrook J., Fritsch E. F., Maniatis T. // *Molecular cloning, a laboratory manual, 2nd edition*, Cold Spring Harbor Laboratory Press, New York, 1989.
- Jbvedas V.K., Margolin A.L., Sherstyuk S.F., Klyosov A.A., Berezin I.V. // *Bioorg. Khim.*, 1977, V.3, P.546-554.
- Morillas M., McVey C.E., Brannigan J.A., Ladurner A.G., Forney L.G., Virden L. // *Biochem. J.*, 2003, V.371, P.143-150.
- Youshko M.I., Shamolina T.A., Guranda D.F., Sinev A.V., Jbvedas V.K. // *Biochemistry (Moscow)*, 1998, V.63, №9, P.1104-1109.
- Takahashi T., Yamazaki Y., Kato K., Isona M. // *J. Am. Chem. Soc.*, 1972, V.94, P.4035-4037.
- Blinkovsky A.M., Markaryan A.N. // *Enzyme Microb. Technol.*, 1993, V.15, P.965-973.
- Polderman-Tijmes J.J., Jekel P.A., van Merode A., Floris T.A.G., van der Laan J.-M., Sonke T., Janssen D.B. // *Appl. Environ. Microbiol.* 2002, V.68, P.211-218.
- Youshko M.I., Švedas V.K. // *Biochemistry (Moscow)*, 2000, V.65, №12, P.1367-1375.
- Youshko M.I., Chilov G.G., Shcherbakova T.A., Svedas V.K. // *Biochim. Biophys. Acta: Proteins & Proteomics*, 2002, V.1599, №1-2, P.134-140.

Genotoxic Effects of Silver Nanoparticles on Mice *in Vivo*

C.G. Ordzhonikidze¹, L.K. Ramaiyya¹, E.M. Egorova², A.V. Rubanovich^{1,*}

¹ Vavilov Institute of General Genetics, Russian Academy of Sciences

² Institute of General Pathology and Pathophysiology, Russian Academy of Medical Sciences
Science-Technology Company "Nanomet"

*E-mail: rubanovich@vigg.ru

ABSTRACT The toxic and genotoxic effects of silver nanoparticles were studied on injected mice (BALB/c line) *in vivo*. A water solution of silver nanoparticles (SNP) with particle sizes of 9 ± 6 nm was obtained by means of the original method of biochemical synthesis. The effect of the SNP solution was compared to those of AOT (anionic surfactant used as SNP stabilizer) and silver nitrate (i.e. Ag^+ ions) introduced as water solutions. In studies of the toxic effects, the death of mice was registered 12-24 hours after injection only at two maximum doses of SNP (equivalent to 0.54 and 0.36 gAg/l). It is shown that the toxic effect decreases in the sequence SNP>AOT>>AgNO₃. The LE₅₀/30 values for SNP and AOT are equal to 0.30 ± 0.07 gAg/l and 13.3 ± 2.1 gAg/l, respectively. Genotoxic effects were assessed by the abnormal sperm heads test and neutral Comet assay. The frequencies of abnormal sperm heads (ASHs) did not differ after treatment by SNP and AOT, but both were significantly higher than those found with AgNO₃ and in control mice. Comet assay showed an increase of the DNA percentage in the comet tail in spleen cells after the injection of SNP and AOT in concentrations of $\frac{1}{2}$ LE₅₀/30. Tail DNA % was 32.8 ± 1.3 and $26.3 \pm 1.7\%$, respectively, vs $16.2 \pm 0.7\%$ for the untreated control. To sum up, these tests showed that the genotoxic effects of the SNP solution are associated with the presence of AOT rather than SNP.

Keywords: silver nanoparticles; mice; lethal effect; germ cells; primary DNA breaks.

Abbreviations: silver Nanoparticles – SNP; Aerosol –OT – AOT; lethal effect- LE; abnormal sperm heads – ASH.

INTRODUCTION

The use of nanotechnological products in human activities has been steadily increasing in recent years. Because of this, it is of vital importance to study the biological effect of various nanoparticles and nanocomposite materials, and especially their effects on animal and human organisms. The main issue is to elicit the toxicity of nanoparticles for humans and thus, the potential risk in the use of nanoparticle- and particle-based products.

Of utmost interest are metal nanoparticles and their biological effect, since these particles are most often used in novel products in various fields in manufacturing and medicine. During the last decade, abundant data has been obtained on both the positive (therapeutic effect) and negative (increase in the appearance of various diseases) effects of metal nanoparticles on living organisms [1, 2]. Silver nanoparticles are one of the most popular objects of research, since they have been actively used in the manufacturing of various consumer goods, such as dietary supplements, clothes, household appliances, toys, etc. Silver particles were mainly studied in bacteria, so as to elicit the particle's antimicrobial activity [3–5] or using *in vitro* cell cultures (for example [6]). There are also some data on the effect of silver nanoparticles on human fibroblasts [7]. Over all, data on the effect of silver and other metal nanoparticles are sparse. Up to now, there have been practically no data on the biological and genetic effect of silver nanoparticles introduced into a mammalian organism.

EXPERIMENTAL PROCEDURES

This work studied the toxic and genetic effects of silver nanoparticles (SNPs) on mice *in vivo*. We used water dispersions of SNPs obtained by biochemical synthesis [8]. The particle size was 9 ± 6 nm. It was previously shown that water dispersions

showed a pronounced antibacterial and antiviral effect [9], as well as a strong antibiotic effect on the slime mold *Physarum polycephalum* [10].

The initial concentration of dispersed SNPs (further named SNP preparation) was $5 \cdot 10^{-3}$ g-ion/l as calculated for Ag^+ . The effect of the SNP preparation was compared with that of the dispersant---an anionic surfactant (AOT), which acts as a stabilizer of the SNP preparation---and also with the effect of Ag^+ ions in equivalent concentrations. Aqueous solutions of AOT (initial concentration - 15 mM) and silver nitrate (initial concentration - 5 mM) were used.

The tests were conducted on male and female laboratory mice of the BALB/c line. The mice were 3-4 months of age and weighed 30–35 g. In order to determine the survival rate and lethal dose (the dose that causes the death of 50% of the subjects), the animals were divided into 4 groups of 16 mice each. The mice of the first group received a single intraperitoneal injection of the SNP solution (in distilled water, 0.2 ml). The concentration of SNP was varied by dilution of the initial preparation 0; 1.5; 2; 3; 5; 7; 10; and 100-fold. Thus, the SNP concentrations in the solutions were 5; 3.3; 2.5; 1.6; 1.0; 0.7; 0.5; and $0.05 \cdot (10^{-3}$ g-ion /l).

The second group was injected with AOT solutions in accordance with the dosages in Group 1. The equivalent concentrations of AOT for the above-mentioned SNP solutions are 15; 10; 7.5; 5.0; 3.0; 2.5; 1.5; and 0.15 mM.

Mice from Group 3 were injected with an aqueous solution of AgNO₃ at 5; 0.5; and 0.05 mM concentrations, which is equivalent to the 0; 10; and 100-fold SNP dilutions.

Group 4 was a control group and was injected with 0.2 ml of distilled water.

The lethal effect of the injected solutions was determined according to the standard protocol. All the injected mice were

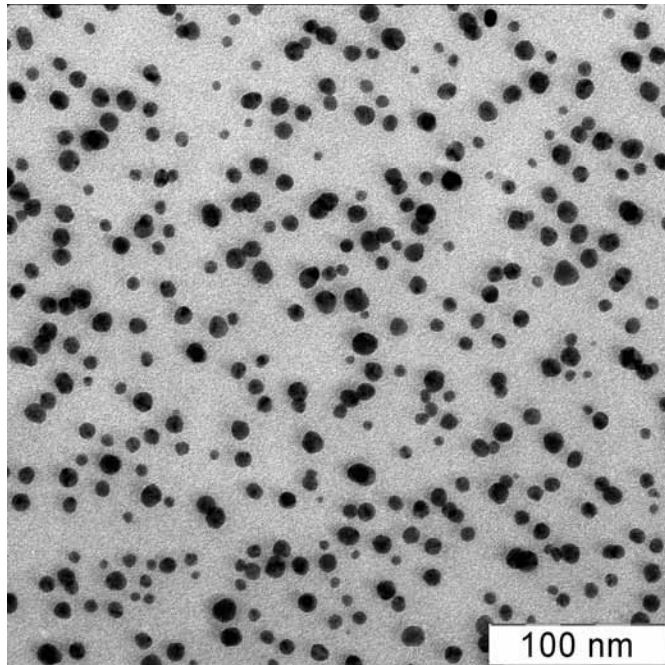


Fig. 1. Electronic microphotograph of silver nanoparticles

kept in an animal facility for 30 days, and animal death was monitored daily. Observation of the physical condition of the animals showed that during the first hours after injection, mice from Group 1 with the highest dosages (5 and $3.3 \cdot 10^{-3}$ g-ion/l) of SNPs exhibited a decreased motor activity and convulsions, followed by paralysis of the hind limbs. The animals died 12–24 hours after injection of the preparation. The other mice from this group with a lower dosage of the drug showed less prominent motor depression and toxic syndrome during the first few hours after injection, and the overall condition of these mice did not differ noticeably from the control animals. Mice in this group also died despite lowered concentrations: however, the survival time of these animals was greater (9–10 days) than that of the animals that received higher doses (1–3 days).

In Group 2, mice died only at high doses of AOT (15; 10; and 7.5 mM). Lower concentrations of AOT did not cause animal death. Groups 3 and 4 did not exhibit any animal death during the 30 days of observation.

The relationship between mouse death and the concentration of the injected solution is presented in Fig. 1.

The data presented in Fig. 1 show that a mouse's survival rate decreased monotonously as the concentration of the SNP solution was increased, with the exception of the lowest concentration ($0.05 \cdot 10^{-3}$ g-ion/l – 100-fold dilution). Regressive analysis yielded $LD_{50/30}$ figures for the preparations studied. $LD_{50/30}$ for the SNP preparation was $(2.74 \pm 0.67) \times 10^{-3}$ g-ion/l at an AOT concentration of 8.3 mM (1.8-fold dilution). Extrapolation of these data shows that $LD_{50/30}$ for AOT without silver nanoparticles is 29.9 ± 4.8 mM. Thus, we can assume that the toxic effect of the dispersed silver nanoparticles is 3.6-fold stronger than that of the dispersant. The absence of

animal death in the group that received silver nitrate injections shows that, of the three studied reagents, Ag^+ ions have the lowest toxic effect. Thus, the toxic effect decreases in the following sequence: SNP > AOT >> $AgNO_3$.

In order to study the effect of the SNP preparation on mammalian reproductive cells, we monitored the number of emerging sperm with anomalous head morphology (AHS) 21 days after injection of the preparation. This method allows to detect the deleterious effects of a preparation on reproductive cells in the early premeiotic stage of gametogenesis, namely the pachytene of the first meiotic division. It is assumed that AHS emergence is caused either by gross chromosomal aberrations, such as translocations, or by point mutations and small deletions or, alternatively, by somatic damage. It was shown earlier that certain physical (ionizing radiation, microwaves) and chemical factors (cyclophosphamide, cadmium chloride, zinc chloride, etc.) increased the AHS index and also caused a decrease in testicular mass, an increase in pre-implantation pregnancy loss and a decrease of effective copulations caused by damage to premeiotic male reproductive cells in mice. These data point not only to a mutagenic, but also to a cytolytic and/or cytotoxic effect of these factors [11]. The method for determining the AHS index is fairly simple; it does not require a large number of animals and can be used as an estimate of the mutagenic effect of various drugs [12].

Figure 2A shows AHS index data obtained after injection of AOT and SNP solutions into male mice. The AOT and SNP concentrations were 5.0 mM and $1.6 \cdot 10^{-3}$ g-ion/l, respectively (the initial samples were diluted 3-fold). The data presented in Fig. 2A indicate that the deleterious effect of SNPs and AOT was greater (approximately 1.5 fold) than that of the control injections. Notably, there was no marked difference between the effects of AOT and SNPs.

The same animals were also tested for primary DNA damage using neutral gel-electrophoresis of individual cells (DNA-comets). Damaged DNA (DNA with single and double strand breaks) is a sign of oxidative stress and cell death. The DNA-comet method can provide some information about

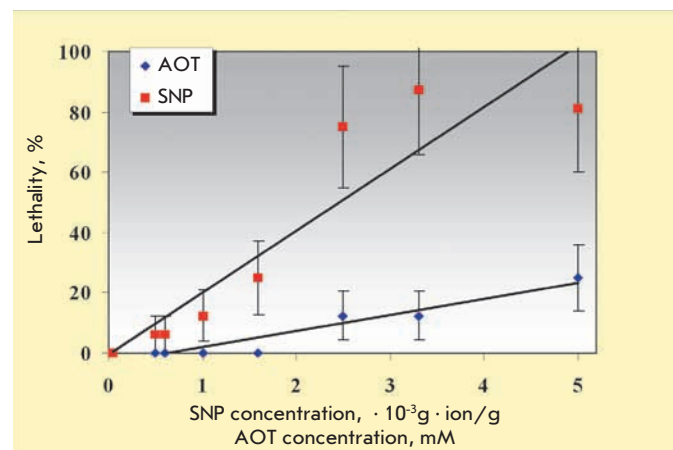


Fig. 2. Death rate dependence on the concentrations of SNP and AOT for injected mice

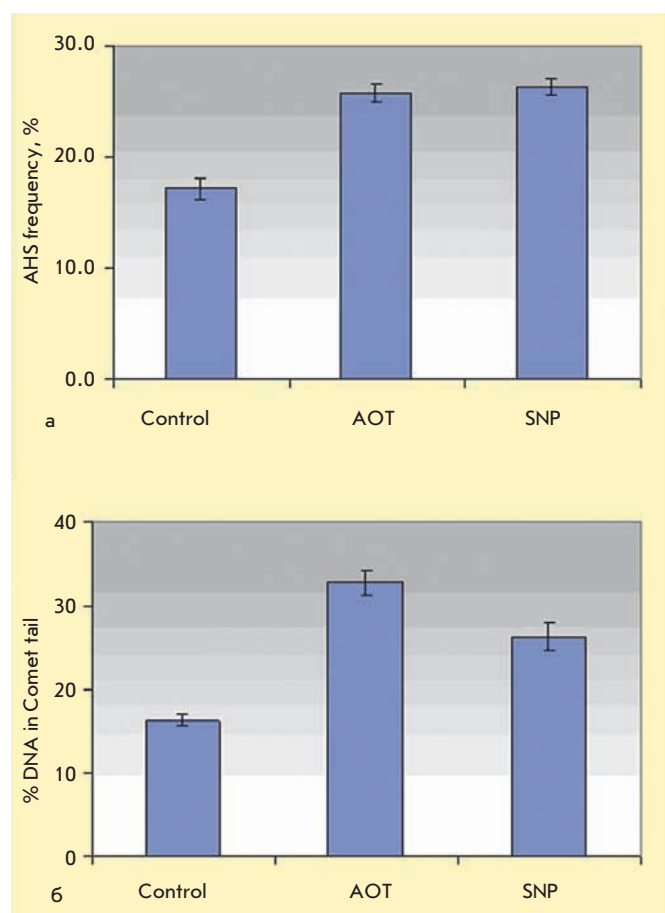


Fig. 3. AHS frequency (a) and % of DNA in Comet tail (b) after injection of SNP and AOT solutions (0.17g/l and 2.2 g/l, respectively; dilution in 3)

induced mutagenesis by visualizing the relative amount of damaged DNA. This electrophoretic method is based on the different mobility of the whole DNA and its possible degradation fragments obtained from lysed cells trapped in an agarose gel.

This test was performed according to the standard protocol [13]. We chose the spleen as a target organ because of

its specific function in the circulatory system (lymphocytes, monocytes, and macrophages that are accumulated in the white and red pulp can easily be damaged by nanoparticles).

Using the DNA-comet method, we measured the level of DNA damage after injecting the SNP preparation and the AOT solution (concentrations $1.6 \cdot 10^{-3}$ g-ion/l and 5 mM, respectively). We did not detect any increase of the DNA percentage in the comet tails up to 48 hours after the injections. Figure 2B shows the data averaged for 14 mice and 7 time points (3, 5, 7, 9, 12, 24, and 48 hours). The figure shows that the AOT dispersant had a damaging effect on DNA (the portion of DNA that exited the cell into the “comet tail,” was 32.8 ± 1.5 %), which is greater than the respective value for SNPs (26.3 ± 1.7 %). Both of these values exceeded the control value (16.2 ± 0.7 %).

It was previously shown for the mold *Physarum polycephalum* [10] and *E.coli* bacterium [9] that an aqueous dispersion of silver nanoparticles had a more pronounced toxic effect than equivalent concentrations of silver ions or solutions of the surfactant, which was a component of the SNP preparation.

CONCLUSION

This work shows for the first time that dispersed silver nanoparticles obtained via biochemical synthesis have a lethal effect on mammalian organisms when injected *in vivo*. The lethal effect of the dispersed nanoparticles is approximately 4 times greater than that of the AOT dispersant alone (See Fig. 1), while injection of equivalent doses of silver ions is followed by 100% survival of the tested animals. Using a calculation taking into account the mouse’s body weight, we estimate that the dose of silver nanoparticles that causes 50% lethality is 1.9×10^{-6} mg per gram of body mass. ●

This work was supported by funding provided by “Nanomet” in accordance with the R&D (contract № 8418-16/09). The authors thank corresponding RAS member prof. N.K. Yankovskiy (IOGEN RAS) for thorough discussion of our results and V.S. Lysenkova, senior technician of the ecological genetics laboratory of IOGEN RAS for technical assistance in this project.

REFERENCES

- Salata O.V. // *Journal of Nanobiotechnology* 2004, 2:3doi:10.1186/1477-3155-2-3.
- Oberdorster G., Oberdorster E., Oberdorster J. // *Environmental health perspectives*. 2005. V.113. N 7. P.823.
- Loponov A.N. *Silver*. 2005, St. Petersburg: Agat, 399 p.
- Neal A.L. // *Ecotoxicology*. 2008. V.17. P.362.
- Sondi L., Salopek-Sondi B. // *J. Colloid Interface Sci.*, 2004, v.275, p.177.
- Braydich-Stolle L., Hussain S., Schlager J.J. et al. // *Toxicological sciences*. 2005.V.88. Iss.2. P.412.
- Hui Yang, Chao Liu, Danfeng Yang, Yuashan Zhang and Zhuge Xi. // *J. Appl. Toxicol.*, 2009, 29, p.69-78.
- Egorova E.M., Revina A.A., Kondratieva V.S. Patent RF №2147487.
- Egorova E.M., Revina A.A. // *Colloids and Surfaces A*. 2000. V.168. P.87.
- Egorova E.M., Revina A.A. // *Colloid Journal* (Russia) 2002. V.64. P.334
- Egorova E.M., Revina A.A., Rostovschikova T.N., Kiseleva O.I. // *Vestnik of Moscow State University. Ser.2 Chemistry*. 2001. V.42. P.332
- Egorova E.M., Revina A.A. // *Journal of Physical Chemistry* (Russia). 2003. V.77. P.1683.
- Egorova E.M., Revina A.A., Rumyantsev B.V. et al. // *Journal of Applied Chemistry* (Russia). 2002. V.75. P.1620.
- Egorova E.M. // *Nanotechnics* (Russia). 2004. №1. C.15.
- Matveeva N.B., Egorova E.M., Beylina S.I., Lednev V.V. // *Biophysics*. 2006. V.51. P.758.
- Pomerantseva M.D., Ramayya L.K., Vilkina G.A. // *Genetica*, 1980, T.XYI, N8, p.1397-1403.
- Bruce W.R., Furrer R., Wyrobec A.J. // *Mutat.Res*. 1974, V.23, N3, p. 381.
- Olive P.L., Banath J.P. // *Int.J.Radiat.Biol.*, 1993, V.64, N4, p. 349-358.

Combining Two Technologies for Full Genome Sequencing of Human

K.G. Skryabin¹, E.B. Prokhortchouk^{1*}, A.M. Mazur¹, E.S. Boulygina¹, S.V. Tsygankova¹, A.V. Nedoluzhko¹, S.M. Rastorguev¹, V.B. Matveev², N.N. Chekanov³, D.A. Goranskaya³, A.B. Teslyuk¹, N.M. Gruzdeva¹, V.E. Velikhov¹, D.G. Zaridze², M.V. Kovalchuk¹

¹ Russian Research Centre Kurchatov Institute

² Institute of Carcinogenesis, Blokhin Cancer Research Center, Russian Academy of Medical Sciences

³ Bioengineering Center, Russian Academy of Sciences

* E-mail: prokhortchouk@biengi.ac.ru

ABSTRACT At present, the new technologies of DNA sequencing are rapidly developing allowing quick and efficient characterisation of organisms at the level of the genome structure. In this study, the whole genome sequencing of a human (Russian man) was performed using two technologies currently present on the market - Sequencing by Oligonucleotide Ligation and Detection (SOLiD™) (Applied Biosystems) and sequencing technologies of molecular clusters using fluorescently labeled precursors (Illumina). The total number of generated data resulted in 108.3 billion base pairs (60.2 billion from Illumina technology and 48.1 billion from SOLiD technology). Statistics performed on reads generated by GAI and SOLiD showed that they covered 75% and 96% of the genome respectively. Short polymorphic regions were detected with comparable accuracy however, the absolute amount of them revealed by SOLiD was several times less than by GAI. Optimal algorithm for using the latest methods of sequencing was established for the analysis of individual human genomes. The study is the first Russian effort towards whole human genome sequencing.

Keywords: human genome, sequencing technology, single-nucleotide polymorphism

Abbreviations: Indel SNP – insertion/deletion type single-nucleotide polymorphism.

INTRODUCTION

Progress in the development of novel DNA sequencing technologies allowing rapid and accurate determination of the distinctive features of an individual at the level of the structure of his genome has made genomics one of the most rapidly developing scientific disciplines. At present, three advanced technologies of DNA sequencing are in use: pyrosequencing based on the “sequencing-by-synthesis” principle and commercialized in a next-generation Roche GS-FLX capillary genome sequencing system (454 Life Science Inc./Roche), cyclic sequencing by oligonucleotide ligation and detection (SOLiD™, Applied Biosystems), and high-throughput DNA molecular cluster sequencing-by-synthesis using proprietary fluorescently labeled oligonucleotides (Illumina GAI Genome Analyzer, previously known as SOLEXA). These platforms have already demonstrated their significant worth: in the last two years, five new genomes have been read – in addition to the reference genome sequence that was determined by several leading research groups at institutes in the U.S., United Kingdom, and Canada in the course of 10 years and at a cost of 3 bln dollars [1] – such as the genomes of outstanding biologists of our times [2, 3], that of a Nigerian man [4, 5], a Chinese [6], and a Korean [7], not to mention thousands of other eukaryotic and prokaryotic species sequenced [8]. All these projects only became possible thanks to advanced technologies allowing cost-effective and high-throughput sequencing. We can expect an exponential increase in the amount of completely sequenced genomes in the near future; in particular, the recently launched 1,000 Genomes Project brings together

multidisciplinary research teams from institutes around the world, including the United Kingdom, China, and the United States (<http://www.1000genomes.org>). Yet, despite considerable progress in the reading of DNA, the sequencing of large genomes, such as the human genome, has yet to become trivial. At present, there is no standardized approach to the analysis of these genomes, and how the performance of the above-mentioned technologies is assessed remains far from objective.

In the present work, we were the first in Russia to perform the full-genome sequencing of a Russian male using two advanced DNA sequencing technologies: Sequencing by Oligonucleotide Ligation and Detection (SOLiD™, Applied Biosystems) and DNA molecular-cluster sequencing with fluorescently labeled precursors (GAI, Illumina). In this work, we intended to optimize the algorithms of full-genome sequencing data acquisition, processing, and representation.

The general principles and terms governing large-scale advanced sequencing are as follows: initially, genomic DNA is subjected to fragmentation to obtain fairly short strands (200 to 1,000 basepairs (bp)); then these fragments (*shotguns* – by analogy with the quasi-random firing pattern of a shotgun) are linked with oligonucleotide adapters, followed by PCR-amplification using specific enzymes and cloning for creating shotgun libraries. The library protocols for the GAI and SOLiD™ systems are described on the manufacturers' websites. Further procedures result in primary nucleotide sequences from each of the two ends of a DNA fragment from the library. These DNA sequences are called *reads*. The length of a read differs

depending on the platform and is 36 nucleotides for GAI and 25 nucleotides for SOLiD™. Thus, each DNA fragment in the library is characterized by two reads with length and direction depending on the technological platform used. Then, the reads obtained by the sequencing of the shotgun library entries from the two ends (*paired-end reads*) are mapped to the human reference genome (hg18). This process, called *read mapping*, specifies the coordinate of the read in the genome. The mapping allows to draw the histograms of coverage and distance between paired-end reads and also to identify single-nucleotide polymorphisms (SNPs) and short insertions/deletions (indels). Moreover, the distance between the reads and their orientation might provide information on more considerable structural rearrangements in the studied genome. For instance, the essentially greater distance between the mapped reads compared to the overall length of the DNA fragments used for construction of the library is indicative of deletion between the reads in the analyzed genome, as compared to the reference one. Similarly, improper orientation of the reads that is discordant with the library design may suggest inversions in the analyzed region. Thus, large-scale genome sequencing using advanced technologies allows to determine short polymorphic loci and detect regions likely associated with great genomic abnormalities. However, the latter can be described only by *de novo* assembly of the reads into extended *contigs*, which was beyond this study.

MATERIALS AND METHODS

DNA sample. The male whose genome was to be sequenced was chosen from data contained in the Principal Component Analysis (PCA) of ethnic groups of the Russian Federation. One thousand three hundred eighty-two individuals representing 32 ethnic groups were genotyped by no less than 300,000 SNPs using high-density DNA microarrays. The group of ethnic Russians comprised 285 samples provided by Prof. D. G. Zaridze. The male whose genome was chosen for sequencing was characterized by principal components positioning him within the group of ethnic Russians on the 2D plot of the first vs. the second principal component (PC1-PC2). This PC1-PC2 area had no intersections with areas of other closely related ethnic groups (data in press).

Sample preparation. Genomic DNA was isolated from the arterial blood lymphocytes of a Russian male (patient N with renal cell carcinoma, see above, from the Blokhin Cancer Research Center, RAMS). The blood sample was collected with informed consent. The DNA was fragmented on a HydroShear® DNA shearing device (Genomic Solutions®, USA) to the average fragment size of 500 to 1,000 bp. Construction of genome libraries and all subsequent manipulations were carried out using supplemental reagent kits in accordance with the manufacturers' protocols. Both genome libraries were suitable for paired-end reading. Following adapter ligation, the genome library constructed for an Illumina Genome Analyzer II (GAI) DNA Sequencing Platform (Illumina, USA) was divided into two parts: the first was frozen, and the second one was used for PCR amplification (hereinafter Amplification No.1). When sequencing of this library was accomplished in nine flow cells, the second part was thawed and also subjected to PCR (Amplification No.2). These samples were sequenced in five flow cells.

The same fragmented DNA was also implicated in the construction of genome libraries suitable for paired-end reading on a SOLiD™ v.2 System (Applied Biosystems, USA) (hereinafter SOLiD). Following emulsion PCR in the reaction product (DNA immobilized on magnetic beads) was applied onto flow cells in which the ligase chain reaction was carried out. In each sequencing cycle, the enzyme (ligase) attached a fluorescently labeled oligonucleotide to the 5'-end of the substrate complex. Once identified, the fluorescence moiety was removed to produce the substrate complex elongated by five nucleotides. Over all, nine flow cells were used for sequencing.

Sequencing. Decoding of genetic information was performed using two platforms commercialized by Illumina, Inc., and Applied Biosystems, Inc. The first platform uses detection of fluorescently labeled nucleotides incorporated *in situ* into surface molecular clusters. This technology is embodied into an Illumina GAI DNA sequencing platform. The length of the reads is 36 nucleotides from each end, and 14 flow cells were used. The second technology, based on ligase chain reaction, is embodied into a SOLiD sequencing system. The length of the reads is 25 nucleotides from each end, and nine flow cells were used.

Genotyping of unfragmented genomic DNA was carried out using the Infinium technology on Illumina Human610-Quad BeadChips, according to the manufacturer's recommendations. The chips were scanned using an Illumina iScan System. Quality control has shown high conformity with the control parameters (call rate 99.7 %). Over all, allelic variants of 588,702 SNPs were reliably found. The SNP list is presented on the manufacturer's site (http://www.illumina.com/documents/products/marker_lists/marker_list_human660W_quad.zip).

GAI data processing. Analysis of the obtained images and their conversion to the DNA sequence was performed using the Illumina Genome Analyzer Pipeline v.1.4.0 software suite. The mapping of sequences to the human reference genome (hg18) was performed using the Eland program supplied with Genome Analyzer Pipeline and SOAPaligner/soap2 v.2.20 alignment program developed at the Beijing Genomics Institute (<http://soap.genomics.org.cn/>) (hereinafter SOAP, Short Oligonucleotide Analysis Package). The library of paired-end reads obtained from GAI is available at <http://www.rus-siangenome.ru>, the web site of the Project. This library allows to examine the localization and direction of the reads in genome browsers, such as the UCSC Genome browser or Ensembl Genome Browser available on demand. Nucleotide mismatches and short indels in the studied genome compared with the reference genome were calculated using the SOAPaligner/soap2.

SOLiD data processing. Pre-processing of SOLiD data for mapping was carried out using SOLiD software (provided as a component of the SOLiD System). Mapping of the sequences was carried out in initial color space using a SOLiD System Analysis Pipeline Tool (Corona Lite) v.4.0r2.0 and, following conversion from the color space to the FASTQ format, using a Burrows-Wheeler Aligner (BWA) v.0.5.1 installed on a computer cluster mounted at the RRC Kurchatov Institute. From 20 to 250 processing units were implicated in the data processing, depending on calculation complexity. Barring

Table 1. General statistics on the analysis performed. Percentages are in compliance with total read counts for either GAI (orange) or SOLiD (blue)

		GAI (SOAP)	SOLiD (CoronaLite)
Total nucleotides		60,290,962,560	48,151,787,550
Total reads		1,674,748,960	1,926,071,502
Total not mapped reads		17.41%	32.65%
Total single-end reads		13.99%	48.53%
Unique		5.75%	31.36%
Multiple		8.24%	17.17%
Mismatches	0	54.85%	55.67%
	1	16.95%	23.13%
	2	28.20%	21.20%
Total paired-end reads		68.60%	18.82%
Unique		51.16%	12.20%
Multiple		17.44%	6.62%
Mismatches	0	74.29%	28.14%
	1	16.32%	21.86%
	2	9.39%	50.00%
		SOAP	CoronaLite
Insertions (nucleotides)	Total	0.93%	-
	1	0.58%	-
	2	0.19%	-
	3	0.08%	-
	4	0.08%	-
Deletions (nucleotides)	Total	0.81%	-
	1	0.50%	-
	2	0.17%	-
	3	0.07%	-
	4	0.07%	-

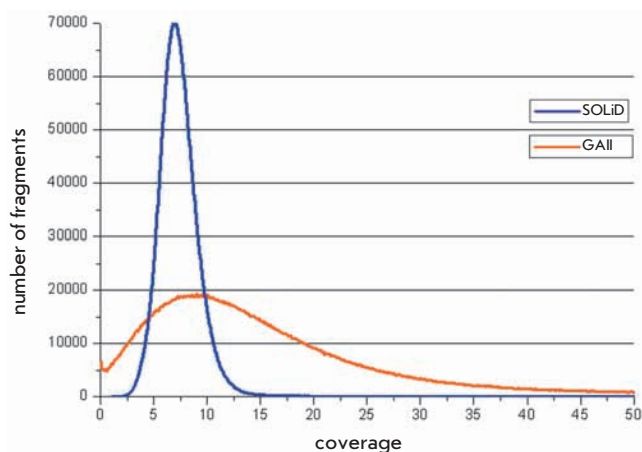


Fig. 1. Genome coverage depth; GAI and SOLiD data are shown in orange and blue colours respectively

BWA, all calculations, both for GAI and SOLiD data, were carried out on dedicated computer of the above-mentioned computer cluster. Similarly to the GAI data, the data from SOLiD can be found on the site of the Project, <http://www.russiangenome.ru>, and viewed in genome browser.

Original processing methods. Calculations of coverage density, distances between the reads in paired-end libraries, and sequencing errors compared with those of genotyping on SNP microarrays were carried out using the authors' codes written on Perl. These codes are available on demand.

RESULTS

Processing of GAI and SOLiD datasets.

The total amount of data from Amplification No.1 and Amplification No.2 genome libraries, which passed the internal filters, was 60.2 Gbp or 1,674,748,960 reads for GAI and 48.1 Gbp or 1,926,071,502 reads for SOLiD. The mapping of reads to the human reference genome allowed to draw the coverage histogram (Fig. 1). To do that, the genome was broken into tandem segments, 500 bp each, and the number of reads per segment was calculated from the mapping data (using Eland for GAI and BWA for SOLiD). This number was multiplied by the read's length (36 for GAI and 25 for SOLiD) and normalized by the fragment's length (500). The histogram of the GAI dataset has the shape of a Maxwell distribution with a peak at eightfold multiplicity and a tail shifted to tens of thousands of bp. This extremely high coverage largely corresponds to centromeric regions. Nucleotide-by-nucleotide analysis has shown that the GAI dataset covers, at least once, 2,033,881,571 nucleotides or 66.03% of the genome. General statistics of read mapping is presented in Table 1. It is worth noting that the amount of unmapped reads in the SOLiD dataset compared with the GAI one is almost twofold higher (32.65%).

The score of nucleotide mismatches and short indels was only performed for unique alignments. Alignments were assumed as unique when mapped to the reference genome only once and, hence, characterized by unique location. About 82.6% of 1.67 billion reads corresponded (with minimal mismatch) to the reference genome sequences. The other 17.4% of reads were classified as unmapped. Random sampling inspection of 164 reads from this group has shown that none of them can be mapped to the reference genome with minimum distortion (less than two misalignments, indels no longer than four nucleotides). Over all, 13 out of 164 reads were attributed to other genomes: most to simian genomes, two reads to the genome of *Mus musculus*, and two reads to the genomes of *Danio rerio* and *E.coli*, respectively. However, all these fragments were not completely consistent with the DNA sequences of these organisms, whereas the major portion of unmapped reads showed complete, if rather short (no more than 25 nucleotides), coincidence with various human genome sequences.

Checksum of matches between allelic SNP variants detected by sequencing and by genotyping on DNA microarrays.

Using DNA microarrays, allelic variants of 588,705 SNPs with distinct locations were determined in the studied genome. For evaluation of the accuracy of SNP determination by the sequencing method all reads encompassing the

Table 2. Comparison of homozygous SNPs detected by genotyping and inferred from sequencing data

		GAI (Eland)	SOLiD (BWA)	Both platforms (Eland or BWA)
Total homozygous SNP's on microchip		409,760		
Tests performing	coverage (≥ 1)	302,919	394,373	404,564
	coverage (≥ 5)	250,353	238,130	349,309
	coverage (≥ 10)	194,016	74,902	270,890
		After performing all tests		
Coverage ≥ 5	Amount	242,201	218,974	331,873
	Percentage	96.74%	91.96%	95.01%
Coverage ≥ 10	Amount	188,708	71,999	261,537
	Percentage	97.26%	96.12%	96.55%

positions of microarray SNPs were selected (mSNP). The number of SNPs to which at least one read from the GAI and SOLiD datasets was mapped was 581,596 (98.8% of the total number of mSNPs). The reads from the GAI platform involved 437,056 SNPs (74% of mSNPs); those from SOLiD, 566,952 (96,3%). Sequencing of the genome libraries obtained at the stages Amplification No.1 and Amplification No.2 on the GAI platform resulted in reads overlapping 333,647 (56.7%) and 372,483 (55.6%) mSNPs, respectively (Fig. 2). Homozygous mSNPs from Illumina 660W chip (totally 409,760) were only chosen to estimate the accuracy of allelic variant prediction by sequencing. Also, only mSNPs, to which no less than 1.5 or 10 reads were mapped, were chosen. Their numbers are given in corresponding rows (coverage (≥ 1), coverage (≥ 5), and coverage (≥ 10)) in the Table 2. The same table mSNP percentages and the numbers of allelic variants that are closely similar – on different platforms and at different coverage levels – to the ones predicted by genotyping on DNA microarrays. We have found that sequencing allows to determine about 81% of mSNPs with an accuracy of no less than 95 % (Table 2, grey column “Eland or BWA,” coverage ≥ 5).

Analysis of information on paired-end reads.

The data on the relative position of paired-end reads on the reference genome was used for estimating the number of structural rearrangements in the studied genome. The plot of the number of paired-end reads mapped to hg18 versus the length of the reference sequence between the ends is shown on Figure 3. One can see on the figure that the lines significantly differ in shape between the two platforms, which can reflect the essential difference in the protocols used for construction of paired-end libraries. It is noteworthy that the GAI library distribution pattern of the distance between the paired ends shows local peaks at approximately 70 bp and 300 bp, as well as a high peak at 700 bp. These peaks are very likely associated with the reads mapped to DNA repeats of discrete lengths falling into these areas. Such a situation has

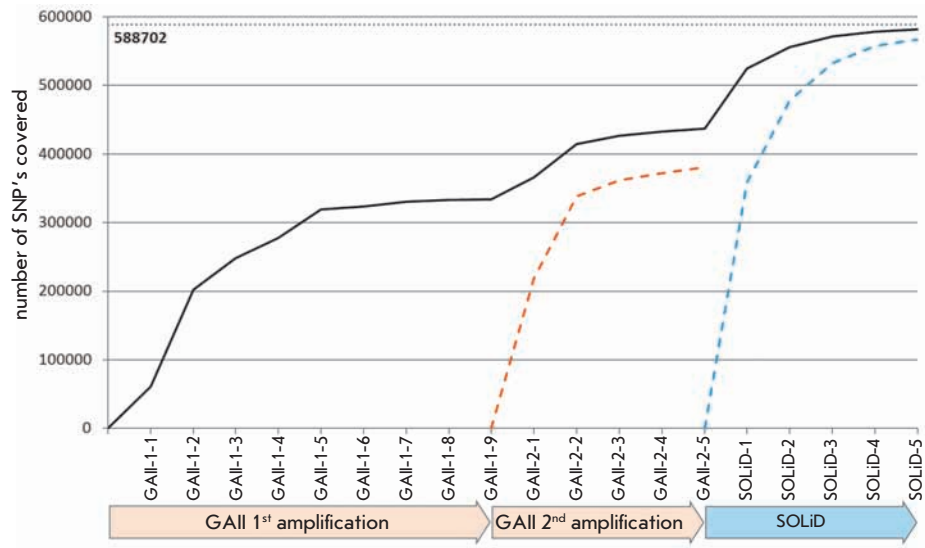
already been described above as a problem encountered in the analysis of coverage histogram using ELAND. The plot of the distribution of the distances between reads in the SOLiD library shows a single peak at about 1,000 bp.

We also analyzed the possible three variants of the paired-end read's relative position and direction. The first variant is when the reads are mapped to the reference genome in accordance with the logic of library design. This means that for the GAI platform two reads face each other, if the 5'–3' direction is taken as the forward direction. As for the SOLiD platform, the logic of library design dictates unidirectional orientation of paired reads. The second and third variants are deviations from the “normal” position, which are only possible when essential rearrangements exist in the corresponding genome area. This leads to either one or two inversions in the paired-end read mapped to the reference genome. In

Table 3. Distribution of reciprocal paired end reads' position and orientation. A single inversion is considered in the case when one of the paired-end reads mapped to the human reference genome ver. hg18 is reversed as against its expected orientation. A double inversion is the case when both reads are inverted. GAI and SOLiD data are shown in orange and blue colours respectively

		GAI	SOLiD
Paired-end reads mapped to different chromosomes		3.18%	4.56%
Reciprocal position of paired reads:	Normal	96.12%	95.22%
	Single inversion	0.48%	0.14%
	Double inversion	0.22%	0.07%
Insertion sizes are in a permissible range		93.06%	95.43%

Fig. 2. Coverage of genotyped SNPs by sequencing data and its saturation throughout the experiment. Shown is the number of SNPs at least once covered by reads mapped to the human reference genome ver. hg18 (by Eland for GAI and BWA for SOLiD data). Every step along the abscissa corresponds to the data from another sequencing run. Saturation only by GAI 2nd amplification runs is shown by orange dashed line. Saturation only by SOLiD runs is shown by blue dashed line



accordance with this definition, all paired-end reads were subdivided into three classes: normal, carrying one inversion, and carrying two inversions (Table 3). Those reads that are mapped to different chromosomes of the reference genome were assigned to a separate class. A small excess of the reads mapped to different chromosomes is explained (in the case of SOLiD) by the presence of the blunt-end ligation stage that is implicitly fraught with errors associated with the covalent linkage of two fragments of different chromosomes. In general, both platforms are near-equal in the percentage of improperly oriented reads.

DISCUSSION AND CONCLUSIONS

We have performed the full genome sequencing of an ethnic Russian male. The characteristic property of this work is the combination of two technological platforms: namely GAI and SOLiD. It is worth noting that data acquisition took approximately the same amount of time (eight weeks) on each platform. During that time, 14 launches (by one flow cell) of the GAI sequencer and 5 launches (by two flow cells) of the SOLiD sequencer were carried out. This allowed to identify 60.2 bln and 48.1 bln nucleotides, respectively. All working parameters of the technological platforms were within the nominal range (as stated by the manufacturers). The data were processed on a computer cluster of the RRC Kurchatov Institute, to which they were transmitted through a high-speed fiber optic cable. This required a special software module ensuring the integrity of the transmitted data. Primary data processing took about 10 weeks.

The main difference between the GAI and SOLiD datasets is in the evenness of the genome coverage by the reads produced by the sequencers. By our estimates, the GAI reads cover about 75% of the genome; and SOLiD reads, 95%, despite the fact that the GAI dataset was bigger compared with the SOLiD dataset. The same trend is observed when analyzing the histogram of genome coverage density (Fig. 1). The peak value of the SOLiD dataset several times exceeds that of the GAI dataset. However, while the number

of reads generated by SOLiD tends to zero already at a coverage density of 20 and more (blue line), the GAI dataset (orange line) has about 10,000 fragments in this range. Thus, the reads generated by the SOLiD system cover the reference genome essentially more uniformly than those generated by GAI. This reflects the quality of the shotgun library construction. The necessity to build several libraries, preferably differing in fragment lengths, for uniform coverage on the GAI platform was noted in studies performed by other research groups, particularly in the sequencing of the genome of a Korean [7]. Moreover, sequencing of two different PCR products of the same primary shotgun obtained at the preliminary step of library construction results in nonidentical, although largely superposed, coverage patterns. For instance, the curve of mSNP filling with reads achieved a plateau after nine launches of GAI with the Amplification No.1 library (Fig. 2). The addition of the quasi-independent Amplification No.2 library resulted in a bounce of the plot, which reached another plateau after five launches of the sequencer. Thus, continuation of launches with the Amplification No.1 and Amplification No.2 libraries could not result in a consequent increase of genome coverage. The addition of the SOLiD dataset to the GAI dataset has solved the undercoverage problem. We suppose that this problem could also be solved on the GAI platform by building an additional library with another mean fragment size.

The sequencing accuracy was estimated by comparing the data on allelic variants of homozygous SNPs determined by sequencing with those determined by genotyping on a DNA microarray. When the SNP is covered by at least ten reads, the error in SNP determination is near equal in two platforms and is about 4–5%. However, the number of homozygous SNPs reliably identified by SOLiD compared with GAI is several times lower, because of insufficiently thick, although smooth, genome coverage by the reads.

The obtained data will allow to identify all SNPs of the given genome, SNP calling, and compare SNP allelic variants predicted by GAI and SOLiD. The analysis of the data on SNPs

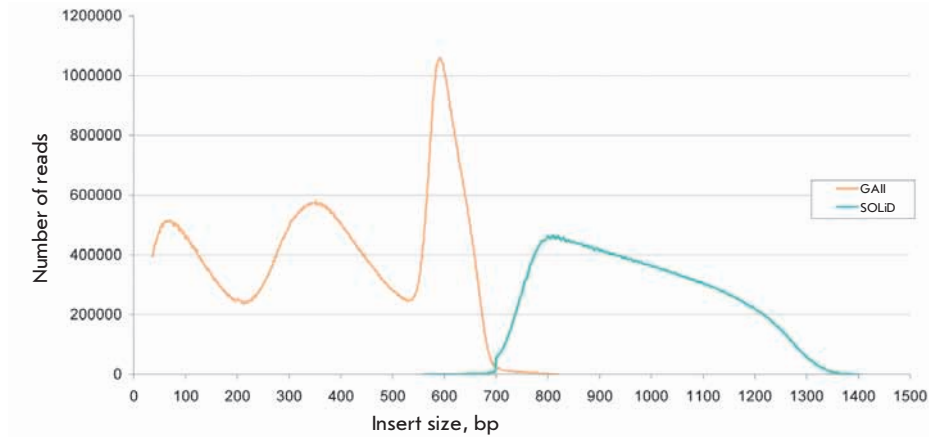


Fig. 3. Paired-end reads' predicted insert size distribution; GAll and SOLiD data are shown in orange and in blue colors respectively. Reads were mapped to the human reference genome ver. hg18 (by Eland for GAll and Corona Life for SOLiD data). The predicted insert size is a genomic distance between locations of mapped reads in each pair

with fixed coordinates, whose allelic variants were determined by both sequencing and genotyping on a DNA microarray, shows that these results should largely have a high predictive potential and overlap no less than 95% of the time. Moreover, a comparison of all polymorphisms of the studied genome with those of already known genomes, particularly Craig Venter's and James Watson's genomes, is of undisputed interest. The second possible task would be to assemble *de novo* extended contigs from those reads that were not mapped to the reference genome. They may be indicative of DNA sequences that are not represented in the reference genome hg18, and further characterizing the genome we are studying.

Our data provide a background for further functional analysis of this genome. In particular, the data on hetero-

zygous SNPs in expressed blood cell mRNAs will allow to determine transcripts with biallelic expression bias. This association between the epigenetic and genetic components of the given genome is of certain interest for further studies. However, this genome first has to become a wholesome model object, which can be achieved via immortalization of N's somatic cells followed by construction of the cell line. This will allow all interested research groups to use exhaustively the data presented in this work. ●

The study was supported by the Russian Federal Agency for Science and Innovation.

REFERENCES

- Lander E., Linton L., Birren B., Nusbaum C., Zody M., *et al.* // *Nature* 2001. V. 409. PP. 860–921.
- Levy S., Sutton G., Ng P., Feuk L., Halpern A., *et al.* // *PLoS Biology*. 2007. Sep 4;5(10):e254.
- Wheeler D., Srinivasan M., Egholm M., Shen Y., Chen L., *et al.* // *Nature*. 2008. V. 452. PP. 872–876.
- Bentley D., Balasubramanian S., Swerdlow H., Smith G., Milton J., *et al.* // *Nature*. 2008. V. 456. PP. 53–59.
- McKernan K., Peckham H., Costa G., McLaughlin S., Fu Y., *et al.* // *Genome Research*. 2009. V. 19. № 9. № 1527–1541.
- Wang J., Wang W., Li R., Li Y., Tian G. // *Nature*. 2008. V. 456. PP. 60 – 65.
- Ahn S., Kim T., Lee S., Kim D., Ghang H., *et al.* // *Genome Research*. 2009. V. 19. № 9. PP. 1622–1629.
- <http://www.ncbi.nlm.nih.gov/guide/genomes/>

Cell Regulation of Proliferation and Differentiation *ex vivo* for Cells Containing Ph Chromosome in Chronic Myeloid Leukemia

N.I. Grineva*, T.V. Akhlynina, L.P. Gerasimova, T.E. Manakova, N.G. Sarycheva, D.A. Schmarov, A.M. Tumofeev, N.M. Nydenova, L.Yu. Kolosova, T.I. Kolosheynova, L.G. Kovaleva, S.V. Kuznetsov, A.V. Vorontsova, A.G. Turkina

GU National Research Center for Hematology, Russian Academy of Medical Sciences

*E-mail: nigrin27@mail.ru

ABSTRACT Cell regulation of Ph⁺ cell proliferation and differentiation has been studied *ex vivo* in various chronic myeloid leukemia (CML) patients. The regulation is provided by alternation of effective stages of proliferation and maturation with inhibition of Ph⁺ cell proliferation by accumulating neutrophils under apoptosis blockage. The alternation of stages consists of switching stage 1 (effective proliferation) to stage 2 (effective maturation) and proceeds according to the 1/2 -1/2/1 or 2/1-2/1/2/1 schemes. The kinetic plots of alternations pass through control points of crossing plots, where the parameters of proliferation and maturation are equal. The indices of P/D efficiency (ratio of proliferation and maturation rates) are 1.06 ± 0.23 and don't depend on time, alternation order, or sources of Ph⁺ cells - CML patients. During stages alternation, conversely, the parameters of Ph⁺ cell proliferation and maturation vary. The proliferation stages are characterized by increased proliferating cells content, a decreased number of neutrophils, and apoptosis induction. At the maturation stages, conversely, apoptosis is inhibited, the number of mature neutrophils increases, while immature Ph⁺ cells decrease. High content neutrophils inhibit the proliferation of Ph⁺ cells and impair their own maturation by inversion of maturation order, probably through a feedback mechanism. The regulation differences *ex vivo* reveal three types of Ph⁺ cells from various individual CML patients, distinguished by the number and duration of alternating stages of proliferation and maturation. Ph⁺ cells types 1 and 2 have one prolonged stage of effective proliferation or effective maturation with efficiency indices $P/D^1 = 1-20$ or $P/D^2 \leq 1$. At the same time period, the proliferation and differentiation of the Ph⁺ cells type 3 proceeds with repeated alternations of stages with $P/D^1 = 1-4$ or $P/D^2 \leq 1$. Type 1 Ph⁺ cells (~20%) were isolated from patients in advanced stages of CML, while Ph⁺ cells types 2 and 3 (30 and 50% correspondingly) were isolated from CML chronic phase patients sensitive to chemotherapy.

Keywords: Regulation of Ph⁺ cell proliferation and Ph⁺ cell by mature cells, cultivation of hematopoietic Ph⁺ mononuclear cells, kinetics of Ph⁺ cell proliferation and differentiation *in vitro*, Ph⁺ cells apoptosis, Ph⁺ cell distribution in cell cycle phases, Ph⁺ cell proliferation and differentiation efficacy, inversion of accumulation order for maturing neutrophils, chronic myeloid leukemia.

Abbreviations: P&D - proliferation and differentiation; CML - chronic myeloid leukemia; Ph - Philadelphia chromosome; MM, B, S, - metamyelocytes, band and segmented neutrophils; Ph⁺ cells - hematopoietic cells with Philadelphia chromosome; PB - peripheral blood; BM - bone marrow; CP, AP, BP - chronic phase, accelerated phase and blastic phase of CML; FCS - fetal calf serum.

INTRODUCTION

Leukemias accounts for 1% of all deaths and 4-10% of deaths from cancer. The prevalence of leukemias and lymphomas varies from 3 to 9:100 000, depending on the geographical region. Unfavorable radiation and ecological environment can increase it by 1.5 logs. In the U.S., leukemias are the major reason of death in children before 15.

The majority of leukemias result from genetic disturbances: chromosomal aberrations, translocations, inversions, deletions, and various mutations [1-3, 6].

Philadelphia-positive (Ph⁺) cells, expressing active tyrosine kinase p210 or p185 (oncoproteins, products of bcr/abl gene), are involved in chronic myeloid leukemia (CML) pathogenesis. It results in reciprocal chromosomal translocation t(9;22)(q34;q11) in the polipotent hematopoietic stem cell. Proliferation and differentiation of this cell leads to replacement of normal hematopoietic cells by their monoclonal neoplastic Ph⁺ counterparts, thus promoting the development and progress of CML [1- 8, 10,12].

The CML clinical course varies among different patients. The cellular and molecular mechanisms of these differences

remain unclear. Current knowledge of CML course and progression *in vivo* is based upon analyses of averaged values of various parameters obtained at different moments and CML phases. CML undergoes a chronic phase (CP), accelerated phase (AP), and an acute rapidly progressing blast phase (BP) with an inevitable fatal outcome. Current CML therapy is based upon highly specific targeted drugs, tyrosine kinase inhibitors (TKI), specifically blocking p210 - imatinib and its analogues. Imatinib allows to extend life by 6 years in 88% of patients., of which 66% continue treatment. In 14% of those patients, CML progresses, while 5% of them interrupt treatment because of toxicity. The toxicity is associated with additional bcr/abl gene mutations, leading to therapeutic resistance. Despite the development of a new generation of TKIs, the problem remains unsolved, because none of them kills the resting leukemia stem cells. Fewer than 5% of CML chronic phase patients are cured, while the majority eventually relapse [6]. There is a need for another strategy in dealing with leukemia stem cells.

Despite the extensive research of Ph⁺ cells both in cultures and *in vivo* [4, 5, 7-12, 15-20,23,27], the processes taking place

in the cells of newly diagnosed CML patients and in those in progression remain poorly studied. There is no unified conception of the biological and molecular processes underway in CML both *in vitro* and *in vivo* and their interaction. Little is known about the patterns of proliferation and differentiation (P&D) of Ph⁺ cells even *in vitro*.

Researchers have often noted that the cellular processes in cells isolated from CML patients differ from those in cell lines. The number of proliferating hematopoietic progenitors in CML is decreased, while the number of nondividing mature neutrophils is higher than in normal cells [13,14,20-23,26]. Ph⁺ stem cells proliferate less actively than healthy donor stem cells, while more differentiated Ph⁺ cells are accumulating more actively [20, 23, 24, 27]. This leads to suggest that the cause of CML is the imbalance between stem cells self-renewal and myeloid progenitors maturation, rather than Ph⁺ cells proliferation under the action of p210^{bcr/abl}, as follows from many other studies [4-7,10-12, 15, 16, 19, 24-29, 31].

It should be mentioned that the question of apoptosis block in Ph⁺ cells is also not obvious. Evidently, our knowledge of cellular and molecular regulation of Ph⁺ cells in CML is still incomplete, though the oncogenic potential of bcr/abl oncogene, including enhanced cell viability, proliferation activation, and resistance to apoptosis, is studied in details in various studies [7, 4,17,18, 25-27, 29, 30, 35, 43-45].

The influence of different p210 mutations on proliferation, apoptosis, and the malignant properties of Ph⁺ cells transfected with different bcr/abl mutated variants was studied in [27]. The activation of proliferation and apoptosis inhibition by p210 was shown to be independent processes closely linked to different bcr/abl mutations (including those responsible for intracellular signaling via various STAT proteins). It was hypothesized that those mutations can influence the severity of proliferation activation and apoptosis inhibition and shift their relative impact. In Ph⁺ cells of CML patients, these properties are not studied due to lack of suitable models and experimental approaches, though the bcr/abl mutations are actively studied and used in practice for CML diagnosis and treatment optimization [19, 29, 30-32]. Based on these considerations, we have suggested that the kinetics of Ph⁺ cells P&D can reveal the differences in Ph⁺ cells P&D regulation of individual patients *ex vivo*.

The goal of our study was to clarify the mechanism of Ph⁺ cells P&D in individual CML patients *ex vivo* in suspension cultures by means of the previously developed kinetic method of study [33, 39]. The pattern of CML cells proliferation and differentiation in cultures is similar to that of enhanced myelopoiesis in CML *in vivo*, allowing to estimate the efficiency of P&D, influence of growth factors, and the expression of bcr/abl and differentiation antigens in Ph⁺ cells.

MATERIALS AND METHODS

Materials: heparin (Flow, UK); Lymphoprep, alpha-MEM media (MP Biomedical, USA); DEPC, Hepes, Tris, PBS, fetal bovine serum (FBS), sodium citrate, laurilsarcosyl (ICN, USA); trypan blue stain, L-glutamine and 2-mercaptoethanol (Serva, Germany), penicillin and streptomycin (OAO "Biochimik" Saransk Russia); G-CSF (F.Hoffmann-La Roche Ltd, France);

tabletted PBS (10 mM phosphate buffer+ 0,13 M NaCl + 2,7 mM KCl, pH 7,4), NPZ "EKO-servis" Russia.

Ph⁺ mononuclear cells used for the study were prepared from the peripheral blood (PB) or bone marrow (BM) of CML patients in chronic phase (CP), accelerated phase (AP), and blastic phase (BP) before or under treatment. Leucocytes and granulocytes are the main CML mononuclears content; namely, they have been investigated here. Characteristics of Ph⁺ cells and CML patients from whose PB and BM the mononuclears were isolated are given in Tables 1-3. In Ph⁺ cells, bcr/abl RNA types as b3a2, b2a2, or e1a2 were assayed by RT-PCR (reverse transcription-polymerase chain reaction) according to [33, 35].

Ph⁺ mononuclear cells were isolated from 10–15 ml of PB or 1-2 ml of BM aspirate (from superior iliac crest) of CML patients at different phases of the disease. The heparinized (50 IU/ml) material was centrifuged for 30 min at 1500 rpm over Phycoll or Lymphoprep (1.077 or 1.119 g/sm³). The resulted light fraction was washed twice with PBS at pH 6.8 and once with a α -MEM media and then re-suspended in α -MEM media for analysis and cell cultivation. This fraction contained progenitor cells (blasts), lymphocytes, granulocytes, and monocytes, as well as some quantities of mature neutrophils, typical for CML mononuclear cells. The viable and dead cell number was counted 3 times in smears, stained by 0.2% trypan blue according to Romanovsky, with consequent cell counting in Goryaev's chamber.

Ph⁺ cells cultivation was performed according to [33]. Cell suspension of $2 \div 8 \cdot 10^6$ cells/ml was incubated with α -MEM media containing 10-20% fetal calf serum (FCS, 2 mM L-glutamine, 10^{-4} M 2-mercaptoethanol, 100 U/ml penicillin and 50 U/ml streptomycin, 25 mM HEPES-NaOH, pH 7.2-7.4 in 25 sm² plastic flask for 2-3 h) nonadherent cells were then centrifuged at 1500 rpm for 7 min and re-suspended with the same medium to $0.8 \div 1.4 \cdot 10^6$ cells/ml and transferred to 24- or 96-well plates (12 wells per probe) and incubated at 37 °C with 100% humidity and 5% CO₂ without FCS for 2 h. Then 10-20% of FCS was added, and the cells were cultured for 6-14 days, selecting each sample from separate probes. Every point was tested in triplex. Ph⁺ cells in the collected samples were washed from FCS by centrifuging in α -MEM media and analyzed for their morphology, cellular composition, distribution in the cell cycle phases, and apoptosis. We had separately determined that 2-hour incubation of cells without FCS with consequent incubation with FCS diminishes cell proliferation during the next ~6 hours. After that, the proliferation rate is restored, and 12 hours later the cellular composition becomes the same as in the cultures with FCS.

The prepared probes were analyzed for the number of viable and dead cells and cell morphology in 3 zones of smear (100 cells in each) according to Romanovsky, the individual cell morphology was identified according to Abramov [34], and we also calculated the percentage of each cell type in every sample. We calculated cell concentration in samples in 10^6 cells/ml and obtained plots for the kinetic curves of accumulation and depletion of different Ph⁺ cells and their subpopulations: proliferating cells (blasts, promyelocytes and myelocytes) – P and nondividing mature neutrophils: MM, B and S (metamyelocytes, band and segmented neutrophils) – D. The plots reflected the rate of production (or

accumulation) of each differentiating cell type, transforming to the next subpopulation for this cell differentiation line. The mean error was $\pm 5 \div 11\%$. We also studied apoptosis and cell cycle phase distribution by flow cytometry.

Ph-chromosome in PB and BM cells of our patients was identified cytogenetically (for 100 mitoses) or by FISH in the cytogenetical laboratory of the National Research Center for Hematology.

Flow cytometry for the analysis of the cell cycle phase distribution of the cultured Ph⁺ cells selected during the incubation of Ph⁺ cells samples (5,000 cells each) after the isolation in Phycoll gradient were centrifuged for 7 min at 2000 rpm and 4°C, washed with PBS, and accurately fixed by cold 70% ethanol for 30 min at 4°C. Before measurement, the suspension was washed with PBS, centrifuged, and the pellet was incubated in 0.5 ml of PBS containing 5 mcg/ml of propidium iodide and 50 mcg/ml of ribonuclease A for 30 min at room temperature in a dark place [36, 37]. The cell measurements were done with flow cytometer EPICS-XL. Cells of granulocytes gate were analyzed by direct (FSC) and side (SSC) light scatter with simultaneous registration of the FL2 fluorescence by amplitude and impulse square (it allows to exclude aggregated cells, conglomerates, and debris) in the linear and logarithmic scales. Simultaneously apoptotic cells are detected. FL2-H particles with hypodiploid DNA located as a separate peak leftward of the peak of diploid cells (decrease of cell size not more than by 2 logs) were considered to be apoptotic. The percentage of apoptotic granulocytes was estimated within the granulocytic gate, where there is no cell debris. The DNA histograms from the same cell samples were analyzed for cell cycle phase distribution (S, G2/M). Its analysis was done with the help of previously developed specialized software. It was based upon an algorithm developed for asynchronous proliferating cell populations (SFIT-method) [38].

The kinetic plots of Ph⁺ leucocytes (number increase and death of Ph⁺ cells) were obtained using the percentage of viable and dead cells (see above) measured in 10⁶ cells/ml. The proliferation rate of leucocytes in whole and granulocytes in particular was estimated from the kinetic plots of its accumulation and depletion under cultivation and also from the sum of its subpopulations, defined by morphology. The kinetic curves showed that the accumulation and consumption of each studied type of cells parallels its transformation into the next subpopulations. The depletion of segmented neutrophils means its death.

Differentiation plots of Ph⁺ leucocytes and its subpopulations – lymphocytes, all myeloid cells, granulocytes, and its subpopulations, blasts, promyelocytes, myelocytes, metamyelocytes, band and segmented neutrophils – were defined, calculating the concentration of corresponding cells in samples and multiplying its fraction (defined by morphology in smears) by 10⁶/ml. The cell morphology in smears was assessed as already mentioned. The cellular composition was assessed in smears in 3 zones of 100 cells each. It should be noted that besides morphology Ph⁺ CML cells № 1.1 and № 2.6 were identified by CD antigens expression according to [39], where the results of identification and kinetics of cells according to both antigen expression and morphology are given and results are shown to be coincident [33, 39].

Kinetic plots of P/D efficiency. P/D of granulocytes in Ph⁺ cell cultures was defined as the P/D efficiency index that is the rate ratio of accumulation and depletion of proliferating cells (P cells sum) and neutrophils matured without dividing (D cells sum) under Ph⁺ mononuclear cells cultivation as defined above.

The P/D efficiency index was defined as the ratio of P cells (proliferation rate) and D cells (differentiation rate). It is equivalent to the ratio between the concentration of these cells, according to consideration $V_p/V_d = K_p [P] t / K_d [D] t = K_p / K_d \times [P] / [D]$, where V_p and V_d are the rates of P and D cells accumulation; K_p and K_d , constants of rate; and $[P]$ and $[D]$, cell concentrations. $K_p / K_d = K$ is the constant of relative P/D efficiency.

RESULTS

In order to study the differences in P&D in the Ph⁺ cell cultures of patients with different CML phases, we obtained kinetic plots for 34 samples of Ph⁺ mononuclear cells from the PB and BM of 23 CML patients under cultivation in strictly identical conditions. Ph⁺ mononuclear cells from PB and BM contain hematopoietic cells capable of self-renewal at P&D and to P&D for 2-3 cycles, forming the full set of Ph⁺ cells [33, 39].

The majority of mononuclear cells in CML are leucocytes [1, 7]; thus, we considered them as Ph⁺ cells. The leucocytes subpopulations include granulocytes, lymphocytes, monocytes; the granulocyte subpopulation contains myeloid precursors (blasts), promyelocytes, myelocytes and mature neutrophils: metamyelocytes (MM), band (B), and segmented (S) neutrophils.

The characteristics of Ph⁺ cells and corresponding CML patients are given in Tables 1-3. In cell samples obtained under cultivation, we obtained the kinetic plots of CML leucocytes proliferation and death and differentiation of Ph⁺ granulocytes subpopulations: blasts, promyelocytes, myelocytes, MM, B and S. In the same samples, we also assayed plots for apoptosis and distribution of Ph⁺ cells in the cell cycle phases (Fig. 1-4). The experiments showed a typical for CML elevated granulocytes content [1, 6, 14]. The shape of the plots indicates the rate of production or accumulation of the corresponding cell type, their transformation to next subpopulations, and final death. Thus, the resulting plots reflect the main differentiation processes (Fig.1-4 and Tables.1-3). The leucocytes differentiation coincides with known CML data [1-3, 21, 22, 40-42].

Besides the differentiation rate of separate subpopulations (Fig.1-4 a), we studied the plots of overall proliferation and maturation rate. In the 1st case, these were the plots of accumulation of P cells differentiating in parallel with proliferation and including blasts – myeloid cell precursors, promyelocytes, and myelocytes. In the 2nd case, they were the plots of accumulation of neutrophils – D cells: the sum of MM, B, and S matured without dividing (Fig.1-4 b).

TYPES OF CML CELLS

Myelopoiesis (P&D of myeloid division of hematopoiesis) begins with the proliferation and differentiation of hematopoietic progenitor cells and their immature progenies. Then P&D

RESEARCH ARTICLES

Table 1. Proliferation and differentiation of types 1 and 2 Ph⁺ cells in culture

N	Samples of mononuclear cells from PB or BM of CML patients			Proliferation and differentiation parameters					
	Sample number CML PB or BM	Diagnosis, treatment before sampling	Type of bcr/abl RNA, leucocytes x10 ⁹ /L; % blasts	P/D, [immature]/[mature]	P/D duration, days	[S]/[M]	S x10 ⁶ cells/ml, max/at day	Neutrophils x10 ⁶ cells/mlmax /at day	Apoptosis(cell death) % /at day
1	2	3	4	5	6	7	8	9	10

Type 1 Ph⁺ cells, P/D¹ ≥ 1. Effective proliferation (accumulation rate of immature Ph⁺ cells is higher than neutrophils maturation rate). Concentration [immature] > [mature].

1	1.1 PB	CP	b3a2 L115; 3 %;	12 - 4 4 - 1	14	0,14 - 0,0 - 0,27	0,17	0,53/ 7 -0,32 -0,65/ 14	(28 / 7 50 / 14)
2 3	1.2 BM PB	CP, HU*	b3a2 L 72 blasts 3%	1,2 -1,6 2,4 - 1	8 8	0,41 0,5	0,2 0,2 / 3	0,67 0,58 / 3	(33 / 4 30 / 9)
4 5	1.3 KM PB	AP Chem**	b2a2 blasts 17%	3 - 9 1 - 13	> 8	0,1-0,54 0,1-0,31	0,13 / 8	0,17/ 0-4	(14 / 8)
6	1.4 Blood	lymphoidBP, Chem**	b2a2 blasts 30%	>20- 2	> 6	0,0	0/ 0-6	0,4 / 6	(47 / 6)
7	1.5 PB	CP, HU*	b3a2 L 175 Blasts 5%	1,2-2,5	< 8	0,14-0,53	0,08 / 4	0,32-0,2 / 4, 8	(21 / 7)

Type 2 Ph⁺ cells, P/D² ≤ 1. Effective maturation (neutrophils maturation rate is higher than accumulation rate of immature Ph⁺ cells). Concentration [mature] > [immature].

1	2.1 PB	CP→BP HU*	b3a2, L 188-145, 2 %	0,3-0,8	> 6	1,0-1,62 -0,4	0,47 / 2	1,2 / 2	40 / 1 4 / 3 55 / 6
2	2.2 BM	CP HU* 4 days leter	b3a2	0,8-1,5	> 5	0,8-0,4-1,0	0,5 / 2	0,7/1 0,6/5	(4 / 1 11 / 5)
3	2.3 PB	CP	b3a2	0,4-1,0	~ 6	0,4-0,8	0,25 / 3	0,7/3	(2 / 3 19 / 7)
4	2.4 PB	CP, 2 risk group,***	b3a2, L 165, blasts 2%	0,3-1, 2	~ 7	0,73 / 4 2,4 / 8	0,86 / 8	2,4 / 4 1,3 / 8	(2 / 4 33 / 11)
5	2.5 PB	CP	b2a2	0,1-1,1- 1,6	~ 7	1,4 / 5	0,7 / 5	1,13 / 5	(3 / 2 2 / 5 17 / 7)
6 7	2.6 PB BM	CP	b3a2	0,2-1,0 0,3 -1,0	>9	8 / 6 24 / 11	0,41 / 6	0,7 / 6 0,1 / 11	(2 / 6 24 / 11)
8 9	2.7 BM PB	CP	b2a2	1,5 1,5	~ 4-6	- -	- -	1,1 / 3 1,3 / 7	(15 / 1 2 / 2-7 2 / 3 12 / 8)

The treatment before sampling: *hydroxyurea; **chemotherapy. *** interferon
CP, AP and BP – chronic phase, accelerated phase, and blastic phase – CML phases.

continues with neutrophils maturation without dividing [3, 40, 41]. In this study, we obtained and analyzed *ex vivo* P&D kinetic plots in 34 samples of Ph⁺ cells from the PB and BM of 23 CML patients. They demonstrated evident differences in the P&D of Ph⁺ cells from individual CML patients.

The kinetic plots of proliferation and maturation reflect the rate of two crucial P&D processes: proliferation with differentiation and differentiation without dividing (maturation of neutrophils). The rate ratio of proliferation (P cells ac-

cumulation) and maturation (D cells) reflects the efficiency of Ph⁺ cells proliferation relative to its maturation without dividing and is expressed in the P/D index, an indicator of P/D efficiency. One can see from the kinetic P and D cell plots that the proliferation rate can be higher or lower than that of maturation; thus, P&D can give advantage to proliferation (effective proliferation) or to maturation (effective maturation). In the first case, the P/D efficiency index P/D¹ ≥ 1; in the second, P/D² ≤ 1 (Fig.1-4C, Tables 1-3).

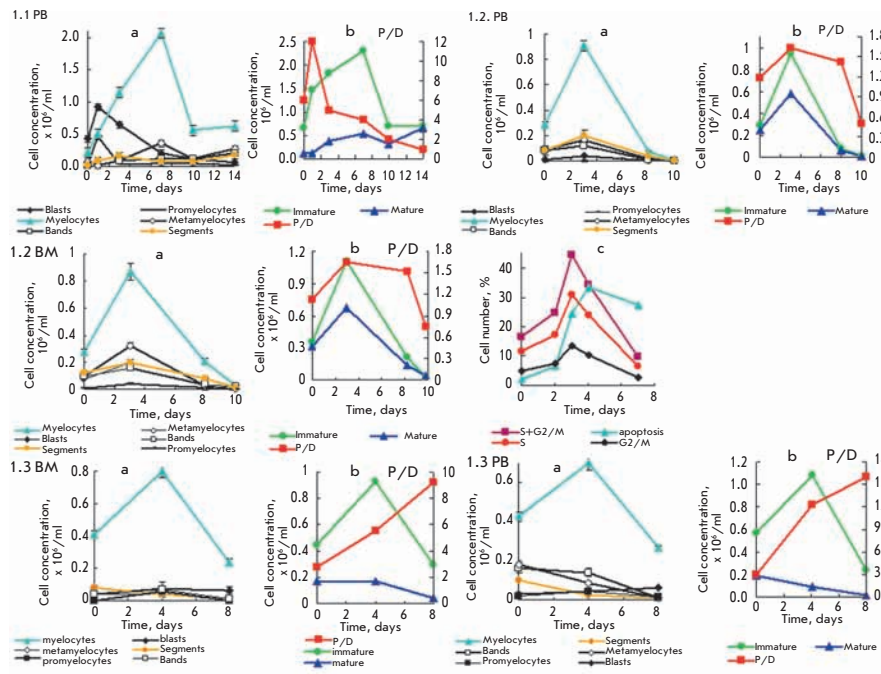


Fig. 1. Kinetics plots of proliferation and differentiation (as accumulating and consuming) of type 1 Ph⁺ cells subpopulations in the suspension culture of mononuclear cells isolated from bone marrow (BM) or peripheral blood (PB) of chronic myeloid leukemia (CML) patients a – differentiation of granulocytes subpopulations with formation of proliferating myeloid precursor cells (blasts), promyelocytes, myelocytes, and neutrophils matured without dividing: metamyelocytes, bands, and segmented; b – accumulation and consumption of immature proliferating cells, P cells, and neutrophils matured without dividing, D cells, as well plots of proliferation and maturation efficiency as ratio of P and D cells accumulation rates – P/D index; c – plots of apoptosis and Ph⁺ granulocyte distribution in the cell cycle phases in culture. 1.1 PB – PB Ph⁺ cells from CML CP patient #1.1 before treatment, showing fast progression to CML BP. 1.2 PB and 1.2 BM – Ph⁺ cells from CML CP patient #1.2; 1.3 BM and 1.3 PB – Ph⁺ cells isolated from patient with CML AP under treatment (rapid progression to blastic phase). Samples characteristics, as well as proliferation and differentiation data of type 1 Ph⁺ cell, are shown in Table 1

In 20% of CML patients, P&D is associated with continuous effective proliferation only. These are type 1 Ph⁺ cells, their $P/D^1 = 1-20$ (Fig. 1 b, Table 1). P&D of Ph⁺ cells from other CML patients is associated with effective maturation only with $P/D^2 \leq 1$. It corresponds to type 2 Ph⁺ cells (~30% of cases) (Fig. 2 b, Table 1). During P&D, these indices vary in their limits of P&D. The Figures 1, 2 and Table 1 show that the continuance of P&D of Ph⁺ cells types 1 and 2 (7–14 days) is comparable with that for 1–2 cell cycles of hematopoietic cell clon [3, 40, 41].

The P&Ds of most samples of Ph⁺ cells are characterized by repeated alternations of efficient proliferation and efficient maturation with the prevalence of the proliferation rate and high P/D index ($P/D^1 \geq 1-4$), or with the prevalence of the maturation rate and low P/D index ($P/D \leq 1$) (Fig. 3, b, Fig. 4, b and Tables 2, 3). These Ph⁺ cells constitute type 3 of Ph⁺ cells. The alternations of proliferation (1st stage) and maturation (2nd stage) are clearly identified by crossings of the kinetic plots of accumulation of proliferating (P [immature]) and maturing (D [mature]) cells (Figs.3 and 4b, Table 3). Proliferation and maturation can be accompanied by 1–3 such alternations that are changing each other according to the following schemes: 1/2 -1/2/1 or 2/1-2/1/2, with frequency from 1 to 5 days, rarely 0.2–6 days (Tables 3 and 4). In alternation according to schemes 1/2 - 1/2/1, the rates and, hence, cell concentrations, are changing in the following series: [immature] > [mature] → [mature] > [immature] → [immature] > [mature] (Figs.3: 3.1- 3.4, b). In alternation according to schemes 2/1 - 2/1/2, the rates and, hence, cell concentrations, are changing in the following series: [mature] > [immature] → [immature] > [mature] → [mature] > [immature] (Figs.4: 3.10-3.12, b). P&D of Ph⁺ cells type 1 and 2 can be considered as a particular case of prolonged alternation of proliferation or maturation of Ph⁺ cells type 3.

At the moment of crossing of effective proliferation and effective maturation plots, the alternation of stages, rate of Ph⁺ cells accumulation, and other P&D parameters takes place. At the moment of crossing, the proliferation and maturation rates, as well as the indices of their efficiency, become equal. At that moment, $P/D = 1.06 \pm 0.23$ regardless of scheme, time point, alternation sequence or cell concentration (Tables 2 and 3). These moments can be called critical or equilibrium points. Other characteristics that may be expected to be equal (concentrations of immature cells, mature neutrophils ($0.4-0.7 \times 10^6/\text{ml}$) and segmented neutrophils only ($0.1-0.4 \times 10^6/\text{ml}$)) in fact were similar (Table 3). However, the prominent variations ($1.53-0.1 \times 10^6/\text{ml}$ for BM 3.5, PB 3.10, 3.11 and BM 3.12 in Table 3) suggest the concentration's dependence.

During alternations, the cell characteristics are continuously changing (Table 2). The parameters increase or decrease, passing maximums and minimums. These events are usually asynchronous (Fig. 1-4 and Table 2). The pattern of P/D indices changing depends upon the stage. They vary within the limits $P/D^1 = 1-4$, $P/D^2 \leq 1$. The only exception was BM № 3.3, where P/D^1 decreased from 11 to 1 3 days before it reaches a steady-state and the maturation stage begins. Under maturation, stage P/D^2 varies from 0.1 to 1, until a new proliferation stage with $P/D^1 > 1$. It means that the rates of proliferation and maturation – their P/D efficiency indices and concentrations of proliferating and maturing cells (especially segmented neutrophils) – are rather significant parameters of Ph⁺ cells P&D.

Analysis of the kinetics of granulocytes subpopulations development, the kinetics of apoptosis, and distribution of Ph⁺ cells in the cell cycle phases in combination with differences in P/D index and the rates of proliferation and differentiation stages of Ph⁺ cells P&D revealed the following.

RESEARCH ARTICLES

Table 2. Parameters of alternation for efficient proliferation and efficient maturation stages under condition of proliferation and differentiation of 3 type Ph⁺ cells in the culture

Item №, Figure №	№ XMJI (*)	Parameters of proliferation stage Max or interval (days)				Parameters of maturation stage Max or interval (days)				[S]/ [myelocytes]	Apoptosis (death) % / at day
		P/D ¹	10 ⁹ cells/ml			P/D ² Min	Max, 10 ⁶ клеток/мл				
			[immature]	[mature]	[S]		[[imma- ture]	[mature]	[S]		

Alternation stages according to 1/2 or 1/2/1 schemes with changing accumulation rates and Ph⁺ cell concentration
[immature] > [mature] → [mature] > [immature] → [immature] > [mature]

1, Fig.3	3.1 BM (1/2/1)	1,4-1,1 1,2-2,3	0,35-0,5 0,64- 0,5	0,29-0,5 0,63-0,2	0,1-0,25 0,3 -0,1	0,9 (at 3 day)	0,71 (at 3 day)	0,8 (at 3 day)	0,4 (at 3 day)	0,5-0,2	5-12
2, Fig.3	3.2** PB (1/2)	1,3-1,2	0,4	0,30	0,05	0,85	0,4-0,3	0,53	0,16	0,5 -0,25	< 5
3	3.3 BM	11- 2	0,6-0,46	0,02-0,5	0,01-0,4	0,1	0,5-0,1 0,5- 0,1	0,6 (3 day)	0,4-0,2	0,1-2,5	(7 at 11 day)
4 Fig.3	PB (1/2)	1,32-1,5	0,5-0,8	0,4-0,5	0,14 -0,25	1,0-0,2		0,6 (7 day)	0,3-0,2	0,1-1-2,5	
5 Fig.3	3.4 BM (1/2/1)	1,2 1,53	0,4-0,8-0,5- 1,1	0,34	0,17	0,52-0,34	0,57-0,34	1,1	0,7	0,4-1,3-2	(18 at 5 day)
6	3.5 BM (1/2)	3,43-1,0	1,95	1,65	0,03- 0,37	1,0-0,3	1,53-0,4	1,53-1,21	0,37-0,5	0,19- 0,25-2,0	(30 at 11day)
7	3.6 BM (1/2)	1,61	0,61	0,52	0,24	0,28	0,13- 0,56	0,52 -0,4- 0,6	0,24	0,13 -0,5-1,0	(22 at 11 day)
8	3.7 BM (1/2)	2,5-1,4	0,65-0,60	0,55	0,3	0,7-0,4	0,55	0,83	0,53	0,1-0,5	(30 -35 at 10-11 day)
9	PB (1/2/1)	2,2-1,2 1,2-1,4	0,51-0,54 0,38-0,32	0,5 -0,54 0,4-0,2	0,05-0,19 0,15-0,07	0,78	0,57- 0,38	0,73	0,27		
10	3.8 BM (1/2)	3,9-1,5	0,53-0,35	0,14-0,35	0,02-0,22	1,5-0,5	0,35-0,23	0,35-0,44	0,2-0,4	0,04-0,3	(45 at 7 day)

Alternation stages according to 2/1 or 2/1/2 schemes with changing accumulation rates and Ph⁺ cell concentration:
[mature] > [immature] → [immature] > [mature] → [mature] > [immature]

11, Fig.4	3.10* PB (2/1)↔	1,4 -3,1-2,0	0,3-0,6 -0,4	0,3-0,2	0,17-0,25	0,2-0,5	0,13-0,3	0,5-0,8-0,3	0,35-0,7	8,5-0,3	34-22 at 2-5 days
12, Fig.4	3.11 PB (2/1/2) ↔	1,15-1,63	0,8-0,6	0,6-0,8-0,1	0,43-0,3	0,33-1,15 0,75-0,8	0,2-0,8, 0,6-0,8-0,1	0,5-1,3-0,8, 0,8-0,1	0,16-0,75-0,4, 0,39-0,06	1,2-0,2	35 at 1 day, 63 at 4 days
13, Fig.4	3.12 BM (2/1/2/1/2) ↔	1,0-1,32 1,0	0,7-0,78 0,24	0,6 0,24	0,14-0,35 0,12	0,5-0,7-1,0 1,0-0,8- 1,0	0,6-0,7 0,6 0,2	0,4--0,95 0,6	0,14-0,5 0,34 0,1	0,9-0,2	7, 22 at 6, 11 days
14	3.13 PB (2/1) ↔	1,7-2,5	0,39	0,4-0,2	0,2-0,1	0,45-1,7	0,42	0,9-0,4	0,3-0,2	0,9-1-9	(48 at 8 days)
15	3.14 BM (2/1) ↔	1,0-1,34	0,6-0,5	0,4	0,4 - 0,2	1,1-0,7	0,4-1,0	0,4-1,37	0,12-0,92	0,3-1,0	(45 at 5 days)

Footnote: *alternation scheme; **acceleration phase; BM – bone marrow, PB- peripheral blood; brackets [] show cell concentration; hyphen - interval of values on the kinetic plots; Figure ↔ shows that 1st stages data for alternation 2/1 and 2/1/2/1 are given earlier than that of 2d stages according to the subtitle of the table.

Proliferation and differentiation of type 1 Ph⁺ cells prolonged effective proliferation with P/D index >1-2 ÷ 20 and amount of Ph⁺ cells in S+G2/M phases ≥ 20-45%; the proliferation rate prevails over maturation (Figs.1.2, 1.3.a ,b, c). P&D is associated with enhanced accumulation of myelocytes, promyelocytes, and/or blasts and low accumulation of neutrophils, maturing without dividing (Figs.1,

a, b,: 1.1-1.3 and 1.4 in Table 1) and with a high level of apoptosis (Fig.1, c). The lower neutrophils content characterizes the higher P/D index (Fig.1: 1.1-1.3, b) and CML progression (Table 1). These Ph⁺ cells were isolated from the PB and BM of CML patients at all phases of the disease during progression and amounts to one-fifth of all studied samples of Ph⁺ cells.

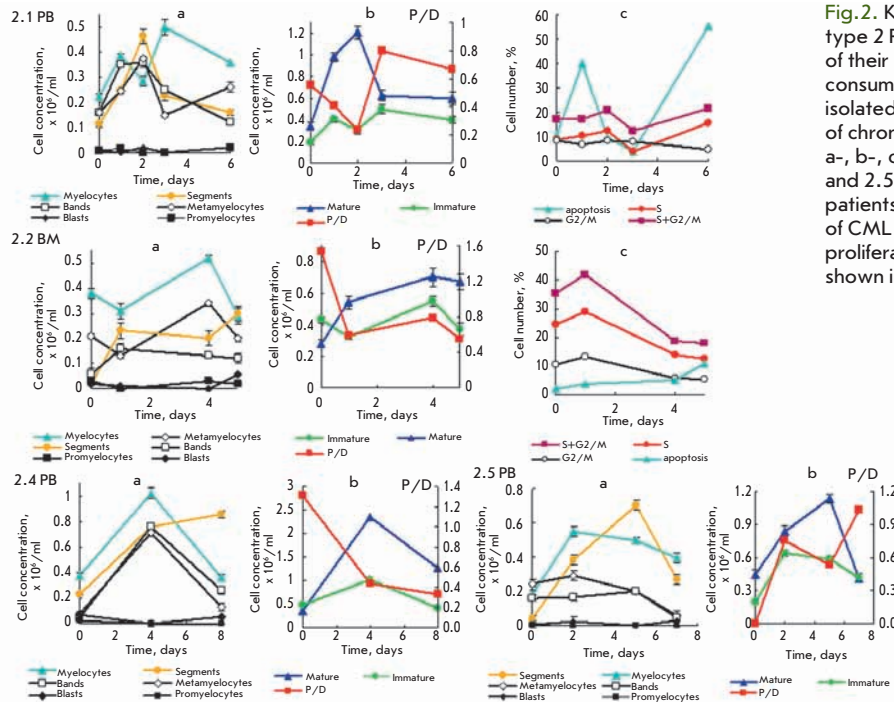


Fig.2. Kinetic plots of formation and transformation of type 2 Ph⁺ cells subpopulations as well as kinetics plots of their proliferation and maturation (as accumulating and consuming) in the suspension culture of mononuclears isolated from bone marrow (BM) or peripheral blood (PB) of chronic myeloid leukemia (CML) patients a-, b-, c- the same as in Fig. 1. 2.1 PB, 2.2 BM, 2.4 PB and 2.5 PB – samples of cells from corresponding CML CP patients. The prolonged observation didn't reveal signs of CML progression. Samples characteristics, as well as proliferation and maturation data of type 2 Ph⁺ cells, are shown in Table 1

Proliferation and differentiation of type 2 Ph⁺ cells is characterized by low efficiency (index $P/D \leq 1$) with the prevalence of the maturation rate over the proliferation rate and apoptosis blocking. P&D of type 2 Ph⁺ cells is associated with prominent neutrophils accumulation inhibiting the Ph⁺ cells proliferation (Fig.2, Table 1). Apoptosis inhibition during neutrophils maturation is not synchronous to myelocytes accumulation and induction of apoptosis. The decrease of the P/D index is associated with neutrophils accumulation; the increase of P/D, with depletion of segmented neutrophils (Figs.2, a b, c; 2.1 – 2.7, Table 1). The mature neutrophils concentration is significantly higher than that of proliferating Ph⁺ cells.

An important characteristic of proliferation is the distribution of cells in the cell cycle phases. The kinetic plots of type 2 Ph⁺ granulocytes distribution in the S and G2/M phases show that their changing during proliferation and maturation depends upon their source – CML patient. Figure .2.1, 2.2, c shows that the portion of proliferating Ph⁺ cells in phases $S+G2/M \leq 20-40\%$; and in G2/M phase, ~ 15%. Another proliferative index – $S/(G2+M)$ ratio is 1.2. The maximum percentage of cells in $S+G2/M$ at day 1 was 43 %; and in S phase, ~ 30%, rapidly declining to 12%. P/D index at the same time is 0.6-0.8 (Fig. 2.2 b). The percentage of proliferating types 1 and 2 Ph⁺ cells is similar (Fig. 2.2 c and Fig. 1.2 c). But the duration of elevated percentage of cells in $S+G2/M$ phases in type 2 Ph⁺ cells is shorter than in type 1, where the portion of proliferating cells $\geq 43\%$ is maintained for nearly 6 days with the $S/(G2+M)$ ratio ~2.4-2.7 and P/D efficiency 1.2-1.6. In terms of P&D of type 2 Ph⁺ cells, that means the prevalence of maturation over proliferation. Under cultivation, the percentage of proliferating malignant K562 cells (derivate of CML blastic phase) in $S+G2/M$ phases reaches $44 \pm 3\%$ with little changes

in the next 7 days. The $S/(G2+M)$ ratio at that time reaches 4.4. For comparison with these values, the P&D of type 2 Ph⁺ cells is characterized by a decreased percentage of cells in $S+G2/M$ phases (< 20 – 40%) with a shortened duration of proliferation (<3 days) and low P/D efficiency ≤ 1 (Figs.2.1 and 2.2, b, c).

The kinetic plots of type 2 Ph⁺ leucocytes differentiation show ~80% granulocytes content, while the percentage of myelocytes is not more than 25%. The percentage of mature neutrophils, especially segmented, reaches ~40% (Figs.2.1-2.5, b and 2.1-2.6 in Table 1). During P&D, all neutrophils are actively accumulated with little consumption. The maximal concentration of segmented neutrophils at times is twice that of myelocytes (Figs 2.1-2.5 a, b, c, and Table 1), emphasizing the role of mature cells (especially segmented neutrophils) in the regulation of P&D of type 2 Ph⁺ cells.

In summary, the population of type 2 Ph⁺ leucocytes increases by 2-4 times under low P/D efficiency and a proliferation rate lower than that of maturation. It is associated with the accumulation of neutrophils (especially segmented), the decrease of the myelocyte content, and apoptosis inhibition. It means that the excessive accumulation of neutrophils (due to apoptosis blockage) inhibits the proliferation of their own precursors and the differentiation of neutrophils (metamyelocytes, band and segmented), probably through a feedback mechanism.

The ratio $[S]/[M]$ is usually a small value; during proliferation and differentiation of type 2 cells, it increases, exceeds 1, and reflects the accumulation of segmented neutrophils, synchronous to the decrease of myelocytes accumulation. The neutrophils accumulation reaches a maximum at day 2 or 4 (Fig.2: a, b) that corresponds to a minimum accumula-

Table 3. The characteristics of crossings at alternation of proliferation and maturation stages during P&D of type 3 Ph⁺ cells in culture. See notes to Table 2

N Fig.	N CML patient	Stage alternation scheme	Crossings of accumulation rate plots for proliferating [immature] (stage 1) and maturing cells [mature] (stage 2)			Cell concentration in crossings, 10 ⁶ cells/ml							
			Time of crossing, Day		Stages duration, days		P/D indices in crossing points N			Mature and immature		Segmented (S)	
			1	2	1	2	1	2	3	1	2	1	2
1 Fig.3	3.1 M	1/2/1	2;	3,5	2;	1,5 2,5	1,1;	1,2;	-	0,5;	0,65	0,24;	0,34
2 Fig.3	3.2 PB	1/2	2,5;	-	2,5;	>4,5	1,06;	-	-	0,4;	-	0,1;	-
3	3.3 BM	1/2	2,5;	5,5	2,5;	8,5	~ 2;	-	-	0,49;	-	0,26;	-
4	PB	1/2			5,5;	5,5	1,0;	-	-	0,53;	-	0,38;	-
5 Fig.3	3.4 BM	1/2/1	0,2;	6	- 0,2;	5,8	1,15;	1,2;	-	0,37	0,8	0,17;	0,58
6	3.5 BM	1/2	5;	-	5;	3	1,0;	-	-	1,53;	-	0,37;	-
7	3.6 BM	1/2	4,5;	-	4,5;	3,5	1,06;	-	-	0,53;	-	0,24;	-
8	3.7 BM	1/2	2,5;	-	2,5;	8,5	1,5;	-	-	0,53;	-	0,4;	-
9	PB	1/2/1	2,5;	8,5	2,5;	8,5	1,2;	1,25;	-	0,54;	0,4	0,17;	0,15
10	3.8 BM	1/2	5;	-	5;	>2	1,5;	-	-	0,35;	-	0,15;	-
11 Fig.4	3.10 PB	2/1 ↔	3;	-	4;	3	1,4;	-	-	0,33;	-	0,3;	-
12 Fig.4	3.11 PB	2/1/2/1 ↔	5,5;	7,5; 11	2;	5,5 0; 3,5	1,15;	1,15;	0,80	0,75;	0,6 0,1	0,43;	0,3 0,06
13 Fig.4	3.12 BM	2/1/2/1/2 ↔	4,5;	6; 0 8	1,5;	4,5 0; 2,5 1	1,05;	1,0;	1,0	0,62;	0,72 0,24; 0,24	0,13;	0,35 0,11; 0,34
14	3.13 PB	2/1 ↔	4,0		1;	4	1,7;	-	-	0,4		0,2	
15	3.14 BM	2/1 ↔	4,0		1;	4	1,0;	-	-	0,6		0,4	

The mean value of P/D efficiency index at crossings of proliferation and maturation rate plots at alternation stages 1 and 2 is 1.06 ± 0.23 (21,7%). BM – bone marrow, PB- peripheral blood; brackets [] show cell concentration; hyphen - interval of values on the kinetic plots; Figure ↔ shows that 1st stages data for alternation 2/1 and 2/1/2/1 are given earlier than that of 2d stages according to the subtitle of the table.

tion of myelocytes and low P/D index. Figure 2 a, b shows the increase of the P/D index because of D cells concentration decrease at day 4.

The content of type 2 Ph⁺ cells is one-third of all studied samples; all of them isolated from CML chronic phase (CP) patients. The clinical observation of CML patients with type 2 P&D shows prolonged life duration in case of successful therapy.

Proliferation and differentiation of type 3 Ph⁺ cells takes place as repeated alternation of proliferation (stage 1) with maturation (stage 2) and changing of the P/D efficiency index from $P/D^1 > 1$ to $P/D^2 \leq 1$ etc. (Tables 2 and 3). At the first stages of alternation, the myelocytes accumulation under active apoptosis prevails over the accumulation of maturing neutrophils. Second stages of alternation are characterized by efficient maturation with apoptosis blockage and accumulation of enhanced concentration of neutrophils that inhibits further proliferation (Fig. 3, 4 and Tables 2 and 3). It is accompanied by an inversion of the sequence of neutrophils accumulation, as in the P&D of type 2 Ph⁺ cells (Figs. 2–4 a, b, c) considered below.

The transition to the next alternation stage 1 according to schemes 1/2/1 and 2/1/2 is associated with apoptosis induction, depletion of segmented neutrophils, and accumulation of immature Ph⁺ cells. Thus, at the crossing of the proliferation and maturation plots, the effective proliferation changes to effective maturation and vice versa; the characteristics

and parameters of the same (but already passed) stage are restored.

The P/D efficiency indices depend upon the alternation stage, as discussed above (Tables 2 and 3). At the first stages, index $P/D^1 > 1$; at next stages, $P/D^2 \leq 1$.

The alternation of the highly effective proliferation of Ph⁺ cells and relatively ineffective neutrophils maturation results in increasing and decreasing P/D efficiency (Table 3 and Figs. 3 and 4). It allows to maintain a moderate efficiency of P&D on the whole. It corresponds to alternating changing of type 3 Ph⁺ cells parameters (Tables 2 and 3). Taken together, these results mean alternation between the efficient proliferation and efficient maturation of the participation of segmented and other mature neutrophils in the regulation of type 3 Ph⁺ cells P&D through inhibition of Ph⁺ cells proliferation at the maturation stage. A similar regulation can be seen during P&D of type 2 Ph⁺ cells.

Alternating efficient proliferation and efficient maturation according to 1/2 scheme, beginning from stage 1 (Fig. 3.6, Table 2), have P&D characteristics compatible to type 1 Ph⁺ cells. The content of myelocytes and proliferating cells in S+G2/M phases increases, > 30–45%, at days 1-2 (Fig. 3), while the neutrophils content drops. At stage 2 of alternation, the segmented neutrophils concentration increases, the P/D² index drops ≤ 1 , the myelocytes content drops, and the percentage of Ph⁺ cells in S+G2/M phases becomes $\leq 25\%$. At the beginning of the next stage 1 – the proliferation stage that

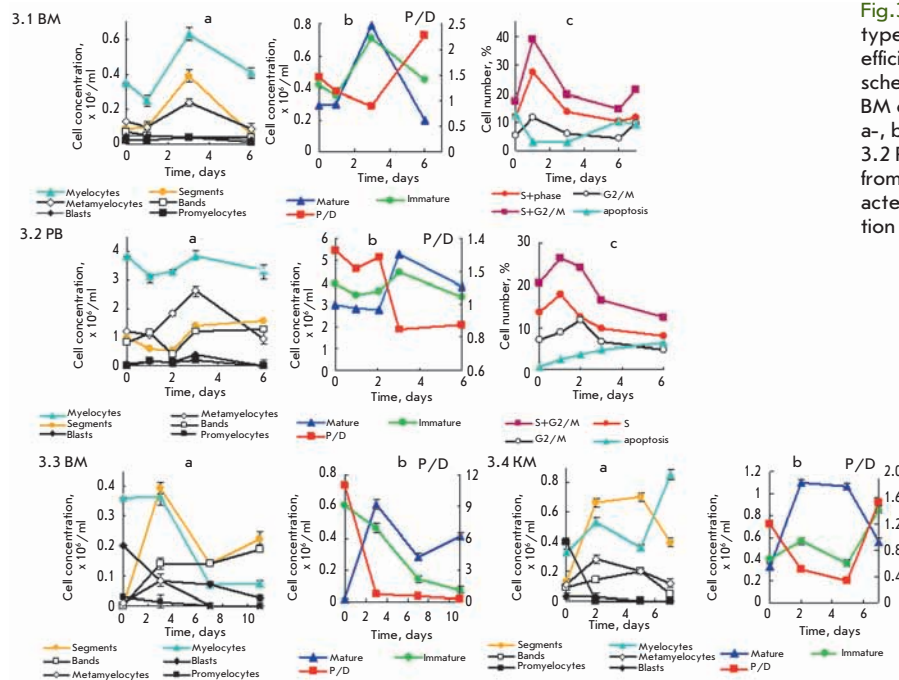


Fig.3. Kinetic plots of proliferation and maturation of type 3 Ph⁺ cells with alternating efficient proliferation and efficient maturation stages according to 1/2 or 1/2/1 scheme in the suspension culture of mononuclear from BM or PB of CML patients a-, b-, c- the same as at Fig. 1. Ph⁺ cell 3rd type 3.1 BM, 3.2 PB, 3.2 BM, 3.3 BM and 3.4 BM – samples of cells from corresponding CML CP patients. Parameters characteristics of efficient proliferation and efficient maturation for this Ph⁺ cell type are shown in Tables 2 and 3

happened at day 7 – the percentage of cells in the S+G2/M phases again begins to increase (Fig. 3.1 c) and other parameters of stage 1 restore. The alternation of stages 1 and 2 is analogous to the P&D of Ph⁺ cells types 1 and 2, correspondingly, but their duration is shorter. The overall duration of stage 1 of Ph⁺ cells type 3 is 4-5 days – 3 times as much as in its stage 2 (Table 3).

The second stage of alternation of the P&D of type 3 Ph⁺ cells – Ph⁺ cells corresponds to P/D² ≤ 1 values, high segmented neutrophils concentrations, apoptosis block, and minimal concentration of myelocytes (Fig. 4 a, b, c and № 3.10-3.14 in Table 2). Under apoptosis induction and decrease of segmented neutrophils concentration, accumulation of myelocytes and the P/D¹ index again begin to increase. The maximums of P/D¹ and myelocytes accumulation coincide with minimum segmented neutrophils concentration at day 5. The mature neutrophils accumulation leads to a drop in P/D¹, decrease of myelocytes concentration, and establishment of the efficient maturation stage. These properties are coherent with the characteristics of the P&D of type 2 Ph⁺ cells and other patient cells with P&D type 3, according to the 2/1/2 alternation scheme (Fig. 4 a, b, c; other samples data in Table 2).

Inhibition of apoptosis and myelocytes proliferation by segmented neutrophils is clearly seen in Figures 3 and 4 a, b, c, etc. in Table 2. At the same time, the more active the Ph⁺ cells proliferation and apoptosis, the higher is the P/D¹ index (for example, Fig.3 a, b, c, and Table 2).

The distinctive feature of type 3 Ph⁺ cells P&D according to the 2/1/2 scheme, beginning with stage 2, is in a prompt and significant accumulation of segmented neutrophils at the beginning of P&D and in prominent apoptosis blocking, especially Ph⁺ cells proliferation inhibition and distortion of matu-

ration. The apoptosis rate reaches a maximum at 2-5 days and equals ~35% (Fig.4, a-c), corresponding by P&D time to the depletion and complete death of segmented neutrophils. At the beginning, maximum cell proliferation at the maturation stage is due to the accumulation and depletion of segmented neutrophils; the second maximum of Ph⁺ cells proliferation is already observed due to transition to the efficient proliferation stage after the death of segmented neutrophils (Fig.4: 3.10, c).

The percentage of cells in particular cell cycle phases for 2/1 and 2/1/2 alternations is also linked to segmented neutrophils accumulation and corresponds to a decreased cell number in S+G2/M phases. In Figure 4, a-c, one can see maximums at day 1 ~10% (Fig.4: 3.10 BM), ~25 % (3.11 PB), and less than 5% (3.12 BM). At days 2-10, the portion of this pool significantly decreases. Under the next short proliferation stage at days 10-11, it slightly increases with a simultaneous increase of the P/D¹ index. Taken together with the decrease of P/D², it means the inhibition of Ph⁺ cells proliferation by neutrophils, more prominent with a more frequent alternation of prolonged maturation stages.

Figures 4: 3.12, a-c indicate that under alternation of 2/1/2/1 stages inhibition of proliferation proceeds with P/D² = 0.1-0.9. Under these alternations, the cell number in the S+G2/M phases is 5 % lower for eleven days. At low proliferation at the maturation stage, the apoptosis rate is less than 10%; at days 6-11, it increases only to 20% and only at the transition to stage 1. Overall, the time Ph⁺ cells № 3.12 stay under the effective maturation stage is three times longer than at stage 1. It leads to the most prominent inhibition of proliferation and apoptosis in type 3 Ph⁺ cells. In Figures 4: 3.11 a, b, c, maximum apoptosis corresponds to minimum

segmented neutrophils accumulation, maximum myelocytes accumulation, and P/D index rise. Type 3 of P&D was found in ~50% of the Ph⁺ cells studied (all from CML chronic phase patients). Among them, 2/3 constitute Ph⁺ cells with alternation of 1/2 and 1/2/1 stages.

Apoptosis block segmented neutrophils accumulation and inhibition of Ph⁺ cells types 2 and 3 proliferation by neutrophils during effective maturation. The maturing neutrophils differentiate in the sequence M→MM→B→S with consequent apoptosis of segmented neutrophils. The accumulation of maturing neutrophils decreases in the order S>B>MM [3]. If the transport of neutrophils to other tissues is impossible, they should die by apoptosis; therefore, the observed prominent S accumulation means the block of apoptosis.

The maturation of Ph⁺ cells type 2 and at 2nd alternation stages of Ph⁺ cells type 3 is associated with apoptosis blockage, elevation of mature neutrophils concentration, inhibition of myelocytes accumulation and decrease of the percentage of cells in S+G2/M and/or shortening of their staying in those phases. These events explain the inhibition of proliferation by neutrophils with blocking of apoptosis. It should be noted that apoptosis induction is usually observed under depletion of cytokines in cultural media [16–18, 25, 26, 29, 43–45]. During maturation stages of P&D for Ph⁺ cell, apoptosis is blocked without addition of cytokines in cultural media.

At S concentration ~2–3x10⁵ cells/ml and high S+G2/M cells content, the apoptosis rate is low (<5–10%); at day 5, it hardly reaches 10% (Fig. 2.2 a,c). Cell death in other cases (by trypan blue staining) is also inhibited. In cases 2.2–2.4 BM and 2.6 PB, the apoptosis level didn't exceed 10–20% (Table 1). As one can see in Figure 2.1 a, c, at day 1 of proliferation, the maximum apoptosis level reached 40% and 60% at day 6 with synchronous increase of myelocytes accumulation. It means that inhibition of myelocytes accumulation depends upon apoptosis blocking not directly but is rather mediated by the neutrophils accumulating due to the lack of apoptosis.

During P&D of type 2 Ph⁺ cells, segmented neutrophils are even capable of accumulating with a separate peak at days 2–6 (with [S] > 0,2–0,3x10⁶/ml), as can be seen in Figures 2 a, b. It corresponds to a low apoptosis level and minimum myelocytes accumulation (Figs. 2.1 b, c). The myelocytes accumulation plot is broken for the period of significant S accumulation (with a maximum at day 2), while the P/D index decreases. Myelocytic proliferation is inhibited until the segmented neutrophils concentration decreases due to apoptosis. After the death of a significant part of S, the pattern of myelocytes accumulation restores and the P/D index increases (Figs. 2.1 b), but it remains ≤1. The maximum of S accumulation (Figs. 2.1 a, b, c) at days 1 and 5 also corresponds to the minimum of myelocytes accumulation, lowering of P/D, and activation of apoptosis. A similar S accumulation (at days 2–5) inhibits myelocytes accumulation and decreases P/D, as can be seen in Figures 2.5 a, b. The death of S at day 5 restores the increase of myelocytes concentration and P/D value. Figures 2.2 a, b show that maximum S accumulation corresponds to minimum accumulation of myelocytes; i.e., maximum inhibition of myelocytes accumulation. At the same time, maximums of myelocytes accumulation and apoptotic activity in Figures

2.1 a, b, c coincide. The same rule applies during P&D of type 3 Ph⁺ cells.

Thus, the low P&D efficiency of Ph⁺ cells type 2 and type 3 at stages of effective maturation leads to prominent neutrophils accumulation, decrease of the proliferation rate, in comparison with the maturation rate, and decrease of Ph⁺ cells proliferation. Inhibition of Ph⁺ cells proliferation by neutrophils takes place under alternation of all stages of effective maturation and means the participation of neutrophils in the regulation of Ph⁺ cells P&D.

The regulation of P&D Ph⁺ cells types 2 and 3 consists of interdependent synchronous and asynchronous processes. Apoptosis inhibition with accumulation of neutrophils, maturing without dividing, proceeds asynchronously to myelocytes accumulation, apoptosis induction, and increasing efficiency. Myelocytes accumulation is synchronous with apoptosis induction, proliferation activation, and increase of P/D efficiency.

Inversion of mature neutrophils accumulation order. Neutrophils maturing without dividing (D cells) by definition differentiates in the order M→MM→B→S with consequent apoptosis of S [3, 40, 41, 45]. Under an equal differentiation rate, it can be estimated to have the same order of accumulation. In some cases – at P&D of type 1 Ph⁺ cells and effective proliferation stages in type 3 Ph⁺ cells – this order is actually observed.

During alternation at maturation stages of Ph⁺ cells types 2 and 3, maximum and minimum values in the kinetic plots of neutrophils accumulation point to a change of their accumulation order (Figs. 2, 3 and 4, a). It means that the accumulation rate of S, B and MM changes nonuniformly, and that their accumulation order is inverted. At that time, the accumulation of myelocytes decreases adjacent to the corresponding neutrophils peaks. Moreover, the order of neutrophils accumulation from the same Ph⁺ cells inverts during the maturation process, restoring after the transition to proliferation stages.

Figure 2, a shows that order changes with increase or decrease of the neutrophils accumulation rate are frequent. It can be illustrated by the crossings of the neutrophils accumulation order plots with the inversion of their direction. For example, the order M>B>MM~S, visible at day 1, consequentially inverts to S >MM~B>M at day 2, to M>B~S>MM at day 3, and to M>MM>S>B at day 6. In this figure the equal neutrophils accumulation rates after the crossing of plots at days 4–5 leads to an almost normal neutrophils accumulation order – M>>MM>S>B. In Figure 2:2.1, b one can see the usual accumulation order – M>MM>B>S – at day 0, but the inversion of this order leads to M>S>B~MM at day 1, then to M>>MM>S>B at day 4, and later to S accumulation with depletion of M, MM, and B. In Figure 2:2.4, a nearly equal accumulation rates of 3rd neutrophils at day 4 (M>B~S~MM) and at day 8 result in S>>M>B>MM. In Figure 2:2.5, a, accumulation order M>S>MM>B at day 2 transforms into S>M>>MM~B at day 5.

These results mean that the inversions of accumulation order, i.e. accumulation rate of neutrophils as a whole and their fractions, are frequent. The restoration of the usual neutrophils accumulation order corresponds to the increase of myelocytes content and P/D efficiency index and promotes

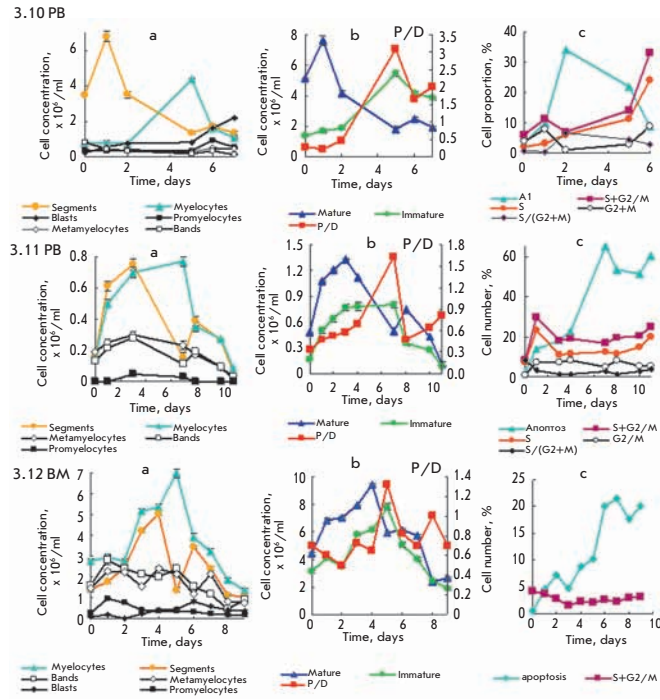


Fig. 4. Kinetics plots of proliferation and maturation of subpopulations of 3rd type Ph⁺ cells that perform by alternating efficient proliferation and efficient maturation stages according to 2/1 or 2 /1/2 schemes in the suspension culture of mononuclears from BM or PB of CML patients. Parameters and characteristics of efficient proliferation and efficient maturation of this Ph⁺ cell type, are in table 2 and 3

the transition to effective proliferation. Rate change and inversion of neutrophils accumulation order, synchronous to the decrease of P/D efficiency index (Figs. 1 and 4, d; 2, 3 and 4, b), points to the direct participation of neutrophils in the P&D regulation of Ph⁺ cells types 2 and 3 with inhibition of their proliferation.

Under maturation stage of Ph⁺ cells type 3 with alternation according to schemes 1/2, 1/2/1 in Figure 3, a the inversion of the accumulation order manifests as transformation from M>MM>S>B to M>S>MM>B and then again to M>MM>S~B. At alternations according to schemes 2/1 and 2/1/2, the inversion of neutrophils accumulation order is seen in Figure 4, a. The neutrophils concentrations are decreasing in row [S]>>[M]>[B]>[MM], as well as P/D index (Fig. 4, b).

Thus, there is inversion of order for consequent neutrophils maturation with their maximum neutrophils accumulation, whose initial order of neutrophils maturation is able to restore by transition to the alternation of the efficient proliferation stage. The extent of this inversion and restoration of the neutrophils accumulation order depends upon the duration of maturation and proliferation stages. Inversion of the accumulation order of M, MM, B, and S is another interesting property of Ph⁺ cells types 2 and 3 in culture. The rate

and extent of neutrophils accumulation is changed: at first it is S; then, B and MM.

Under effective maturation, the neutrophils accumulate in significant quantities due to apoptosis blocking (Fig. 2). They participate in the regulation of Ph⁺ cells types 2 and 3 P&D not only by inhibiting Ph⁺ cells proliferation, but also by inhibiting neutrophils maturation. Inhibition of neutrophils differentiation in Ph⁺ cells types 2 and 3 is revealed at inversion of the initial neutrophils accumulation order with the increase in their concentration (Fig. 2, 3 and Fig. 4, also in Tables 1-3). Inversion of the initial neutrophils accumulation order leads to gradual change of the neutrophils accumulation sequence M>MM>B>S to S>B>MM>M, and the increase of their concentration in the same order is assumed to proceed by a feedback mechanism. The present inversion of the accumulation order and increase of neutrophils concentration obviously are the result of a stepwise inhibition of the differentiation of every type of neutrophils (by a feedback mechanism) and consequently leads to their accumulation during differentiation of B and MM and, therefore, to inhibition (regulation disturbance) of neutrophils maturation. In other words, S accumulation inhibits the maturation of B, and it in turn inhibits MM maturation. Finally, all neutrophils maturation orders are inhibited, increasing their concentration. At the same time, neutrophils inhibit Ph⁺ cells proliferation and P&D efficiency. The next step, after reaching the “critical” neutrophils concentration, is characterized by apoptosis induction. It releases neutrophils from feedback, restores the initial neutrophils accumulation order, after which their concentration decreases, leading to restoration of the regulation of neutrophils accumulation, leading to increase of the P/D efficiency index and transition to effective proliferation of Ph⁺ cells.

The role of proliferation and maturation alternation. The alternation of effective proliferation and maturation with proliferation inhibitions that have different efficiencies and provide rate advantage to either the proliferation or maturation of neutrophils adds periodicity to this process. Efficient proliferation and myelocytes accumulation are synchronous to apoptosis induction, elevation of P/D¹ >1, but are asynchronous to neutrophils accumulation, inhibition of proliferation by them, inversion of the neutrophils accumulation order and changes of P/D² within ≤ 1 (Fig. 2-4, Tables 1-3).

It should be noted that in the regulation of Ph⁺ cells P&D, both synchronous and asynchronous processes could be observed. The apoptosis inhibition with neutrophils accumulation without dividing proceeds asynchronous to myelocytes accumulation and apoptosis induction. Myelocytes accumulation is synchronous to apoptosis induction, proliferation activation with increase of P/D index, and to restoration of maturation and proliferation regulation.

At the crossing points, P&D parameters equalize. During types 1 and 2 Ph⁺ cells P&D, its regulation at a given differentiation cycle becomes “irreversible.” Modification of type 3 Ph⁺ cells properties is reversible: during stages alternation the properties inhibited at the previous stage can be restored. By P&D regulation the type 3 Ph⁺ cells seems to be more “normal” than type 1 or 2. All types of Ph⁺ cells probably depend upon the inherited properties of the bcr/abl gene in different CML patients. Of note, the P&D regulation of Ph⁺ cells in all

three types of Ph⁺ cells from different CML patients differs in the quantitative characteristics of P/D efficiency, apoptosis inhibition, S accumulation, and the extent of Ph⁺ cells proliferation inhibition. Under P&D of Ph⁺ cells types 1 and 2, its P/D efficiency index varies in the limits $P/D^1 > 1$ or $P/D^2 \leq 1$, correspondingly (Fig. 1 and 2, Table 1). It points to some extent of P&D regulation of Ph⁺ cells in limits $P/D^1 > 1$ or P/D^2 without changing the initial advantage of the proliferation or maturation rate.

The differences in P&D regulation are related to the properties of Ph⁺ cells from different CML patients at different phases of the disease. Evidently, P&D efficiency is determined by myeloid cell precursors in Ph⁺ mononuclear cells fraction, inheriting the mutations in *bcr/abl* gene and p210 tyrosine kinase from different CML patients. The mutations capable of determining carcinogenicity in Ph⁺ cell lines, cell viability, activity of proliferation, and apoptosis blockage are studied in [7, 17, 18, 25, 27, 29, 30].

The alternation of Ph⁺ cells proliferation and maturation *ex vivo* with the rise or fall of P&D efficiency participates in the regulation and maintenance of an optimum regimen of Ph⁺ cells P&D. Under apoptosis blocking, it alternately switches inhibition of proliferation by maturing neutrophils, especially S, being the progeny of the same proliferating Ph⁺ cells. The cell regulation by segmented neutrophils probably mediates genetic regulation with participation of the *bcr/abl* gene. These results are concordant with evidence for the separate transforming and antiapoptotic activity of different mutant forms of p210^{*bcr/abl*} tyrosine kinase in cell lines that change signal transduction pathways [8, 9, 15, 17, 18, 20, 27, 30, 31].

As a result, alternation of effective proliferation and effective maturation with the change of increased efficiency indices to very low ones leads to a fall in the overall P&D efficiency of Ph⁺ cells types 2 and 3 and, thus, maintains a moderate, possibly optimum P&D regimen. At maturation stages, P&D regulation is mediated by the increase of neutrophils (especially S) content, proliferation inhibition with a decrease of P/D efficiency indices, and inversion of the neutrophils maturation order in Ph⁺ cells types 2 and 3.

DISCUSSION

The elicited P&D regulation in Ph⁺ cells *ex vivo* reveals the cellular aspects of genetic P&D regulation. As was shown in the present work, cellular regulation is mediated by alternation of Ph⁺ cells proliferation with efficient maturation and is accomplished by proliferation inhibition by maturing without dividing neutrophils – progenies of the same proliferating Ph⁺ progenitor cells under conditions of apoptosis blockage.

The cellular regulation of hematopoietic cells P&D and its proposed mechanism had previously never been considered in the literature [4-7, 9-12, 15, 16, 21-23, 27, 30]. There was no data on the cellular regulation of Ph⁺ cells P&D by mature neutrophils and on the alternation of stages with high and low P&D efficiency until this study. Also, there was no data on the relative proliferation and maturation rate, their efficiency, and its quantitative equivalent (P/D index). The ratio of proliferating hematopoietic cells and mature neutrophils in single PB samples was designated previously as the maturation index [2].

The proliferation of myeloid progenitor cells and the initial stage of Ph⁺ cells P&D regulation had been extensively studied [4, 5, 7, 8, 10-12, 15, 16, 20, 23-25, 26]. It is determined by the proliferative potential and concentration of CD34⁺ myeloid progenitor cells. Activation of Ph⁺ cells proliferation depends upon *bcr/abl* oncogene expression, its mutations, and higher rate of Ph⁺ stem cells transition from G₀ to G₁ cell cycle phase in comparison with cells without Ph chromosome [20, 23, 24]. Regulation of Ph⁺ cells proliferation is characterized to some extent (by the kinetic method of P&D investigation) also in our study. It is clear from the P&D parameters of type 1 Ph⁺ cells and alternations with effective proliferation stages in Ph⁺ cells type 3. The studies [33, 39] have shown the participation of CD34⁺ myeloid progenitor cells and *bcr/abl* expression in Ph⁺ cells from PB sample № 1.1, studied here in P&D regulation of type 1 Ph⁺ cells. According to this, a maximum level of CD 34 antigen and *bcr/abl* expression is coincident with a high, efficient P/D index at the beginning of Ph⁺ cells P&D.

The study [28] on Ph⁺ cells from CML CP patients isolated by flow cytometry has shown a direct relationship between myeloid Ph⁺ progenitors and *bcr/abl* oncogene expression. The authors have established a decrease in the proliferating Ph⁺ cells fraction in S+G2/M phases – by 4 times for myelocytes and by 7 times for B and S in comparison with CD34⁺ cells. A direct linear correlation of the Ph⁺ cells fraction in S+G2/M phases and *bcr/abl* expression was seen in CD34⁺ cells, myeloblasts, and promyelocytes. But later in the course of differentiation, the correlation inverted. It means, that active accumulation of proliferating Ph⁺ cells in S + G2/M phases at the beginning of the proliferation stage is proportional to the rise in *bcr/abl* expression, while during P&D from myelocytes to B and S this proportionality inverts. In this study we didn't estimate kinetic parameters. The data used was obtained from a single probe of each of the 15 CML CP patients. Our results provide a good explanation of this correlation inversion in study [28] by a change in the mechanism of Ph⁺ cells proliferation due to inhibition of Ph⁺ cells proliferation by the neutrophils accumulating under apoptosis blockage.

Taken together, the work of Primo *et al* [28] and our study indicate the paradoxical mechanism of *bcr/abl* action on Ph⁺ cells. The expression of *bcr/abl* in proliferating myeloid Ph⁺ progenitors activates proliferation, but later mature Ph⁺ neutrophils that inherit the same oncogene *bcr/abl* inhibit proliferation. In other words, the promoting proliferation effect of *bcr/abl* oncogene expression in Ph⁺ cells at early stages of P&D is inhibited by the expression of the same oncogene in the progenies of the same Ph⁺ cells at later stages of maturation. It can be explained, for example, by the change in the signal transduction pathways by *bcr/abl* oncogene, p210 tyrosine kinase, and spectrum of STAT proteins formed in the neutrophils or in one of them.

Some studies propose an alternative hypothesis on the development and progression of CML, considering the imbalance of stem cells self-renewal and further cell differentiation, rather than the activation of proliferation due to uncontrolled expression of p210^{*bcr/abl*}, as the main cause of CML [20-24, 13-14, 26]. It is associated with a decreased number of CML progenitor cells, while their progenies in CML are more

abundant in comparison with that in “normal” cells. Distortion of Ph⁺ cells maturation as a primary biological defect in CML was discussed elsewhere rather long ago [13,14,21, 22, 42]. The example of Ph⁺ cells types 2 and 3 P&D in our work shows that such an imbalance is mediated by the neutrophils inhibiting Ph⁺ cells proliferation at maturation stage under apoptosis blockage. The revealed differences in P&D regulation with inhibition of proliferation by neutrophils with apoptosis blockage and alternation of efficient proliferation and efficient maturation of Ph⁺ cells from different patients inheriting different bcr/abl mutations completely explain the contradictions regarding the imbalance of Ph⁺ cells self-renewal and maturation.

Proliferation inhibition by neutrophils during the alternation of proliferation and maturation under apoptosis blockage is assumed to regulate also P&D of hematopoietic cells in normal conditions. In CML, this mechanism can possibly be an attempt by CML cells to escape proliferation activation by bcr/abl expression, leading to CML progression. The threat of CML progression can be seen, for example, at P&D of type 1 Ph⁺ cells under apoptosis induction and absence of neutrophils excess.

CONCLUSIONS

1. The P&D regulation of Ph⁺ cells from different CML patients is performed by alternation of effective proliferation with effective maturation. In the maturation stage, the Ph⁺ cells proliferation is inhibited by accumulating neutrophils with apoptosis blockage interrupting proliferation and maintaining the optimum P&D level.

2. The alternation of stages consists in the switching of stage 1 – effective proliferation – when the proliferation rate

of Ph⁺ cells exceeds that of maturation, to stage 2 – effective maturation – when apoptosis is blocked, and the neutrophils maturation rate is higher than the proliferation rate. The alternation can proceed according to the schemes 1/2 -1/2/1 or 2/1-2/1/2. The alternation stages have control points where its plots cross and the indices of proliferation and maturation are equal. The indices of P/D efficiency (ratio of proliferation and maturation rates) at control points are equal to 1.06±0.23 and don't depend upon time and alternation order, as well as the source of Ph⁺ cells – CML patients. In alternation stages, these indices are permanently changing.

3. In the proliferation stages, the proliferating cells content rises, the number of neutrophils (especially S) decreases and apoptosis is activated. In the maturation stages, on the contrary, apoptosis is blocked, the neutrophils content (especially S) is enhanced, while the number of immature is decreased. In the effective maturation stage, accumulating neutrophils inhibit the proliferation of immature Ph⁺ cells and the order (sequence) of neutrophils maturation is inverted and impaired, probably via a feedback mechanism.

4. The differences in P&D regulation allowed us to identify 3 types of Ph⁺ cells in CML patients differing by the number and duration of alternation stages. There are Ph⁺ cells whose P&D has either one prolonged effective proliferation stage or one prolonged effective maturation stage, or multiple alternations of effective stages of proliferation and maturation. The first corresponds to advanced CML phases; the other takes place in the CML chronic phase with good response to chemotherapy. ●

This work was supported by the Russian Foundation for Basic Research (grant № 06 04-08372-ofi).

REFERENCES

- Abdulkadyrov K.M., Bessmeltsev C.C., Rukavitsin O.A. Chronic myelogenous leukaemia. S-Pb: Special literature, 1998. 463 pp.
- Haematological Guideline.- M. Newdiamed, Ed. A.I.Vorobiev. 2002. V.1, 280 pp
- Blood Patophysiology.- BINOM Publishers, Ed. Ph.D. Shiffman. 2000. 446 pp
- Deininger M.W.N., Goldman J.M., Melo J.V. // *Blood*. 2000. V. 96. P. 3343–3356.
- Deininger M.W.N., Vieira S., Mendiola R., et al. // *Cancer research*. 2000. V. 60. P.2049–2055.
- Medvedeva N.V. Chronic myelogenous leukaemia.- 50-th annual congress of American Haematological society. // *Clinical oncohaematology*, 2009. V. 2. №1. P. 85–88.
- Melo J.V. // *Blood*. 1996. V. 88. P. 2375–2384.
- Holyoake T.L., Jiang X., Eaves A.C., Eaves C.J. // *Leukemia*. 2002. V. 16. P. 549–558.
- Holyoake T.L., Jiang X., Jorgensen H.G. et al. // *Blood*. 2001. V. 97. P. 720–728.
- Jamieson C.H.M., Ailles L.E., Dylla S.J. et al. // *New England J Medicine*. 2004. V. 351. P. 657–667.
- Jaiswal S., Traver D., Miyamoto T. et al. // *Proc.Nat.Acad. Sci. USA*. 2003. V. 100. P. 10002–10007.
- Pasgue E., Jamieson C.H.M., Ailles L. E., Weissman I. L. // *Proc.Nat.Acad. Sci. USA*. 2003.V. 100. P. 11842–11849.
- Strife A., Lambek C., Wisniewski D. et al. // *Blood*. 1983. V. 62. P. 389–397.
- Strife A., Lambek C., Wisniewski, D. et al. // *Cancer Res*. 1988. V. 48. P. 1035–1041.
- Era T., Witte O. N. // *Proc. Nat. Acad. Sci. USA*. 2000. V. 97. P. 1737–1742.
- Guzman M.L., Jordan C.T. // *Cancer Control*. 2004. V. 11, #(№?)2. P. 97–104.
- Bedi A., Zehnbauser B.A., Barber J. et al. // *Blood*. 1994. V. 83. P. 2038–2044.
- Bedi A., Barber J. P., Bedi G.C. et al. // *Blood*. 1995. V. 86. P. 1148–1158.
- Brandford S., Rudzki Z., Walsh S. et al. // *Blood*. 2002. V. 99. P. 3472–3475.
- Buckle A.M., Mottram R., Pierce A. et al. // *Mol. Med*. 2000. V. 6. P. 892–902.
- Clarkson B., Strife A., Perez A. et al. // *Leukemia & Lymphoma*. 1993. V. 11. P. 81–100.
- Clarkson B., Strife A. // *Leukemia*. 1993. V. 7. P. 1683–1721.
- Coppo P., Bonnet M.L., Dusanter-Fourt I. et al. // *Oncogene*. 2003. V. 22(26). P. 4102–4110.
- Traycoff C.V., Haistead B., Rice S. et al. // *Brit. J. Haematology*. 1998. V. 102. P.759 – 767
- Lotem J., Sachs L. // *Leukemia*. 1996. V. 10. P. 925–9313.
- Lugo T.G., Pendergast A.M., Muller A.J., Witte O.N. // *Science*. 1990. V. 247. P. 1079–1082.
- Cortez D., Kadlec L., Pendergast A.M. // *Mol. Cell. Biology*. 1995. #(№?)10. P. 5531–5541.
- Primo D., Flores J, Quijano S. et al. // *Brit.J. Haematology*. 2006. V. 135. P. 43–51.
- Amarante-Mendes G.P., Naekyung Kim C., Liu L. et al. // *Blood*. 1998. V. 91. P. 1700–1705.
- Selleri C., Maciejewski J.P., Pane F., et al. // *Blood*. 1998. V. 92. P. 981–989.
- Sherbenou D.W., Hantschel O., Turaga L. et al. // *Leukemia*. 2008. V. 22. P. 1184–1190.
- Stoklosa T., Poplawski T., Koptyra M., et al. // *Cancer Res*. 2008. V. 68. P. 2576–2580.
- Akhynina T.V., Gerasimova L.P., Sarkisyan G.P. et al. // *Cytology*. 2007. V. 49. P. 889–900.
- Abramov M.G. Haematological album. M: *Medicine*, 1985.344 pp.
- Gerasimova L.P., Manakova T.E., Akhynina T.V. et al. // *Russian Journal of Biotherapy*. 2002. V.1, № 4. P. 29–38
- Pinegin B.V., Yiarilin A.A., Simonova A.V. et al. // Apoptosis evaluation of human peripheral blood activated lymphocytes by cytofluorimetric method with propidium jodide // In: Use of flow cytofluorimetry for estimation of human immune system functional activity. M, *MH. RF*, 2001. P. 48–53
- Shmarov D.A., Kosinets G.I. Methods of blood cycle analysis by flow cytofluorimetry // In: Laboratorial-clinical meaning of blood analysis by flow cytofluorimetry. M.: Medical Informational Agency, 2004.P. 49–65
- Dean P.N. // *Cell Tissue Kinet*. 1980. V. 13. P. 299–302.
- Grineva N.I., Barishnicov A.Ju., Gerasimova L.P. et al. // *Russian Journal of Biotherapy*. 2007. V.6, № 2. P. 21–32
- Kosinets G.I., Kotelnicov V.M. // *Soviet Medicine*. 1983. № 4. P. 3–77
- Kotelnicov V.M., Kosinets G.I., Kasatkina V.V., Kovalevskaya N.P. Kinetics of granulocytopoiesis. // In: Kinetic aspects of haemopoiesis. *Tomsk State University*. 1982. P. 149–211.
- Golde D.W., Cline M.J. // *New Engl. J. Med*. 1973. V. 288. P. 1083–1086.
- Vladimirskaya E.B. Mechanisms of blood cells apoptosis. // *Laboratorial Medicine*. 2001. № 4. P. 47–54.
- Vladimirskaya E.B. Apoptosis and its role in the development of tumor growth. // In: Biological basis of antitumour therapeutics. M.: *Agat –Med*, 2001. P. 5–32
- Dublez L., Eymen B., Sordet O. et al. // *Blood*. 1998. V. 91. P. 2415–2422.
- Goldman J. M., Th'ng K.G., Catovsky D., Galton D.A.D. // *Blood*. 1976. V. 47. P. 381–388.

Deamination of 5-Methylcytosine Residues in Mammalian Cells

E.V. Gromenko¹, P.V. Spirin², E.A. Kubareva³, E.A. Romanova³,
V.S. Prassolov², O.V. Shpanchenko^{1, 3, *}, O.A. Dontsova^{1, 3, *}

¹Department of Chemistry, Lomonosov Moscow State University

²Engelhardt Institute of Molecular Biology, Russian Academy of Sciences

³Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University

*E-mail: olgash@genebee.msu.ru

ABSTRACT DNA demethylation in mammalia occurs after fertilization and during embryogenesis and accompanies cell aging and cancer transformation. With the help of the primer extension reaction, MALDI MS and DNA cleavage by thymine DNA glycosylase deamination of 5-methylcytosine residues has been shown to take place when the model methylated DNA duplexes are treated with nuclear extracts from the cell lines CHO, HeLa, and Skov3. The hypothesis that deamination of 5-methylcytosine is the first stage of demethylation in mammalia has been postulated.

Keywords: deamination of 5-methylcytosine, 5-methylcytosine demethylation, the reaction of primer extension, MALDI MS, thymine DNA glycosylase.

Abbreviations: mC – 5-methyl-2'-deoxycytidine, PCR – polymerase chain reaction, PAAG – polyacrylamide gel, TDG – thymine DNA glycosylase, MALDI MS – matrix-assisted laser desorption/ionisation mass spectrometry.

INTRODUCTION

Active DNA demethylation in mammalia is necessary for the proper development of the organism, formation of the immune response, and memory. DNA demethylation accompanies the emergence of various diseases and aging.

Global demethylation is observed in the paternal pronuclei in the embryos of mice [1], rats, pigs, cows, and humans [2], as well as in their germ cells. It provides epigenetic reprogramming and specific gene expression.

Methylation and demethylation of DNA in the cells of the nervous system has an influence on synaptic plasticity and memory formation. DNA methylation is necessary for the inactivation of the gene *pp1* that suppresses memory processes; whereas active DNA demethylation is associated with activation of the gene *reelin* that promotes the formation of memory in rats [3]. Also, active DNA demethylation is necessary for normal neurogenesis in the embryos of fish *Danio rerio*. Suppression of the expression of protein Gadd45a (DNA-damage-inducible protein 45 alpha) or other proteins involved in the process of demethylation leads to the death of neurons as a result of hypermethylation and suppression of the genes responsible for neurogenesis [4].

Methylation and demethylation of DNA is an important part of epigenetic control during the immune response [5]. Demethylation of the promoter regions of the genes of cytokines *il-2* and *ifn-γ* as a result of CD8 T cells contact with antigens leads to rapid expression of cytokines [6–8].

DNA demethylation accompanies cell aging [9]. A difference in the degree of demethylation in different tissues has been shown in rats. The degree of demethylation is higher in brain tissues than in liver tissues. Age-related decrease in the content of 5-methylcytosine in DNA was also found in such cells as lung and skin fibroblasts: for the latter case, the connection of demethylation to reduced growth potential in culture was shown [10].

Global demethylation of the genome is observed in all examined cancer cells [11–13]. Hypomethylated sites turned out

to be numerous repeats, imprinted genes, tissue-specific genes, oncogenes, and genes associated with the processes of invasion and metastasis of tumors [14, 13]. At the same time, some loci, including many tumor-suppressor genes, are hypermethylated, which results in suppression of their expression [12, 15, 16].

It is clear that DNA demethylation plays an important role in the life of cells, while the mechanism of DNA demethylation and the participants in this process in mammalia have yet to be identified. The aim of this work was to study the mechanism of DNA demethylation in mammalian cells.

MATERIALS AND METHODS

Synthesis of oligodeoxyribonucleotides. Oligonucleotides I–III were synthesized using the amido-phosphite method in automatic mode on a synthesizer manufactured by Applied Biosystems, with commercial reagents and solvents according to the standard protocol:

5'-CATGTCTAACmCGmCGmCGAGAAATGGTAATGTATGGAGT* (I)

5'-CATACATTACCATTCTmCGmCGmCGGTTAGACATGGC* (II)

5'-CATACATTACCATTCT (III), where * is the amino group bound to the 3'-end of oligodeoxyribonucleotides by the linker.

Oligonucleotides

5'-TT(Biotin-T)TTTTTTTGTCTACGATCGAACmCGmCGmCGAGAAGCTTGAT* (IV), 5'-ATACAAGCTTCTmCGmCGmCGGTTTCGATCGTAGACAAAAA* (V) – commercial preparations (“Synthol”).

Cultivation of cell lines and preparation of nuclear extracts.

Continuous cell lines of Chinese hamster ovary (CHO), a human cervical carcinoma (HeLa), and human ovarian adenocarcinoma (Skov3) were grown in a standard DMEM medium containing 10% fetal serum (FCS), 4 mM L-glutamine, 1 mM sodium pyruvate, a streptomycin/penicillin concentration of 100 μg/ml and 100 U/ml, respectively, at 37° C in an atmosphere of 5% CO₂. For subculturing, the cell monolayer was

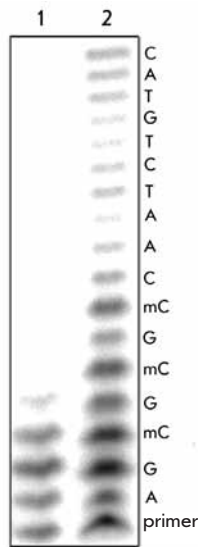


Fig. 1. Primer extension reaction without dGTP for methylated oligodeoxyribonucleotide duplex treated with buffer for extracts (control) (1), or nuclear extract from CHO cell line (2). The partial sequence of model oligodeoxyribonucleotide I is shown on the right

washed with a PBS buffer (10 mM Na_2HPO_4 , 1.5 mM KH_2PO_4 , 137 mM NaCl, 2.7 mM KCl, pH 7.4), a standard solution of trypsin-EDTA (Sigma) was added, and culture bottles were placed in a CO_2 incubator for 3-5 minutes. Then, a medium with FCS was added, and cells were suspended by pipetting. The cells were then placed in the necessary number of culture bottles. The cells grown to the monolayer were collected by centrifugation at 2300 rpm for 10 minutes at 4°C . They were then washed several times with a PBS buffer and lysed in buffer L (20 mM Hepes (pH 7.6), 10 mM NaCl, 1.5 mM MgCl_2 , 20% glycerol, 0.1% Triton X100, 1 mM DTT, cocktail of protease inhibitors). After 2 minutes, the lysate was suspended and centrifuged for 20-30s at 10,000 rpm. An equal volume of buffer NE (20 mM Hepes (pH 7.6), 500 mM NaCl, 1.5 mM MgCl_2 , 20% glycerol, 0.1% Triton X100, 1 mM DTT, a cocktail of protease inhibitors) was added to the pellet, then it was carefully suspended and placed on a shaker for 30-60 minutes. Then, lysate was centrifuged for 10 minutes at 10 000 rpm.

Formation of DNA-duplex and treatment with nuclear extracts from cells. Formation of DNA-duplex I/II proceeded on a PCR thermal cycler by keeping 20 μl of the reaction mixture containing 500 pmol of oligonucleotide I and 550 pmol of oligonucleotide II at 95°C for 3 minutes and then cooling the thermostat from 95 to 35°C at a speed of $0.5^\circ/\text{min}$ and from 35 to 20°C at a speed of $0.25^\circ/\text{min}$.

Then, 50 μl of the reaction mixture containing buffer NE, 10 mM of DNA-duplex I/II, and the nuclear extract corresponding to ~ 50000 cells was incubated at 37°C for 1 hour. After a three-fold phenol deproteinization, the DNA duplex was precipitated with 2.5 volumes of ethanol, with the addition of a 1/10 volume of 3 M NaOAc (pH 5.5) by keeping for 2 hours at -20°C . The pellet was precipitated by centrifugation at 14000 rpm for 10 minutes, washed with 70% ethanol, dried in vacuum, and dissolved in 20 μl of water.

Primer extension reaction. A radioactive label was introduced at the 5'-end of primer III in advance. To do this, a 20 μl reaction mixture containing a buffer solution for T4 polynucleotide kinase (500 mM Tris-HCl (pH 7.6), 100 mM MgCl_2 , 50 mM

DTT, 1 mM spermidine, 1 mM EDTA, 1 mM ADP), 10 mM of primer III, 5 units of T4 polynucleotide kinase (MBI Fermentas, 10 units/ μl) and 10 mM [$\gamma\text{-}^{32}\text{P}$] ATP (GE Healthcare, the specific activity of 220 TBq/mmol or 6000 Ci/mmol) was incubated at 37°C for 1 hour. The enzyme was heat-inactivated for 15 minutes at 75°C . The ^{32}P -labeled primer III was purified by electrophoresis in denaturing 12% PAAG. The area containing primer III with the radioactive label was visualised using autoradiography and excised. Primer III was eluted from the gel by adding 400 μl of a GES buffer (10 mM Tris-HCl (pH 8.0), 1% SDS, 0.5% Triton X100, 50 mM EDTA) supplemented with 400 μl of phenol with vigorous stirring overnight. The aqueous phase was separated by centrifugation at 14000 rpm for 5 minutes, primer III was precipitated as described above and was dissolved in 20 μl of water. The amount of primer III was evaluated either by radioactivity or spectrophotometrically by the absorption of the solution at 260 nm. After that, 4 μl of the solution of DNA-duplex I/II (2.5 mM) and 1 μl of a 10X reaction buffer (500 mM Tris-HCl (pH 8.0), 50 mM MgCl_2 , 10 mM DTT) were added to 3 μl of a purified ^{32}P -labeled primer III (3.3 mM). Hybridization was performed on the PCR thermal cycler by cooling the thermostat from 70 to 42°C at a speed of $0.4^\circ/\text{min}$ with keeping a constant temperature for 21 seconds and increasing the time of every subsequent cycle by 1 second. Then, 5 units of Klenow fragment without exonuclease activity (MBI Fermentas, 10 units/ μl) and 1 μl of a mixture of dATP, dCTP, dTTP (1 mM each), and ddNTP (0.01 mM) were added to the reaction mixture. The mixture was incubated at 37°C for 10 minutes. The reaction was stopped by heating to 75°C for 10 minutes. Analysis of the products of the primer extension reaction was carried out by electrophoresis in denaturing 10% PAAG. The gel was dried in vacuum, exposed to the screen of a BAS CASSETTE 2340, and information from the screen was read on the FUJIFILM FLA 3000 using the BASReader 3.14 computer program.

Treatment of DNA duplexes by thymine DNA glycosylase (TDG). A 20 μl reaction mixture containing 0.5 mM of DNA duplex I/II, ^{32}P -labeled at one of the DNA strands, 10 units of TDG (R & D Systems, 5 U/ μl), and the buffer (10 mM Hepes (pH 7.4), 100 mM KCl, 10 mM EDTA) was incubated at 65°C for 1 hour. Then, a 10 μl of 3-fold alkaline buffer (300 mM NaOH, 97% formamide, 0.2% bromophenol blue) was added to the reaction. It was incubated at 95°C for 10 minutes and then rapidly cooled to $2-8^\circ\text{C}$. The reaction products were analyzed by electrophoresis in 20% PAAG. The gel was dried using a vacuum dryer and exposed to the screen of a BAS CASSETTE 2340. Information from the screen was read on the FUJIFILM FLA 3000 using the BASReader 3.14 program.

Analysis of DNA fragments using mass spectrometry. The biotinylated DNA duplex IV/V was formed as described above for the duplex I/II. Then, it was treated with nuclear extracts from cells, and purification on streptavidin-Sepharose was performed. Streptavidin-Sepharose was pre-equilibrated in buffer R (100 mM Tris-HCl (pH 7.5), 10 mM MgCl_2 , 100 mM KCl, 0.1 mg/ml BSA). For that, 100 μl of a 50% suspension of streptavidin-Sepharose was mixed with 200 μl of buffer R and incubated with weak stirring for 10-15 minutes. Streptavidin-Sepharose was precipitated by centrifugation at 3000 rpm for 3-4 minutes, the buffer was replaced, and beads were resuspended. The procedure described above was repeated

5 times. A 500 pmol of biotinylated DNA duplex I/V was added to the equilibrated Sepharose. The binding of the duplex with Sepharose was carried out under mild stirring at 4° C for 2-12 hours. Then, the resin was washed 6 times by 1 ml of buffer R, as described above. After that, 30 units of restriction endonucleases HindIII (MBI Fermentas, 10 U/μl) and PvuI (MBI Fermentas, 10 U/μl) were added to the reaction mixture and it was kept with low stirring at 37° C for 16 hours. Streptavidin-Sepharose was sedimented by centrifugation at 5000 rpm for 5 minutes. The obtained fragments were separated by electrophoresis in denaturing 20% PAAG. The gel was stained with SYBR Green in a buffer solution TBE (100 mM Tris-HCl, 100 mM H₃BO₃, 2 mM EDTA). The analyzed fragment was eluted from the gel using the method described for the primer. Then, samples were concentrated and desalted on ZipTip C18 (Millipore). The resin was washed with a 50 mM aqueous solution of ammonium citrate, and the oligonucleotide was eluted with a solution of 25 mM ammonium citrate in 50% acetonitrile. On the matrix for MALDI MS 1 μl of the sample, 0.5 μl of a 50 mM solution of ammonium citrate and 0.5 μl of 3-hydroxypicolinic acid (Fluka, 20 mg / ml in acetonitrile) were mixed, then the mixture was dried in air. Mass spectra were obtained on the MALDI TOF/TOF mass spectrometer Ultraflex II BRUKER (Germany), equipped with a UV laser (Nd). Mass spectra were obtained in the mode of positive ions in linear fashion.

DISCUSSION OF RESULTS AND CONCLUSIONS

To study the process of DNA demethylation, complementary oligodeoxynucleotides I and II, each containing 3 residues of 5-methylcytosine in the context of dinucleotide 5'-mCpG, which is typical for the mammalian genome, were designed and synthesized. For the protection against the nucleases

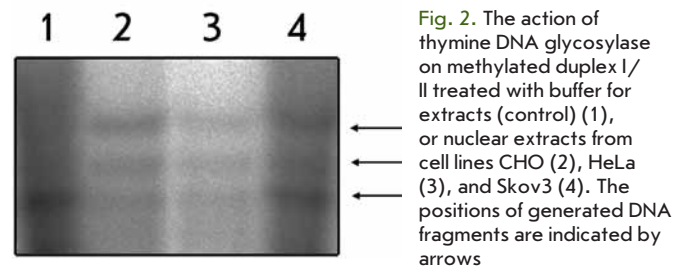


Fig. 2. The action of thymine DNA glycosylase on methylated duplex I/II treated with buffer for extracts (control) (1), or nuclear extracts from cell lines CHO (2), HeLa (3), and Skov3 (4). The positions of generated DNA fragments are indicated by arrows

present in the cell lysates 3'-protruding ends of DNA-duplex I/II carried protective groups.

The extension of primer III was used to monitor changes in the nucleotide sequence. Primer III is complementary to the 3'-terminal region of oligonucleotide I. In the absence of dGTP in the medium and in the presence of ddNTP, the primer had to be extended up to the first 5-methylcytosine residue (lane 1, Fig. 1). After the treatment of DNA-duplex I/II with nuclear extracts from the CHO cell line longer primer-extension products appear (lane 2, Fig. 1). This phenomenon could be explained by the conversion of 5-methylcytosine into thymine; i.e., the deamination process takes place when DNA-duplex I/II is treated with the extract of CHO cells.

To test this hypothesis, DNA-duplex I/II with three 5'-mCpG sites was sequentially treated with the nuclear extract of cell lines (CHO, HeLa or Skov3) and thymine DNA glycosylase, which recognizes T/G-mismatch and cuts the DNA strand containing T (Fig. 2, lanes 2, 3, 4, respectively). After the separation of reaction products, new bands corresponding to the position of 5-methylcytosine residues could be detected by autoradiography. That indicates the deamination of 5-methylcytosine residues.

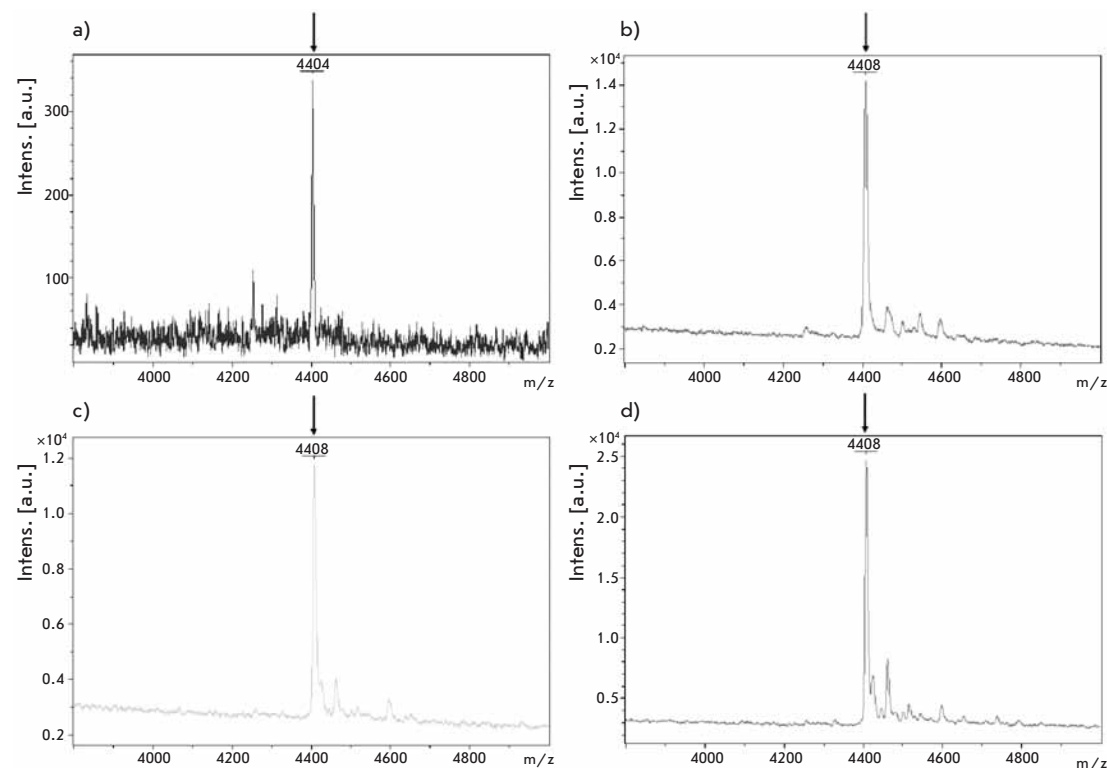


Fig. 3. MALDI mass-spectra obtained for methylated DNA fragment treated with buffer for extracts (control) (a), or nuclear extracts from cell lines CHO (b), HeLa (c), Skov3 (d)

Deamination of DNA duplexes after the treatment with the nuclear extracts of the cells was shown by MALDI MS. For that, we used oligodeoxynucleotide duplex IV/V containing the dinucleotides 5'-mCpG, endonuclease HindIII and PvuI recognition sites, flanking the analyzed fragment (underlined), as well as biotin at the 5'-end for the affinity isolation of the DNA duplex on streptavidin-Sepharose.

The size of the fragment containing three 5'-mCpG sites is 14 nucleotides; its calculated molecular mass is 4404.8 Da (Fig. 3, a).

The spectra of the analyzed fragment derived from the DNA duplex after the treatment with nuclear extracts from cell lines (CHO, HeLa, or Skov3) have a signal at 4408 Da (Fig. 3, b-d), corresponding to deamination of three 5-methylcytosine residues.

Cell lines CHO, HeLa, and Skov3 are derived from cancer cells in which global demethylation of the genome is known to occur. One can suggest that, the deamination of 5-methylcytosine observed as a result of the treatment with nuclear extracts of these cell lines is the first stage of the DNA demethylation in mammalia. It is known that in *Danio rerio* embryos DNA demethylation is a multistep process involving deamination of 5-methylcytosine with the deaminases AID (Activation Induced deaminase)/Apobec (Apolipo-protein B RNA-editing catalytic component), removal of thymine by glycosylase MBD4, and BER-repair [4, 17]. In mammalian cells, there are suitable candidates from the cytosine deaminases family – AID and Apobec1, which are co-expressed with pluripotent genes in oocytes, embryonic germ cells, and embryonic stem cells [18]. However, it is known that these enzymes perform deamination only in single-stranded RNA (Apobec1) or DNA (AID). In addition, AID and Apobec deaminate cytosine residues much more effectively than 5-methylcytosine. DNA methyltransferase Dnmt3b can also perform the role of 5-methyl cytosine deaminase when the concentration of S-adenosylmethionine (AdoMet) is low [19], but this reaction takes place with very low efficiency, making its participation in the global demethylation unlikely. Moreover, the concentration of AdoMet in the cell is usually quite high.

A good candidate for the role of thymine glycosylase is MBD4 glycosylase, which contains methyl-binding and glycosylase domains and removes of thymine from T/G-mismatch [20]. Although MBD4 deficient mice are viable, mutations in the CpG-sites occur more frequently in such mice [21]. MBD4 removes thymine from T/G-mismatch with the formation of

apurine site that is immediately cut by AP endonuclease. In the symmetrically methylated CpG-sites, the deamination of both DNA strands should lead to the formation of the TG/GT region. If MBD4 recognizes such a region then, after the action of AP endonuclease, a double break should form, which would lead to the loss of genetic material. In addition, MBD4 removes uracil from U/G-mismatch faster than thymine from T/G [20].

Another mechanism of DNA demethylation is also known. It is observed in plant *Arabidopsis thaliana* and involves the direct removal of the 5-methylcytosine base using bifunctional glycosylases/liases (ROS1, DML2, DML3, DME) followed by BER-repair [22-25]. There are two suitable glycosylase in mammalia: thymine DNA glycosylase (TDG) and methyl-binding protein MBD4. However, both TDG and MBD4 have weak 5-methylcytosine glycosylase activity compared to the ability to excise thymine [26]. In addition, active demethylation of paternal chromatin was observed in mouse embryos with a knocked-out *mbd4* gene [21].

It is worth mentioning other theoretically possible mechanisms of 5-methylcytosine demethylation. This is the direct removal of methyl groups with the formation of cytosine and excision of one or more nucleotides containing 5-methylcytosine (NER-repair). In vertebrates, there are orthologs of bacterial demethylase (oxidoreductase) AlkB responsible for direct removal of methyl groups from 1-methyladenine and 3-methylcytosine in prokaryotes [2]. However, there is no data indicating that any of the orthologs can carry demethylation of 5-methylcytosine. Histone demethylases (HDMs) are homologous to bacterial AlkB [27], but none of them is involved in DNA demethylation. Earlier, it was shown that methyl-binding protein MBD2b can directly remove the methyl group of 5-methylcytosine with the formation of C and methanol as the reaction products [28]. However, this result could not be reproduced in other laboratories. Moreover, mice lacking MBD2b had a normal phenotype and normal pattern of DNA methylation. As for the mechanism of DNA demethylation by NER-repair, there is no evidence of the occurrence of the process *in vivo*.

Thus, further study of DNA demethylation in mammals is necessary to understand the mechanism of this essential life process. ●

The work was supported by the Russian Foundation for Basic Research (grant № 07-04-00545) in part and the Charity Fund “Science for the Extension of Life.” The authors are grateful to Prof. M.V. Serebryakova for the MALDI MS analysis.

REFERENCES

- Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. // Nature. 2000. V. 403. P. 501–502.
- Morgan H.D., Santos F., Green K., Dean W., Reik W. // Hum. Mol. Genet. 2005. V. 14. P. 47–58.
- Miller C.A., Sweatt J.D. // Neuron. 2007. V. 53. P. 857–869.
- Rai K., Huggins I.J., James S.R., et al. // Cell. 2008. V. 135. P. 1201–1212.
- Reiner S.L. // Hum. Mol. Genet. 2005. V. 14. P. 41–46.
- Bruniquel D., Schwartz R.H. // Nat. Immunol. 2003. V. 4. P. 235–240.
- Kersh E.N., Fitzpatrick D.R., Murali-Krishna K., et al. // J. Immunol. 2006. V. 176. P. 4083–4093.
- Northrop J.K., Thomas R.M., Wells A.D., Shen H. // J. Immunol. 2006. V. 177. P. 1062–1069.
- Vanyushin B.F., Nemirovsky L.E., Klimenko V.V., Vasiliev V.K., Belozersky A.N. // Gerontologia. 1973. V. 19. P. 138–152.
- Wilson V.L., Jones P.A. // Science. 1983. V. 220. P. 1055–1057.
- Ehrlich M. // Oncogene. 2002. V. 21. P. 5400–5413.
- Feinberg A.P., Ohlsson R., Henikoff S. // Nat. Rev. Genet. 2006. V. 7. P. 21–33.
- Wilson A.S., Power B.E., Molloy P.L. // Biochim. Biophys. Acta. 2007. V. 1775. P. 138–162.
- Kisseljova N.P., Kisseljov F.L. // Biochemistry (Mosc). 2005. V. 70. P. 743–752.
- Jones P.A., Baylin S.B. // Nat. Rev. Genet. 2002. V. 3. P. 415–428.
- Jones P.A., Laird P.W. // Nat. Genet. 1999. V. 21. P. 163–167.
- Barreto G., Schafer A., Marhold J., et al. // Nature. 2007. V. 445. P. 671–675.
- Morgan H.D., Dean W., Coker H.A., Reik W., Petersen-Mahrt S.K. // J. Biol. Chem. 2004. V. 279. P. 52353–52360.
- Metivier R., Gallais R., Tiffocche C., et al. // Nature. 2008. V. 452. P. 45–50.
- Hendrich B., Hardeland U., Ng H.H., Jiricny J., Bird A. // Nature. 1999. V. 401. P. 301–304.
- Millar C.B., Guy J., Sansom O.J., et al. // Science. 2002. V. 297. P. 403–405.
- Gehring M., Huh J.H., Hsieh T.F., et al. // Cell. 2006. V. 124. P. 495–506.
- Gong Z., Morales-Ruiz T., Ariza R.R., et al. // Cell. 2002. V. 111. P. 803–814.
- Morales-Ruiz T., Ortega-Galisteo A.P., Ponferrada-Marin M.I., et al. // Proc. Natl. Acad. Sci. USA. 2006. V. 103. P. 6853–6858.
- Penterman J., Zilberman D., Huh J.H., et al. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. P. 6752–6757.
- Cortazar D., Kunz C., Saito Y., Steinacher R., Schar P. // DNA Repair (Amst). 2007. V. 6. P. 489–504.
- Ozer A., Bruick R.K. // Nat. Chem. Biol. 2007. V. 3. P. 144–153.
- Bhattacharya S.K., Ramchandani S., Cervoni N., Szyf M. // Nature. 1999. V. 397. P. 579–583.

Study of Molecular Mechanisms Involved in the Pathogenesis of Immune-Mediated Inflammatory Diseases, using Psoriasis As a Model

E.S. Piruzian^{1*}, V.V. Sobolev¹, R.M. Abdeev², A.D. Zolotarev¹, A.A. Nikolaev¹, M.K. Sarkisova¹, M.E. Sautin¹, A.A. Ishkin¹, An.L. Piruzyan², S.A. Ilyina², I.M. Korsunskaya², O.Y. Rahimova³, S.A. Bruskin¹

¹ Vavilov Institute of General Genetics, Russian Academy of Sciences

² Center for Theoretical Problems of Physicochemical Pharmacology, Russian Academy of Sciences

³ Moscow Municipal Hospital № 24, Department of Health

* E-mail: eleopiru@vigg.ru

ABSTRACT Psoriasis was used as a model to analyze the pathogenetic pathways of immune-mediated inflammatory diseases, and the results of bioinformatic, molecular-genetic and proteomic studies are provided. Cell mechanisms, common for the pathogenesis of psoriasis, as well as Crohn's disease, are identified. New approaches for immune-mediated diseases are discussed.

Keywords: Psoriasis, immune-mediated diseases, gene expression, proteomic analysis, gene interactions, bioinformatic analysis.

INTRODUCTION:

Psoriasis (*Psoriasis vulgaris*, OMMIM 177900) is a chronic inflammatory, recurring, immune-mediated skin disease that involves several other organs and systems. Psoriasis is a complicated genetically based pathology, which involves several groups of genes [1].

The most common clinical manifestations of psoriasis are the appearance of dry, red patches of skin covered with silvery scales. The affected skin is characterized by an increase of the skin cell number and consequent inflammation due to the abnormal keratinocyte differentiation and infiltration of antigen-presenting cells, activation of T-helper cells, and release of proinflammatory cytokines [1, 2]. The appearance of "unaffected" or "uninvolved" skin is normal. However, the gene expression profiling shows that major changes happen in both damaged and undamaged skin of a psoriatic patient, as compared with the skin of a healthy individual [3].

It is supposed that the multi-gene nature of this disease is associated with the presence of several locuses, related to the susceptibility to the disease, known as PSORS1-PSORS9 (Psoriasis Susceptibility) and located on at least 9 chromosomes. Within these chromosomal regions several genes are mapped candidates to be involved in this pathological process [4-6]. In addition, the development of psoriasis may be influenced by several other genomic locuses [7-9]. According to the data appeared in 2008, candidate genes involved in the development of psoriasis may be located on 10 different locuses (PSORS1-PSORS10) [10]. A strong genetic basis of

this disease has been confirmed by family and twin studies that show a high percentage of inheritance (up to 80%), as well as higher concordance (about 70%) in monozygotic twins than in dizygotic twins (up to 30%) [8]. However, just like any other multifactorial diseases, psoriasis is influenced not only by genetic factors, but also by the environment. Many environmental factors may play a role in the development of this disease in susceptible individuals: a mechanical damage skin that may lead to the Kebner's effect (development of psoriatic lesions on the site of epidermal trauma), surgical intervention, UV-radiation, high body mass index, excessive alcohol consumption and smoking, stress factors, including physiological, psycho-emotional and cold stressors[6].

It is well known that bacterial, fungal and virus infections can influence the development of psoriasis [11, 12]. Drugs, such as β -blockers [13], angiotensin-converting enzyme inhibitors [14], antimalarial drugs [15], and lithium [16] can also trigger the development of the disease. Studies show that psycho-emotional stresses can provoke psoriasis development in 23% of the patients, medication - in 16%, physical trauma - in 43%, and infections (over all) - in 14% [17]. The other 4%, perhaps, include other triggering agents.

Therefore, the main reason behind the appearance of psoriasis is unknown; however, it is obvious that this disease is a result of the combined influence of several genetic and environmental factors, as well as the patient's lifestyle. Gene profiling studies show that psoriasis is an immune-mediated inflammatory disease, where epidermal cell structure disbal-

Table 1. Patient's information

	Gender	Age	Psoriasis type	PASI	PsA	Inheritance
1	M	22	P. vulgaris	4.0	No	Denies
2	M	36	P. vulgaris	7.2	No	Father has psoriasis
3	M	23	P. vulgaris	1.8	No	Father and brother have psoriasis
4	M	51	P. vulgaris	9.4	No	Mother has psoriasis
5	M	38	P. vulgaris	2.1	No	Denies
6	M	57	P. vulgaris	4.2	No	Father has psoriasis
7	M	51	P. vulgaris	2.1	No	Denies
8	M	40	P. vulgaris	2.8	No	Father has psoriasis
9	M	34	P. vulgaris	3.2	No	Mother has psoriasis
10	M	49	P. vulgaris	11.4	No	Denies

ance, as well as abnormal cell growth and differentiation, is the result of molecular stress signals that initiate an improper immune response [3].

A study of the molecular mechanisms of the development of psoriasis was performed on the clinical, bioinformatical, molecular-genetic, and proteomic levels. The ethnic backgrounds of psoriasis patients were taken into consideration, and comparison with the development of other diseases (specifically Crohn's disease and atherosclerosis) was performed.

MATERIALS AND METHODS

Skin samples collection from psoriatic patients was done under local anesthetic with the use of dermatological biopsy punch (4mm). Patients received no systematic PUVA/UV therapy for a month before the skin biopsy (Table 1). Biopsy samples from unaffected skin regions were taken at a distance of 3 cm from the damaged skin region [3, 18, 19]. The study was approved by the local ethic committee at the Genetics Institute of the Russian Academy of Sciences (RAS) and was conducted in agreement with the principles of the Declaration of Helsinki.

To grade the severity of psoriasis, a Psoriasis Area and Severity Index (PASI) were used. To calculate overall PASI, local PASI for different body parts were first obtained, using the following formula: PASI = fraction x area x (redness x peeling x thickness). Over all, PASI is equal to the sum of local PASI and can differ in range from 0 to 72. Maximum PASI in this study group was 11.4; minimal, 1.8, meaning that all the patients had an acute form of the disease. The presence or absence of psoriatic arthritis accompanying psoriasis is show in column PsA (Table 1).

Collection of atherosclerosis' autopsies was done from postmortem material. To achieve that, the abdominal aorta with signs of atherosclerosis was taken and samples from the tunica intima layer of atheromatous plaques were collected. Autopsies from the healthy part of the aorta were collected in the same manner. Samples were collected from patients who were treated for different diseases (Table 2).

RNA extraction from biopsies was done by the Qiagen Kit protocol. RNA was cleaned from DNA contamination using DNase Qiagen®.

Reverse transcription was done using Promega's reverse transcriptase M-MLV protocol.

Real-time PCR was done using the fluorescent-labelled oligonucleotide probes. Reaction was done using the chemical supplies manufactured by the company Eurogene. All of the primers and probes were synthesized by the DNA-Syntez company. GAPDH was taken as a control for the target genes expression. Results analysis was done using PCR reactions with the following conditions: reaction efficiency more than 95%, correlation coefficient no less then 0.99, and slope -3.4 ± 0.2 . Method $2^{-\Delta\Delta Ct}$ was used to analyze the results of polymerase chain reaction, according to [20].

To study the **proteomic profiles**, samples of damaged and visually undamaged skin were used. Proteins were extracted and studied using two-dimensional electrophoresis. A silver dye was used to visualize the gel lines. The images obtained were analyzed using the Melanie II program (GeneBio, Switzerland). Protein identification was done using the methods

Table 2. Pathological diagnosis of patients with atherosclerosis

Patient	Age	Gender	Diagnosis
1	80	F	Encephalopathy. Brain atrophy, internal hydrocephalis. Arthrosclerosis of brain arteries. Type II Diabetes. Sclerosis, pancreatic lipomatosis.
2	70	F	Ischemic infraction of right-frontal lobe. Arthrosclerosis of the brain vessels with stenosis up to 50%. Type II Diabetes. Hypertension. Arthrosclerosis of heart blood vessels with stenosis up to 75%.
3	77	M	Infraction of the brain in the area of the middle brain artery. Type II Diabetes. Hypertension III degree. Ischemic hear disease: atherosclerotic cardiosclerosis.
4	67	F	Chronic stomach ulcers. Chronic intestinal ulcers. Diffuse cardiosclerosis. Postmyocardial infraction cardiosclerosis. Arthrosclerosis of the coronary blood vessels with stenosis of 70%. Type II Diabetes. Arthrosclerosis of brain vessels with stenosis of 30%.
5	75	F	Breast cancer. Phlebothrombosis in the right crus. Pulmonary thromboembolism of lung artery. Encephalopathy. Arthrosclerosis of brain vessels with stenosis of 25%. Chronic obstructive bronchitis, diffuse pneumosclerosis, obstructive pulmonary emphysema. Gastritis.
6	80	M	Large postmyocardial infraction cardiosclerosis. Arthrosclerosis of the coronary blood vessels with stenosis of 75%. Secondary hypertension. Bilateral nephrosclerosis. Chronic renal insufficiency.
7	64	M	Chronic alcoholic intoxication. Bilateral bronchial pneumonia. Ischemic heart disease: arthrosclerosis of the coronary blood vessels with stenosis of 40%.
8	87	M	Large postmyocardial infraction cardiosclerosis. Arthrosclerosis of the coronary blood vessels with stenosis of 50%. Hypertension of the small blood-circulation circle. Arthrosclerosis of the brain vessels with stenosis of 30-50%.

of MALDI-TOF mass-spectrometry and nano LC-MS/MS mass-spectrometry at the Proteomic center of the Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences (RAMS).

The GEO DataSets (<http://www.ncbi.nlm.nih.gov/geo/>) database was used to for **bioinformatics analysis**. The GEO DataSets database contains a number of electronic tables where the gene expression microarray results are collected. The MetaCore® program produced by GeneGo Inc (USA) was used for the analysis. The MetaCore® program established that the lowest p-value correlates to the higher possibility of gene involvement into the process. The original threshold for the p-value was established as equal to 0.05.

RESULTS AND DISCUSSION:

THE ROLE OF ETHNIC BACKGROUND IN THE DEVELOPMENT OF PSORIASIS

During the last several decades, extensive knowledge has been collected showing that in addition to individual differences in the metabolism of medical compounds and system reactions to those compounds, as related to their therapeutic and side effects, there are also individual differences in people's susceptibility to diseases. The investigation of the factors that influence an individual's susceptibility to a specific disease and a patient's response to treatment was started in the 1970s under the leadership of L.A. Piruzyan [21, 22]. The necessity of tracking the kinetics of metabolic changes during the process of chemical compound interaction with biological objects was stressed [23].

One's ethnic background also plays a very important role in the development of complicated diseases, such as psoriasis. On average, 2-3% of the world's population suffers from psoriasis. However, these data vary considerably depending on the country, geographic region, and ethnic background of the patient [24]. There is a huge geographic variation in the occurrence of psoriasis (Figure 1) [25]. The huge difference in the disease's incidence between Asia and Europe is proof of the significance of the ethnic component in the development of psoriasis. It is obvious that a regional-ethnic differentiation in the occurrence of psoriasis exists [7, 8, 26]. However, data on the occurrence of psoriasis collected in different populations is somewhat contradictory, which could mean that different analytical methods were used in the col-

lection of the data. In order to make data collection easier and to standardize the analytical methods, we developed an Individual Information Card (IIC), which includes more than 50 grading profiles [27, 28]. The analysis of the genetic structures of samples from Dagestan by the IIC method, as well as family pedigrees, showed high ethnic purity, which is a result of very few marriages across ethnic groups. The consequence of this is a high level of endogamy, which leads to a rise in the degree of homozygosity in the population. Besides, the study of the genealogical trees includes a large number of families with multiple children, which provides an easier way of tracking families' structures in the space of 3-4 generations. For example, the average number of children in the families of Avarecs, Dagestanis, Lezgins, Lakcevs, Azerbaijanis, and Nogaicis is 3; the average number of children in the families of Kalmyks, Tabasarans, and Agul's is 4; but in Russian families the average number of children is only 1. It should also be noted that in such families several generations live fairly close to each other and, therefore, are influenced by similar environmental factors.

Our study showed a significant difference in the presence of complications and duration of the disease depending on the type of illness (Type I – early, Type II – late) in the different ethnic groups of Dagestan. The data obtained in this study show the important role of the ethnic component in the inheritance of such a complicated pathology as psoriasis. The data also show that it is possible to use IIC and a model population from Dagestan to study the type of psoriasis' inheritance with reference to the ethnic component [29].

A psoriasis sensitivity gene study done on samples obtained from Russian, Tatarstan, Bashkortostan, and Hakasiyan ethnic groups in the republics of Bashkortostan and Hakasiya showed that the polymorphic gene locuses HLA-C and HCR are the basic sensitivity markers for psoriasis in the analyzed regions, regardless of the form of the disease and the ethnic group [30].

In order to study the incidence of psoriasis and for gene expression profiling, it is necessary to align all the data according to all the parameters of the study. Several differences in gene expression using Affimetrix-chips in Japanese patients with psoriasis and the results obtained with a similar method in four other study groups show variability in gene expression in individual patients based on age, gender, and environmental factors [3, 31, 32, 33, 34].

ANALYSIS OF BIOLOGICAL MICROARRAYS

Today, the use of methods that allow to systematize and compress a vast pool of genetic information is widespread. This allows to explain different gene interactions. One such method is MetaCore®, a program built by GeneGo Inc. (USA). As a result of gene expression microarray analysis of at least 12,000 genes using MetaCore®, we concluded that 7,563 genes changed expression at least 1.5 times [35]. Figure 2 contains the list of processes changed in psoriasis-. The main processes that change under the influence of psoriasis are the immune response, cell cycle, inflammatory response, proliferation, and others. The process of disease development involves a number of immune system cells, such as T-lymphocytes (Th1, Th17), antigen-presenting cells (APC) – dendrite cells, Langerhans cells, macrophages., and natural killer (NK) cells

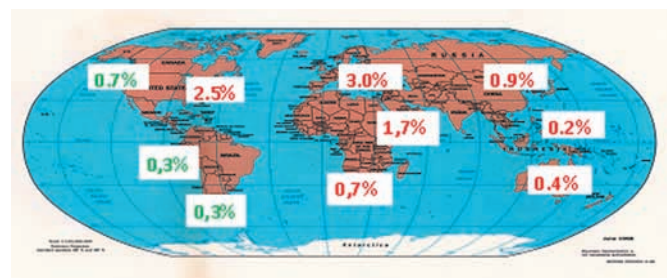


Fig. 1. Frequency distribution of psoriasis case rates in different regions of the world. The percentage of case rates in Aborigines is marked in green

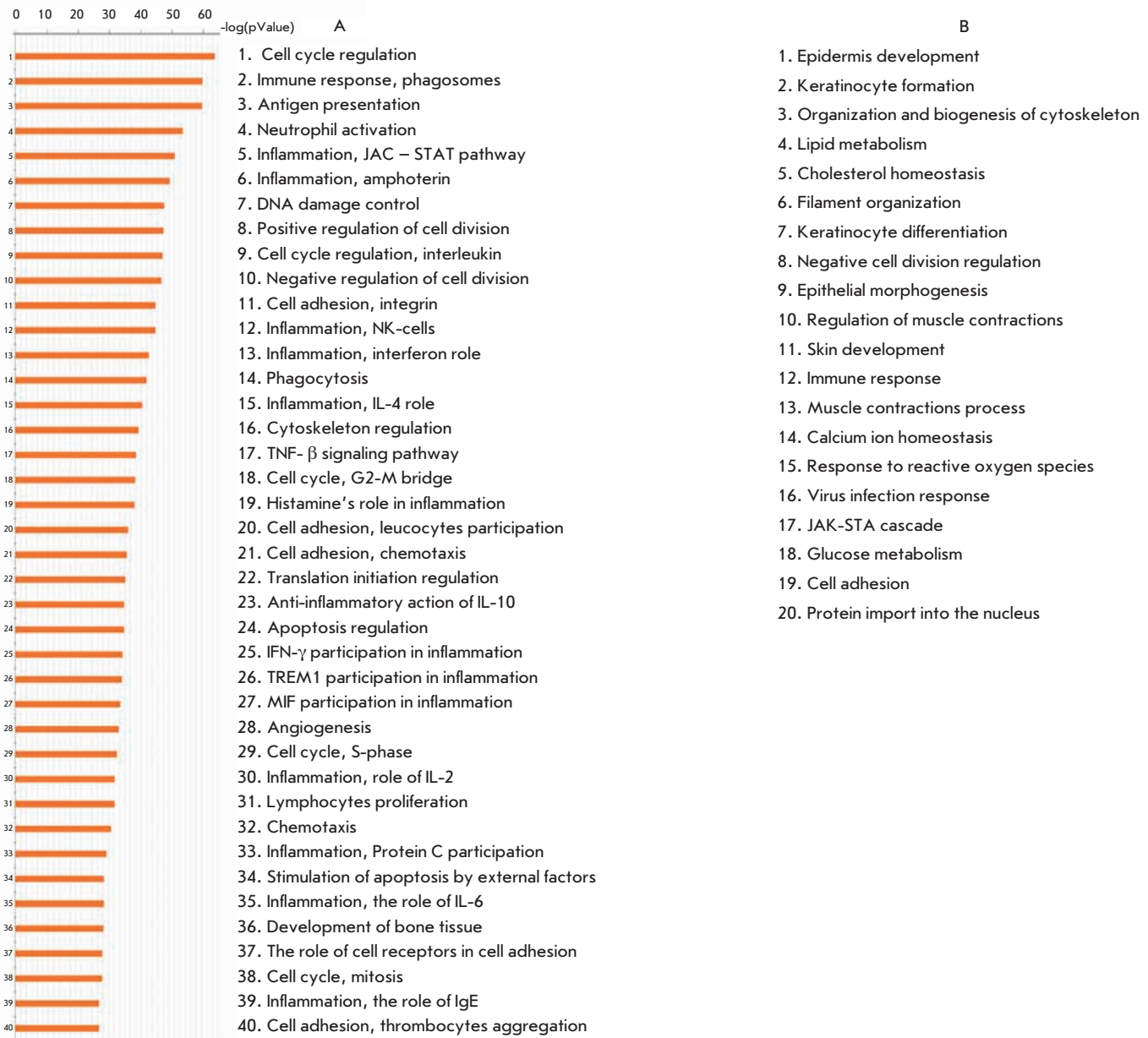


Fig. 2. Forty most altered processes (from the most altered to the less altered) in psoriatic skin compared to the skin of healthy volunteers (A) and twenty most altered processes (from the most altered to the less altered) in psoriatic lesions compared to the unaffected skin (B)

[36]. Cell-to-cell signaling is done with the help of signaling molecules – cytokines, as well as by the interactions of various cell receptors with their ligands. For example, receptors TLR (Toll-like receptors) belong to the family of receptors that promote an immune response when a viral or microbial presence is detected. These receptors are expressed not only on the surface of immune system cells (monocytes, macrophages, dendrite cells, and granulocytes), but also on the surface of the epithelial cells in the respiratory tract and the skin – important regions of host - pathogen interaction [37]. The TLR family includes 11 receptors, and its ligands include li-

popolysaccharides, single-stranded or double-stranded RNA, lipoproteins, flagellin, and unmethylated DNA [38]. Stimulation of these receptors results in stimulation of an entire complex of signaling cascades, including NF-kappaB/Rel, IRF and MAPK – ERK, JNK, and p38. For example, MAPK induces the expression of AP-1 family transcription factors, such as Fos and Jun, and activates transcription activity in the protein complex AP-1, starting several processes leading to inflammation [39]. As a result, the expression of anti-inflammatory cytokines (IL-6, IFN- γ, IL-12 and TNF- α) in both induced and co-stimulating molecules, such as CD80 and

CD86, activates the immune response [40, 41]. Considering that viruses and bacteria are the main triggers of psoriasis, these receptors play an important role in the initiation of the disease [38, 42].

In the entire family of chemokines, a high level of expression was noticed in CCR4, CCR5, CCR6, CCR10, and CXCR3. Chemokines belong to the group of anti-inflammatory cytokines and provoke an influx of Th1 cells into the inflamed regions [43]. Expression of receptors CCR4 and CCR10 usually occurs in CD4+ and CD8+ memory T-cells with CLA + phenotype (cutaneous lymphocyte antigen), which decides on the lymphocyte migration into the lymphatic nodes [44]. A ligand for the receptor CCR4 is CCL17, which is expressed on the endothelial surface of dermal blood vessels, along with ligand CCL27, and provokes the migration of leucocytes into the skin [45]. An expression level of ligand CCL27, which is a ligand for receptor CCR10, occurs in keratinocytes under the influence of anti-inflammatory cytokines IL-1 and TNF- α . In this way, interaction of CCR10-CCL27 involves T-cells in the inflammatory process in the skin [46]. CXCL16 in atherosclerosis can be used as an example of the regulation mechanism of chemokine expression. Expression of this chemokine is induced via IL-18 through the following signaling pathway: MyD88 \rightarrow IrAK1-IrAK4-TRAF6 (tumor necrosis factor receptor – associated factor 6) \rightarrow c-Src \rightarrow PI3K \rightarrow AKT \rightarrow JNK \rightarrow AP-1 [47].

Data shows that the transcription factor AP-1 acts as a common link of the regulatory pathways of TLR-receptors and chemokines.

ANALYSIS OF GENE-TO-GENE INTERACTIONS IN PSORIASIS

Signal transmission from EGF (Epidermal Growth Factor) receptor inside the cell, shown in figures 3a and 3b, can be used as an example of gene-to-gene interaction in psoriasis. Both figures show that genes coding transcription factors, such as c-Fos, c-Myc, c-Jun and ELK1, are activated via signal transmission from EGFR through GRB2 and Shc to SOS, and then through the following pathway: h-Ras \rightarrow c-RAF \rightarrow MEK1/2 \rightarrow ERK1/2 \rightarrow transcription factors. Regardless of the fact that the expression levels is changed only for some genes coding corresponding mediator signals (GRB2, Shc, SOS, h-Ras, c-RAF, MEK1/2), like, for example c-RAF (Figure 3a), expression of transcription factor c-Fos, c-Myc, c-Jun, ELK1, and STAT3 turned out to be a lot higher in cases of damaged skin than in the undamaged skin of psoriatic patients, and higher than in the skin of healthy volunteers [35].

Figure 4 shows the degree of change in the EGF gene signaling pathway expression in the skin of psoriatic patients. We compared the threshold change (Fold Change) of gene expression in psoriasis lesions as compared to the normal skin (Figure 4a) and phenotypically healthy skin of some patients. Fold Change (FC) of expression was established as 2. Figures 4a and 4b show some protein-to-protein interactions that are the part of signaling pathways induced by the Epidermal Growth Factor (EPF) and its ligands. Green arrows show consecutive activation of protein interaction either through bounding or phosphorylation. It is noticeable that the expression of some EGFR-ligands is higher in the psoriasis-damaged skin (HB-EGF and amphiregulin – expression increased 5.77 and 4.96 times, accordingly). Figure 4a shows that each gene

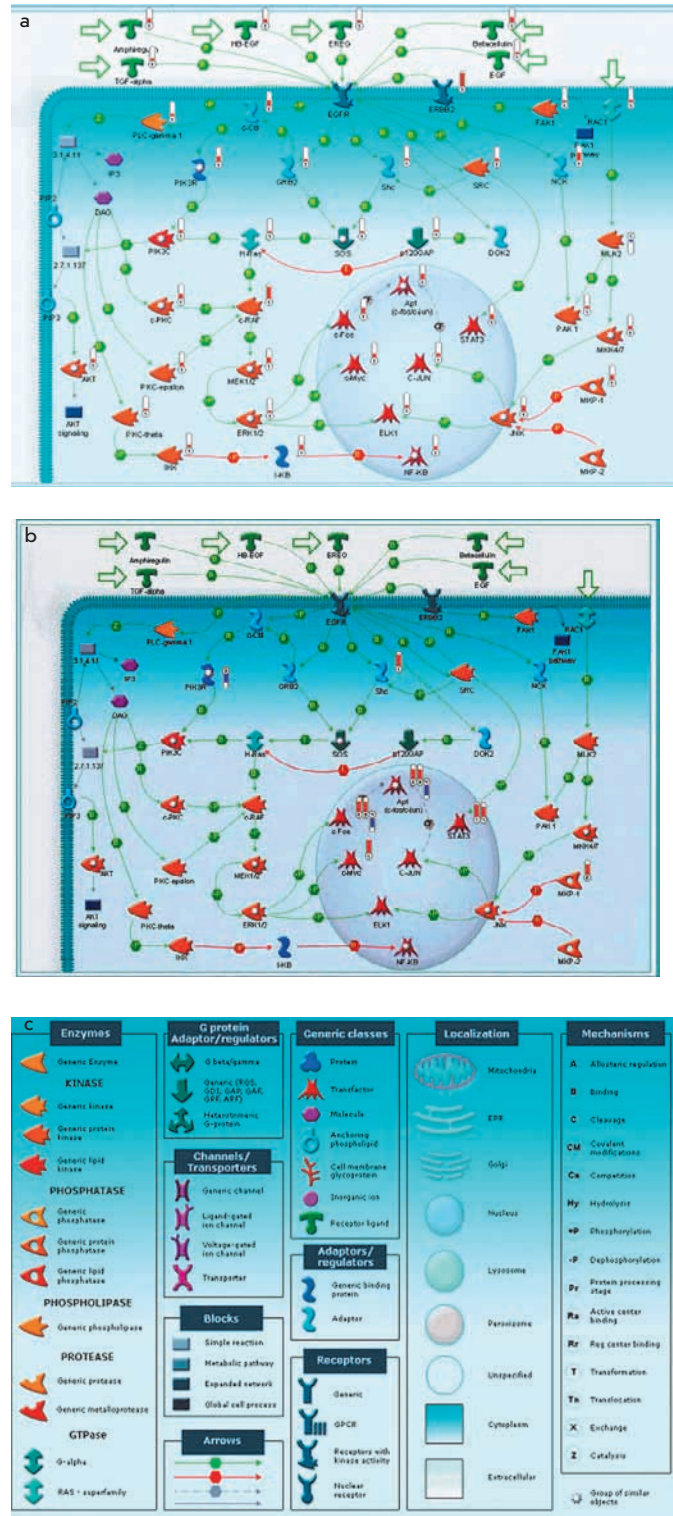


Fig. 3. Signal transduction pathway from EGF to AP-1 and alteration of expression of some genes responsible for this process during psoriasis in skin (A – skin of psoriatic patients compared to the skin of healthy volunteers, B – in affected skin compared to the unaffected skin of psoriatic patients, C – the notation conventions). There are also designations used on the maps that characterize the alteration of the level of expression of genes: - downregulation ; upregulation

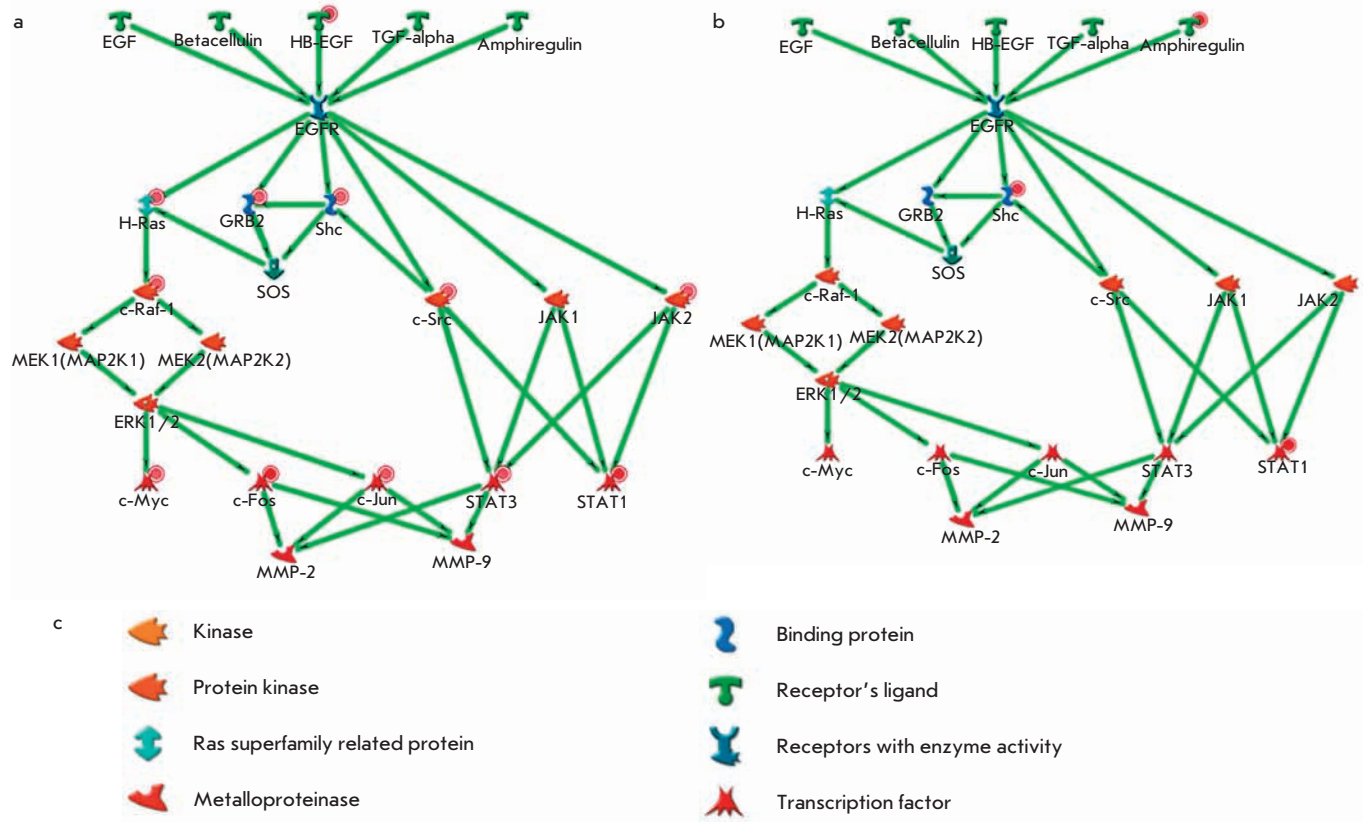


Fig. 4. Level of expression alteration of EGF signal transduction pathway genes in the skin of psoriatic patients: A- as compared to the skin of healthy individuals, B- in affected skin compared to the phenotypically normal skin of the same patients, B - the notation conventions. The hue of the color depends on the absolute value of the expression change (the bigger the change, the more saturated the color)

coding the proteins of the signaling pathway is upregulated; for example, c-Raf kinase with FC = 4.46 and c-Src with FC = 3.96. Moreover, transcription factors themselves are upregulated as well. Therefore, the expression of c-Myc is increased 6 and 15 times; c-Jun, 3.35 times; STAT1, 17 times; and c-Jun, 34 times. The general picture of gene expression changes is shown in Figure 4b. STAT3 is upregulated 8.6 times. This leads to the conclusion that the cells involved in this pathological process are characterized by a significant change in the levels of gene expression, involving a significant number of various molecular pathways and their associated transcription factors in this complicated process. A close look needs to be taken at the overexpression of those transcription factors, since under normal conditions these transcription factors do not show such sudden changes in the expression levels. The uninvolved – meaning visually healthy – skin of psoriatic patients is somewhere in between the skin of the healthy individual and the inflamed skin of the patient, since the difference between samples collected from the visually healthy skin of a psoriatic patient and damaged skin is significant, but not as significant as the difference between the damaged skin of a patient and the skin of a healthy individual.

After the analysis of gene interactions, we concluded that in all of the processes studied the main transcription factors

that changed their expression in psoriasis, only components of transcription factor AP-1 and transcription factor NF- κ B are present. Transcription factor NF- κ B is activated during the immune response. In this particular study, we analyzed the role of components of transcription factor AP-1 [48].

TRANSCRIPTION ANALYSIS

Our group had a significant interest in the comparison of the level expression of genes in the damaged skin regions of psoriatic patients as compared to the expression of the same genes in the visually unaffected skin regions located no farther than 3cm from the affected regions. This kind of comparison was done in several other laboratories around the world [3, 18, 19], and it allows for maximum purity of the experiment. A threshold of change of the expression levels was established for each of the genes equal to 2. Based on a search of the literature and several databases, we established a number of genes that may be important for experimental study. These genes include genes that code transcription factor AP-1 (*C-JUN*, *JUNB*, *JUND*, *C-FOS*, *FOSB*, *FRA-1*, *FRA-2*, and others).

Using a real-time polymerase chain reaction, we were able to analyze the levels of gene expression for 12 genes in the damaged skin and compare it with the visually unaffected

skin of the same psoriatic patient (data not shown). Results of this experiment showed that practically all patients showed a more than doubled expression of gene *FRA-1* in the damaged skin compared to the visually unaffected skin regions (Figure 5). All other genes differed in the direction of change of their expression.

For comparison, we were able to analyze the expression levels of gene *FRA-1* in the samples obtained from the atherosclerotic patients.

Results obtained with real-time PCR showed an increase in the expression of gene *FRA-1* in all patients in the atherosclerosis-affected region of the blood vessels and those suffering from skin psoriasis (Figure 6). At the same time, pathological activation of *FRA-1* expression, as a component of AP-1, can lead to an increase in *IL-2* expression and the consequential stimulation of auto-reactive cytotoxic T-lymphocytes, followed by an increase in the production of anti-inflammatory cytokines and autoantibodies, resulting in the development of a local inflammatory process, which is common in the appearance of psoriasis and atherosclerotic lesions [49, 50]. Increased expression of *FRA-1* can also lead to an increase in the expression of *IL-18*, which induces transcription and expression of MMP9 (matrix metalloproteinase 9) and stimulates production of the active form MMP9. This stimulation leads to an increase in smooth muscles cell migration, which is an important part of the appearance of atherosclerotic plaques [51]. Therefore, we can conclude that it is possible that *FRA-1* plays a crucial role in the pathogenesis of psoriasis and atherosclerosis.

Therefore, with the help of the bioinformatic analysis, the psoriatic gene networks were examined to detect the presence of relatively closed processes, and the main genes were identified: the regulators of the transcription complex AP-1. Transcription profiling of selected genes (*C-JUN*, *JUNB*, *JUND*, *C-FOS*, *FRA-1* and *FRA-2*) based on RNA of 6 chosen genes showed a multidirectional regulation of these genes in psoriasis (data not shown). Gene *FRA-1* stands out from the

other genes analyzed that code the proteins of AP-1 complex (shown on Figure 5). Moreover, change in the expression of this gene before and after treatment showed a clear tendency toward a lowering of the levels of mRNA of this gene, which correlated with the positive dynamic in the patients' illness history (data not shown). Therefore, the transcription activity of *FRA-1* gene complex AP-1 may be a sort of efficiency indicator of treatment at the molecular level.

PROTEOMIC STUDY

The final goal of our study was to analyze the changes occurring in the metabolic pathways which lead to the pathogenesis, since on that level we can identify new targets for medical treatment and new approaches to pharmacological therapy can be found.

With the help of the proteomic analysis of psoriatic skin samples, we were able to establish 10 protein markers present only in the psoriasis-affected skin samples but absent in the unaffected skin samples (Table 3) [52]. A few protein families were of the biggest interest.

The S100 family includes at least 13 proteins, and the genes that code them are localized in cluster form on the 1st chromosome (1q21), which correlates to the psoriasis susceptibility locuse PSORS4. It is known that this protein is highly expressed in the psoriasis-affected skin; however, today this gene is no longer considered a candidate for the development of psoriatic processes. At this point, the exact role of this protein remains unknown.

Proteins SCCA1 and SCCA2 belong to the family of serine proteinase inhibitors. Some studies [53] have shown that protein SCCA1 is expressed in the normal skin, while in the psoriasis affected skin its expression is twice higher.

Protein SCCA2 is not expressed in the control samples, and in the psoriasis-affected skin samples its expression levels are similar to that of SCCA1.

These proteins can be considered as potential targets for the pharmacological compounds in psoriasis treatment [54].

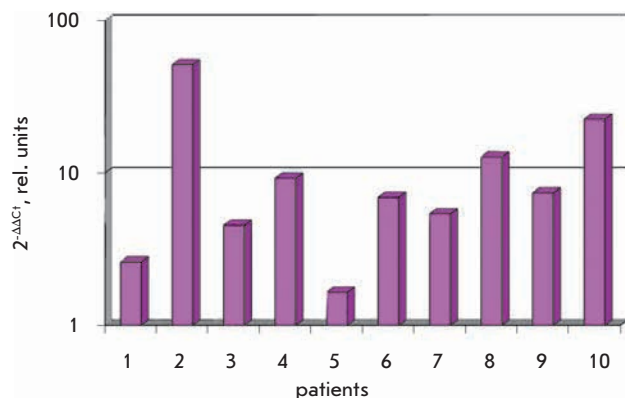


Fig. 5. Level of expression alteration of *FRA-1* in affected psoriatic skin relative to the visually unaffected skin of the same patients (1- absence of expression alteration; <1 – decrease in expression in affected sample; >1 – increase in expression in affected sample)

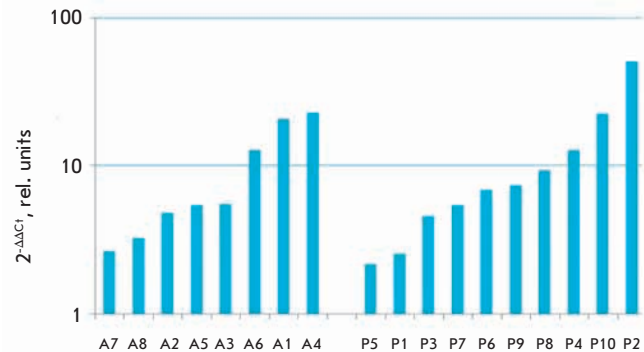


Fig. 6. Level of expression alteration of *FRA-1* gene in atherosclerosis affected vessels compared to visually unaffected (A1-A8) and psoriasis affected skin compared to unaffected (P1-P10). 1- absence of expression alteration; <1 – decrease of expression in affected sample; >1 – increase of expression in affected sample)

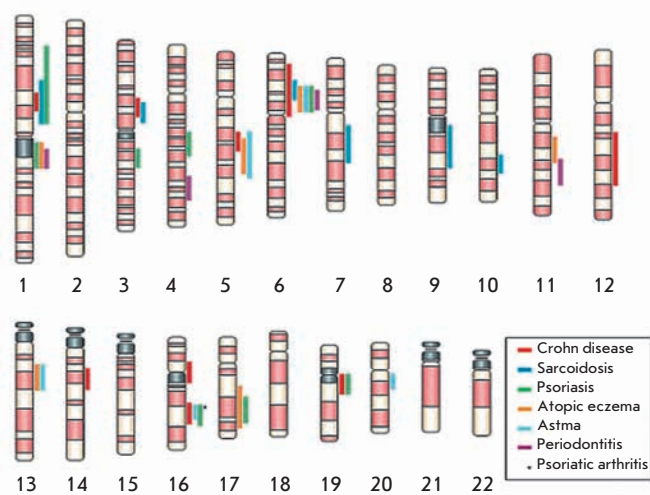


Fig.7. Clustering of linkage regions and disease genes in human inflammatory barrier disease [55]

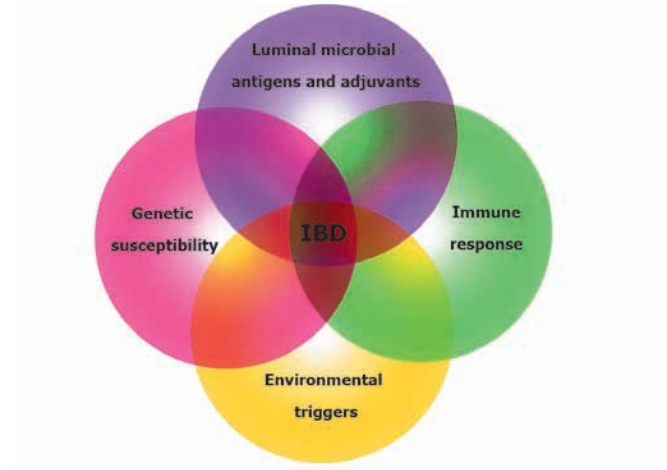


Fig.8. The interaction of different factors (genetic susceptibility, environmental triggers, luminal microbial antigens and adjuvants, immune response) that cause chronic-inflammatory intestine process[56]

COMPARATIVE ANALYSIS OF PSORIASIS AND CROHN'S DISEASE PATHOGENESIS

A multidimensional study of the molecular mechanisms of psoriatic pathogenesis can be considered as a model for the study of the pathogenesis of other immune-mediated inflammatory diseases (IMID – Immune-mediated inflammatory disorders), which are characterized by acute and chronic inflammatory conditions. The most common IMID-phenotypes are Crohn's disease, ankylosing spondylitis, rheumatoid arthritis, psoriasis, uveitis, and psoriatic arthritis. The central role in all of these diseases is played by the cytotoxin TNF- α (tumor necrosis factor - alpha). Figure 7 shows the localization of gene groups responsible for the development of these diseases [55]. Crohn's disease and ulcerative colitis are chronic inflammatory diseases that are generally called IBD (inflammatory bowel disease). Both twin

studies and the generally higher incidence of these diseases in close relatives point to a genetic basis for these diseases. The increase in the number of cases in Europe and North America since the second half of the 20th century is an indication of the strong influence of environmental factors, with relatively low levels of concordance in the monozygotic twins (~50% for the Crohn's disease and 10% for the ulcerative colitis). Even though animal studies do provide clues about the etiology of Crohn's disease, its nature remains a mystery [55, 56].

The most widespread hypothesis about IBD pathogenesis is the presence of a heavy T-cell immune response to certain environmental factors and some pathogenic enterobacteria in the genetically susceptible individuals, which increases the beginning or reactivation of the disease. Figure 8 shows the activation of four independent components, the interaction of which at different points of the disease's

Table 3. Proteins identified as different at involved and uninvolved psoriatic skin [52]

Blot, number	Protein name	%Vol	
		Affected skin	Unaffected skin
1	Keratin 17	1.97 ± 0.92	0.18 ± 0.06
	Keratin 14		
	Keratin 16		
2	SCCA2/SCCA1	0.28 ± 0.09	0.066 ± 0.02
3	Squamous cell carcinoma antigen; SCC antigen	0.35 ± 0.18	0.03 ± 0.01
4	Enolase 1	0.87 ± 0.22	0.40 ± 0.14
5	Superoxide dismutase [Mn]	0.22 ± 0.03	0.11 ± 0.02
6	Galectin 7; Gal - 7	1.14 ± 0.41	0.19 ± 0.01
7	Protein S100 - A9	0.54 ± 0.03	0
8	Protein S 100 - A9	0.15 ± 0.07	0
9	Protein S100-A7 (Psoriasin)	0.51 ± 0.16	0
10	Protein S100-A7 (Psoriasin)	1.36 ± 0.37	0.02 ± 0.03

Table 4. The list of genes common for pathogenesis of psoriasis and Crohn's disease.

Gene	Identification Number	Location	Gene	Identification Number	Location
GNA15	2769	19p13.3	CBX3	11335	7p15.2
GPM6B	2824	Xp22.2	UGT1A6	54578	2q37
IFI44	10561	1p31.1	DEGS1	8560	1q42.11
OAS2	4939	12q24.2	PSME2	5721	14q11.2
FOXC1	2296	6p25	TRAK2	66008	2q33
ZNF207	7756	17q11.2	DNAJC7	7266	17q11.2
IFI35	3430	17q21	CSNK1D	1453	17q25
STAT3	6774	17q21.31	TRIM22	10346	11p15
TXNDC1	81542	14q22.1	IRF9	10379	14q11.2
MRPL9	65005	1q21	UBE2L6	9246	11q12
CASP4	837	11q22.2-q22.3	ETS2	2114	21q22.2
MECP2	4204	Xq28	QPCT	25797	2p22.2
LONRF1	91694	8p23.1	SFPQ	6421	1p34.3
CG018	90634	13q12-q13	UGT1A4	54657	2q37
VKORC1	79001	16p11.2	H2AFY	9555	5q31.3-q32
MIB1	57534	18q11.2	HMG1	3150	21q22.2
RFK	55312	9q21.13	CTSC	1075	11q14.1-q14.3
SOSTDC1	25928	7p21.1	SERPINB5	5268	18q21.3
KIAA1033	23325	12q24.11	IER2	9592	19p13.13
SYNCRIP	10492	6q14-q15	S100A8	6279	1q21
RARG	5916	12q13	ARMET	7873	3p21.1
DDOST	1650	1p36.1	FGFR2	2263	10q26
CDC42EP1	11135	22q13.1	RBPMS	11030	8p12-p11
S100A9	6280	1q21	JUNB	3726	19p13.2
PHGDH	26227	1p12			

development is necessary for its clinical implications. From our point of view, comparable studies of the genetic factors involved in the pathogenic pathways of psoriasis and Crohn's disease can play a significant role in the deciphering of the molecular mechanisms of the pathogenesis of these diseases. A comparative bioinformatics analysis of microarray database GEO Data Sets allowed to prepare a general list of the genes that undergo changes in the cell processes of both diseases.

IDENTIFICATION OF GENES THAT CHANGE THEIR EXPRESSION IN BOTH PATHOLOGIES – PSORIASIS AND CROHN'S DISEASE

Our research was focused on the identification of differentially expressed genes in each data set and comparison of these lists of genes at the systemic level. The approach we took in working on this problem was dictated by the character of the expression data (high noise level, large amount of data to be analyzed) and the attributes of each individual data set taken for the analysis (they were done on different types of microarrays that have different numbers of probes and, therefore, cannot be compared directly) [57].

Originally, **psoriasis data** sets contained information on the expression of 12,626 probes for 8 experiments (4 samples of damaged skin and 4 samples of healthy skin). When probes with badly detectable expression were thrown out, the number of probes was reduced to 5,076. The list of probes with statistically significant differences in expression be-

tween the affected and unaffected skin samples contained 410 probes with a significance level of 0.1.

The data sets for **Crohn's disease** contained information on expression levels for 24,016 probes in 21 experiments (10 intestinal epithelial samples of healthy individuals and 11 epithelial samples of damaged epithelia). The list of probes with statistically significant differences in expression between the damaged and undamaged tissues was 3,850 probes, with a significance level of 0.1. Such a huge difference in the probe lists' sizes was due to the fact that the algorithm to control Type I errors (FDR) depends on the size of the input set: the larger the number of genes in the input set, the larger the number of genes taken with the same p-value distribution that will go through FDR control. In our study, the number of analyzed probes in the Crohn's disease data set is five times that of the psoriasis data.

The resulting data sets of differentially expressed genes were introduced into the MetaCore®. Since microarrays contain not only probes for the genes studied, but also for a large number of EST with unclear roles, in this step the size of the genes lists changed because not all of the probes have a correlating gene in the MetaCore® database, and because some probes correlate to more than one gene. The size of the genes lists changed to 425 genes for psoriasis and 2,033 for Crohn's disease.

These genes lists had 49 common genes, and the level of similarity is rather significant (p-value = 4.94×10^{-2}). Fisher's

test was used to grade this significance; 9,017 genes present in the two study data sets were taken. Identification of these genes was done by comparing the genes lists for both gene microarray sets in MetaCore®.

Forty-nine genes were selected for the further analysis and are shown in table 4. The cell molecular-genetic processes associated with the genes common to psoriasis and Crohn's disease were of a particular interest. Program methods in MetaCore® were used to create Table 5, where the most possible cell processes involving genes from Table 4 are shown. The processes shown in Table 5 can be divided into two categories: those participating in inflammation processes and those participating in the regulation of the cell cycle. In both psoriasis and Crohn's disease, the main pathological focus lies in the focus of inflammation. Cell cycle and apoptosis are disturbed in psoriasis. Keratinocytes in the focus of the inflammation have no time to apoptose due to extra proliferation; thus, they form the skin patches we observe in psoriasis. Similar processes take place in the intestine of Crohn's disease's patients.

REASONABLE APPROACHES TO THE PHARMACOLOGICAL TREATMENT OF PSORIASIS AND POSSIBLE NEW APPROACHES TO THE TREATMENT OF IMMUNE-MEDIATED INFLAMMATORY ILLNESSES

Psoriasis is considered a recurrent incurable illness; its treatment focuses on lengthening the remission periods and reducing the severity of the disease. The drugs used (cyclosporine, system retinoids and fumarates) lead only to a temporary improvement of the patient's condition. Nonetheless, some new approaches are being explored. New therapeutic methods are characterized by a more specific influence on the specific molecular targets that play a key role in the formation of the pathological processes in psoriasis. They include modifiers of the biological response, such as alefacept, efalizumab, etarcept, infliximab, and adalimumab, which specifically target the molecular mediators involved in the immune-pathogenesis of psoriasis (receptors and ligands). For example, the suppression of TNF α activity, a key cytokine in the innate immune response, is achieved through three inhibitors of TNF α (alefacept, etarcept, and infliximab). Etarcept is a two-component protein produced from the ligand-binding fragment receptor TNF attached to the Fc-fragment IgG1. It binds TNF and blocks its interaction with cell-surface receptors, reducing the inflammation process. Infliximab – TNF α human monoclonal antibodies lower the activity of TNF α , which lowers the production of *IL-1* and *IL-6* [58]. Therefore, the new generation of biological approaches to the treatment of psoriasis focus on the specific destruction of targets for the T-cell mediated pathogenic processes.

A number of successful attempts at psoriatic arthritis treatment with bio-modifiers have been documented since 2001 [58, Abstract book of Third EAD International Spring Symposium, Sofia, 2005]. TNF α levels are increased in the intestinal mucosa during Crohn's disease; therefore, TNF α inhibitors are used to treat this disease. This kind of monotherapy by means of bio-modifiers is used for the treatment of psoriatic arthritis and rheumatoid arthritis [59, 60].

Table 5. The common cell processes typical for psoriasis and Crohn's disease

Processes	P-value
Inflammation: interferon signaling pathway	2.19E-03
Signal transmission: WNT signaling pathway	1.20E-02
Regulation of translation initiation	5.66E-02
Morphogenesis of blood vessels	9.76E-02
DNA repair	1.17E-01
Inflammation: amphotericin signaling pathway	1.19E-01
Cell cycle and apoptosis driven proteolysis	1.29E-01
Interleukin regulation of cell cycle in G1-S phase	1.29E-01
Signal transmission: androgen receptor signaling pathways	1.34E-01

However, a number of unusual side effects associated with this treatment have been reported. For example, some researchers have indicated the development of psoriatic lesions on the skin of some Crohn's disease patients [60, 61, 62]. The mechanisms leading to such serious side effects are unclear; therefore, continued research in this area on the molecular-genetic level is necessary.

In conclusion, the main goal of investigations of the pathways leading to the pathogenesis of immune-mediated diseases on the molecular level is to find new pharmacological treatment options. The proteomic studies of skin samples affected by psoriasis and intestinal fragments affected by Crohn's disease performed by our research group are geared toward finding new targets for pharmacological treatment options. It is also necessary to compare the obtained data with the results of the meta-analysis of development pathways for these two immune-mediated inflammatory diseases done by us.

We used psoriasis as an example to show the necessity of an integrated approach to the deciphering of the disease's pathogenesis stages, including clinical and bioinformatical analysis, as well as analysis of metabolic and genetic data, which will assist in the development of simple methods of individual characterization of pathogenesis and finding more effective methods for the treatment of each individual. In this manner, psoriasis is taken as a typical complicated disorder and is considered an adequate model for the study of the pathogenicity mechanisms for other disorders similar to psoriasis in their influence on human life. ●

The work was supported by the Russian Federal Agency for Science and Innovation (federal contract 02.512.11.2231) and the Presidium of Russian Academy of Sciences, as a part of the program "Fundamental Medical Science."

REFERENCES

1. Lowes M.A., Bowcock A.M., Krueger J.G. Pathogenesis and therapy of psoriasis. // *Nature*. 2007. V. 445 P. 866–873.
2. Bowcock A. M. and Krueger J. G. Getting under the skin: the immunogenetics of psoriasis. // *Nat. Rev. Immunol.* 2005. V. 5. P. 699–711.
3. Kulsky J.K., Kenworthy W., Bellgard M. *et al.* Gene expression profiling of Japanese psoriatic skin reveals an increased activity in molecular stress and immune response signals // *J. Mol. Med.* 2005. V. 85. P. 964–975.
4. Barker J. N. Genetic aspects of psoriasis. // *Clin. Exp. Dermatol.* 2001. V. 26. № 4. P. 321–5.
5. Prinz J. C. Psoriasis vulgaris—a sterile antibacterial skin reaction mediated by cross-reactive T cells? An immunological view of the pathophysiology of psoriasis. // *Clin. Exp. Dermatol.* 2001. V. 26. № 4. P. 326–32.
6. Bowcock A. M. and Cookson W. O. The genetics of psoriasis, psoriatic arthritis and atopic dermatitis. // *Hum. Mol. Genet.* 2004. V. 13. № 1. P. R43–55.
7. Asumalahti K., Veal C., Laitinen T. *et al.* Psoriasis Consortium. Coding haplotype analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus. // *Hum. Mol. Genet.* 2002. V. 11. P. 589–597.
8. Sagoo G.S., Cork M.J., Patel R. *et al.* Genome – wide studies of psoriasis susceptibility loci: a review. // *Journal of Dermatological Science.* 2004. V. 35. P. 171–179.
9. Fun X., Yang S., Huang W. *et al.* Fine mapping of the psoriasis susceptibility locus PSORS1 support HLA-C as the susceptibility gene in the Han Chinese population. // *PLoS Genet.* 2008. V. 4. P. 1–10.
10. Duffin K.C., Chandran V., Krueger G.G. *et al.* Genetics of Psoriasis and Psoriatic Arthritis: update and future direction (GRAPPA 2007) // *J. Rheumatology.* 2008. V. 35. № 7. P. 1449–1453.
11. Duvic M., Johnson T.M., Rapini R.P. *et al.* Acquired immunodeficiency syndrome-associated psoriasis and Reiter's syndrome. // *Arch. Dermatol.* 1987. V. 123. P. 1622–1632.
12. Swerlick R.A., Cunningham M.W., Hall N.K. Monoclonal antibodies cross-reactive with group A streptococci and normal and psoriatic human skin. // *J. Invest. Dermatol.* 1986. V. 87. P. 367–371.
13. Gold M.H., Holy A. K., Roenigk H. H. Jr. Beta-blocking drugs and psoriasis. A review of cutaneous side effects and retrospective analysis of their effects on psoriasis. // *J. Am. Acad. Dermatol.* 1988. V. 19. P. 837–841.
14. Wolf R., Tamir A., Brenner S. Psoriasis related to angiotensin-converting enzyme inhibitors. // *Dermatologica.* 1990. V. 181. P. 51–53.
15. Baker H. Psoriasis – clinical features. // *Br. Med.* 1971. V. 3. P. 231–233.
16. Lazarus G.S., Gilgor R.S. Psoriasis, polymorphonuclear leukocytes, and lithium carbonate. An important clue. // *Arch. Dermatol.* 1979. V. 115. P. 1183–1184.
17. Asumalahti K. Molecular genetics of psoriasis. // *Helsinki university biomedical dissertation.* № 27. 2003.
18. Sonkoly E., Wei T., Janson P.C., Saaf A. *et al.* MicroRNAs: novel regulators involved in the pathogenesis of Psoriasis? // *PLoS ONE.* 2007. V. 2. № 7. P. 1–8.
19. Yao Y., Richman L., Morehouse C. *et al.* Type I Interferon: Potential Therapeutic Target for Psoriasis? // *PLoS One.* 2008. V. 3. № 7. e2737. P. 1–14
20. Livak K.J., Schmittgen T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. // *Methods.* 2001. V. 25. № 4. P. 402–408.
21. Podimov V.K., Gladkich S.P., Mishkin V.S., Piruzyan L.A., N-acetyltransferase polymorphism of patients with lupus erythematosus // “The problems of investigation, study and production of new drugs”, Shwiesa, Kaunas. – 1979, P. 52–53.
22. Piruzyan L.A. Metabolic human passport – the basis of new strategy in pharmacology // *Vestnik of RAS.* 2004. V.74. P. 610–618.
23. Piruzyan L.A. On pharmacological metrology // *Bull. USSR Academy of Sci., Biol.* 1990. №2. P. 302–303.
24. Gudjonsson J.E., Elder J.T. Psoriasis: epidemiology. // *Clinics in Dermatology.* 2007. V. 25. P. 535–546.
25. Piruzyan An.L., Abdeev R.M., Sarkhisova M.K. The role of pharmacogenetics and pharmacogenomics in dermatology // *Clinical dermatology and venerology.* 2005. №3. P. 8–13
26. Bowcock A.M. The genetics of psoriasis and autoimmunity. // *Annu. Rev. Genomics Hum. Genet.* 2005. V. 6. P. 93–122.
27. Piruzyan An.L. *Doctoral diss.* 2005. №3. P.352
28. Piruzyan L.A., Goldenkova I.V., Korsunskaya I.M. *et al.* Scientific approach to standardization of psoriatic patient's care (scientific review) // *The problems of standardization in health care.* 2006. №6. P. 7–13.
29. Goldenkova-Pavlova I.V., Piruzyan A.L., Abdeev R.M. *et al.* Population analysis and determination of the ethnic background are necessary in the study of multifactorial diseases (a study using the Dagestan population as a model) // *Genetika.* 2006 V.42, №8. P.1137–42.
30. Akhmetova V.L., Galimova E.S., Yusunbaev B.B., Khusnutdinova E.K. The analysis of the genes responsible for susceptibility to psoriasis in Bashkortostan and Khakasia republics // *Medical genetics.* 2009. №8. P. 29–35.
31. Bowcock A.M., Shanon W., Du F. *et al.* Insight into psoriasis and other inflammatory diseases from large scale gene expression studies. // *Hum. Mol. Genet.* 2001. V. 10. P. 1793–1805.
32. Oestreicher J.L., Walters I.B., Kikichi T. *et al.* Molecular classification of psoriasis disease-associated genes through pharmacogenomic expression profiling. // *Pharmacogenomics J.* 2001. V. 1. P. 272–287.
33. Nomura I., Gao B., Boguniewicz *et al.* Distinct pattern of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis. // *J. Allergy Clin. Immunol.* 2003. V. 112. P. 1195–1202.
34. Zhou X., Krueger J.G., Kao M.-C.J. *et al.* Novel mechanism of T-cell and dendritic cell activation revealed by profiling of psoriasis on the 63, 100-element oligonucleotide array. // *Physiol. Genomics.* 2003. V. 13. P. 69–78.
35. E.S. Piruzian, T.A. Nikolskaya, R.M. Abdeev, S.A. Brouskin, Transcription Factor AP-1 Components As Psoriasis Candidate Genes // *Molecular Biology.* 2007, Vol. 41, No. 6, pp. 1069–1080.
36. Sabat R., Philipp S., Höflich C., *et al.* Immunopathogenesis of psoriasis. // *Exp Dermatol.* 2007. V. 16. № 10. P. 779–98.
37. Medzhitov R., Preston-Hurlburt P., Janeway C.A. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. // *Nature.* 1997. V. 388. P. 394–397.
38. McInturf J.E., Modlin R.L., Kim J. The role of toll-like receptors in the pathogenesis and treatment of dermatological disease. // *J. Invest Dermatol.* 2005. V. 125. № 1. P. 1–8.
39. Hu X., Chen J., Wang Lu *et al.* Crosstalk among Jak-STAT, Toll-like receptor, and ITAM dependent pathways in macrophage activation. // *Journal of Leukocyte Biology.* 2007. V. 82. P. 237–243.
40. Schnare M., Barton G. M., Holt A. C. *et al.* // *Nat Immunol.* 2001. V.2. № 10. P. 947–950.
41. Schjette K.W., Thompson K.M., Nilsen N. *et al.* Cutting edge: link between innate and adaptive immunity: Toll-like receptor 2 internalizes antigen for presentation to CD4+T cells and could be an efficient vaccine target. // *J. Immunol.* 2003. V. 171. № 1. P. 32–36.
42. Perez-Lorenzo R., Zambrano-Zaragoza J. F., Moo-Castillo K. *et al.* IgG class antibodies to heat shock-induced streptococcal antigens in psoriatic patients. // *Int J. Dermatol.* 2003. V. 42. P. 110–115.
43. Krueger J.G. The immunologic basis for the treatment of psoriasis with new biologic agents. // *J. Am. Acad. Dermatol.* 2002. V. 46. № 1. P. 1–23.
44. Reiss V., Proudfoot A.E., Power C.A. *et al.* CC Chemokine Receptor (CCR)4 and the CCR10 Ligand Cutaneous T Cell-attracting Chemokine (CTACK) in Lymphocyte Trafficking to Inflamed Skin. // *J. Exp. Med.* V. 194. № 10. 2001.
45. Lonsdorf A.S., Hwang S.T., Enk A.H. *et al.* Chemokine Receptors in T-Cell-Mediated Diseases of the Skin // *J. Invest. Dermatol.* 2009.
46. Honey B., Alenius H., Muller A. *et al.* CCL27–CCR10 interactions regulate T cell-mediated skin inflammation. // *Nat. Med.* 2002. V. 8. P. 157–65.
47. Chandrasekar B., Mummidi S., Valente A. *et al.* The Pro-atherogenic Cytokine Interleukin-18 Induces CXCL16 Expression in Rat Aortic Smooth Muscle Cells via MyD88, Interleukin-1 Receptor-associated Kinase, Tumor Necrosis Factor Receptor-associated Factor 6, c-Src, Phosphatidylinositol 3-Kinase, Akt, c-Jun N-terminal Kinase, and Activator Protein-1 Signaling. // *The Journal of biological chemistry.* 2005. V. 280. № 28. P. 26263–26277.
48. Turpaev K.T. 2006. Role of transcription factor AP-1 in integration of intracellular signal pathways. // *Mol. Biol.* 40, 945–961.
49. Gonsky R., Deem R.L., Hughes C.C. *et al.* Activation of the CD2 pathway in lamina propria T cells up-regulates functionally active AP-1 binding to the IL-2 promoter, resulting in messenger RNA transcription and IL-2 secretion. // *J. Immunol.* 1998. V. 160. № 10. P. 4914–22.
50. Bantel H., Schmitz M.L., Raible A. *et al.* Critical role of NF-kappaB and stress-activated protein kinases in steroid unresponsiveness. // *FASEB J.* 2002. V. 16. № 13. P. 1832–4.
51. Chandrasekar B., Mummidi S., Mahimainathan L. *et al.* Interleukin-18-induced human coronary artery smooth muscle cell migration is dependent on NF-kappaB- and AP-1-mediated matrix metalloproteinase-9 expression and is inhibited by atorvastatin. // *J. Biol. Chem.* 2006. V. 281. № 22. P. 15099–109.
52. Bruskin S. A. *PhD Thesis*, 2008. P. 120.
53. Takeda A., Higuchi D., Takahashi T. *et al.* Overexpression of serpin squamous cell carcinoma antigens in psoriatic skin. // *J. Invest. Dermatol.* 2002. V. 118, № 1. P. 147–1454.
54. Bruskin S.A., Abdeev R.M., Moshkovskii S.A. *et al.* Psoriasis proteomic research as an approach for identification of potential targets for pharmacotherapy // *Clinical dermatology and venerology.* 2009. №1. P.28–31.
55. Schreiber S., Rosenstiel P., Albrecht M. *et al.* Genetics of Chron disease, an archetypal inflammatory barrier disease. // *Nature.* 2005. V. 6. P. 376–388.
56. Sartor B.R. Mechanisms of Disease: pathogenesis of Crohn's disease and ulcerative colitis. // *Nature Clinical Practice Gastroenterology & Hepatology.* 2006. V. 3. P. 390–407.
57. Piruzian E.S., Ishkin A.A., Nikol'skaya T.A., Abdeev R.M., Bruskin S.A. A comparative analysis of the molecular genetic processes in the pathogenesis and Crohn's disease, 2009, V.46, №1, P. 175–179
58. Chaudhari U., Romano P., Mulcahy L.D. *et al.* Efficacy safety of infliximab monotherapy for plaque-type psoriasis: a randomized trial. // *Lancet.* 2001. V. 357. P. 1842–1847.
59. Baeten D., Kruithof E., Van den Bosch *et al.* Sistematic safety follow up in a cohort of 107 patients with spondyloarthritis treated with infliximab: a new perspective on the role of host defence in the pathogenesis of the disease? // *Ann. Rheum. Dis.* 2003. V. 62. P. 829–834.
60. Sfikakis P.P., Iliopoulos A., Elezoglou A. *et al.* Psoriasis induced by anti-tumor necrosis factor therapy: a paradoxical adverse reaction. // *Arthritis Rheum.* 2005. V. 52. P. 2513–2518.
61. Peramiqel L., Puig L., Dalmau J. *et al.* Onset of flexural psoriasis during infliximab treatment for Crohn's disease. // *Clin. Exp. Dermatol.* 2005. V. 30. № 6. P. 713–714.
62. Umeno J., Matsumoto T., Jo Y. *et al.* Psoriasis during anti-tumor necrosis factor – alpha therapy for Crohn's disease. // *Inflamm. Bowel Dis.* 2007. V. 13. P. 18188–1189.

Rules for authors

GENERAL PROVISIONS

Journal «Acta Naturae» publishes experimental research papers and reviews, and also discussion papers, mini-reviews, short communications, which are related to most topical problems of the fundamental and practical life sciences and biotechnologies.

Journal is published by Park Media Publishing Company in Russian and English languages.

Edition board of the journal "Acta Naturae" would like to ask authors to follow the following rules. Manuscripts, which are not consistent with those rules, will be sent back to authors without consideration.

The maximal size of review together with tables and references should not be more than 60 000 character (it is about 40 pages of A4 format and with 1.5 lines interval, Times New Roman, size 12), and 16 figures.

The size of experimental research paper should not exceed 30 000 characters, (20 pages of A4 format, including tables and references). The number of figures should not exceed 10. Papers of larger size could be accepted only in case of preliminary agreement with the editors.

Short communication should include task design, experimental data and conclusions. The size of short communication should not exceed 12 000 characters (8 pages of A4 format, including tables and references, no more than 12 references). Number of figures should not exceed 4.

Manuscript should be presented as a set of files: main text, summary, list of references, comments to figures, tables must be in Microsoft Word 2003 for Windows format, figures must be in JPG or TIF format, each figure should be represented as a separate file.

Manuscript should be sent to editor board as an electronic version; it is possible to do it on a CD-ROM.

MANUSCRIPT DESIGN

Manuscript should have:

Title of the manuscript. It must not be very long or very short or low-informative. It should represent the main result, the essence and the novelty of the work.

Authors initials and surnames with the references to the organisations in which authors work. Author, which is responsible for the communication with the editorial board and correction, should be indicated by the reference, and his correspondent e-mail should be represented on the bottom of the first page.

Full name of the scientific organisation and agency.

Summary. Structure of summary must be very clear and reflect following: task design; experimental methods, possibility of the practical application, new possible tasks and problems. 20 lines.

Keywords. Keywords should reflect: the subject of research, method, object, specificity of this work.

Abbreviations.

Introduction.

Chapter "Experimental part."

Chapter "Results and discussion."

Chapter "Conclusions." At the end of this chapter names of organisations, that provided financial support, should be indicated (and in brackets number of grants should be indicated).

Chapter "References." References in text should be given in square brackets, for example [1].

RECOMMENDATION ABOUT TEXT FORMATTING:

We recommend using Microsoft Word 2003 for Windows editor.

Font should be Times New Roman, and standart size is 12.

Interval between lines is 1.5

We do not recommend using more that one space between words.

Strictly do not use automatic reference option, automatic hyphenation or automatic block of hyphenation, automatic list, automatic indentation etc.

It is recommended to use Word for making tables (Table – Add Table) or MS Excel. Tables, which are created manually (by using a lot of empty spaces, but not cells), should not be used.

Always, it must be an empty space between surnames and initials: A. A. Ivanov (except from the names of authors in the title of the manuscript, in this situation empty spaces must be between initials as well, i.e. A. A. Ivanov).

Everywhere in the text, except the date of manuscript's arrival, all dates should be in format "date.month.year," according the examples: 02.05.1991, 26.12.1874 and so on.

Do not use full stop after title of the manuscript, authors names, addresses, headings and subheading, titles of tables, comments to figures, dimensionalities (s for second, g for gram, min for minute, h for hours, d for days, ° for degrees).

Full stop must be after: footnotes (and particularly in tables), comments to the table, short annotations, abbreviations (y. for years, t. melt. for the temperature of melting, but not for inferior indexes: T_{melt} for temperature of melting, $T_{ph.tr}$ for temperature of phase transition). Exception: mln for million does not need full stop.

For decimal fractions use full stops, but not commas (0.25, but not 0,25).

Dash "-" must be between two empty spaces, symbols for "minus", "interval" or "chemical bound" do not need to be between two empty spaces.

Use only "x" as a symbol for multiplication. Use symbol "·" only in case if there is a number in the right from his symbol. Symbol "·" is used for the complex compounds in chemical formulas, and also for the non-covalent complexes (DNA-RNA and so on)

Do not use inverted commas for quotes.

All numerical data should be represented as tables.

In formulas, use the letters of Latin and Greek alphabets.

Latin names of genera and species must be in italic font, for higher taxons use regular font. For the names of viruses and bacteriophages in Latin transcription use regular font.

Names of genes (except the yeast genes) should be in italic font, for names of proteins use regular font.

Names of nucleotides (A, T, G, C, U), amino acid residues (Arg, Ile, Val etc.) and phosphates (ATP, AMP etc) should be written in Latin transcription using regular font.

Numeration of nucleotides and amino acid residues should be written without hyphen (T34, Ala89).

Use international system of units SI for measurement units.

For molecular mass, use daltons (da, kda, mda).

For the number of nucleotide pair, use abbreviations (b.p, t.b.p.).

Biochemical terms (and particularly, names of enzymes) should be according the international rules of IUPAC.

Using of abbreviations of terms in the text should be minimized.

Repetition of the same data in text, tables and graphs is not allowed.

RULES FOR ILLUSTRATIONS

Figures for manuscripts should be represented as separate files.

The resolution of illustrations must be 300 dpi or higher for the colourful or halftone pictures, and 600 dpi for the black-white illustrations.

It is not allowed to keep in files the ways of clipping paths that are not in use, and also to keep additional channels (Path and Alpha Channel), and also to use additional layers.

File's formats should be in Adobe Photoshop up to version 7.0. (including this version), TIFF, JPG.

REVIEWING, PREPARATION TO PUBLISHING, AND THE QUEUE FOR PUBLICATION

Papers will be published accordingly to the process of receiving. The queue for the publications will be established accordingly to the date of approval for publication. Members of Editorial Board have a right to recommend the papers, which were classified as a high priority and had high marks of reviewers, to the faster publication. Manuscripts, which arrived to Edi-

torial Board, are reviewed by the experts from Editorial Board, and, if it is required, could be sent to the external review.

The prerogative of Editorial Board is to choose the reviewer. The manuscript will be sent to the review of the experts in this area of research, and the Editorial Board will make a decision about the fate of this manuscript: it could be accepted as it is, it could need an improvement and it could be rejected.

If manuscript does not fit to the above requirements, it could be rejected by the decision of Editorial Board.

If manuscript was sent back to authors for improvement as a decision of reviewers and editors, it will be reviewed second time, and after that Editorial Board will consider the possibility of the publication of this manuscript again. The date of the arrival of manuscript and the date of the acceptance of this manuscript to publishing will be indicated at the beginning of the paper.

Returning of the manuscript for improvement back to the authors does not mean that this manuscript is accepted to publishing. After improvement, this manuscript will be considered by Editorial Board again. The author should return back the text after improvement together with the original variant of manuscript and also with the answers to all of the comments. The date of arrival of the final version of the manuscript to Editorial Board is considered as date of arrival of the manuscript.

The improved manuscript should arrive back to Editorial Board during 1 week after author accepted comments; otherwise this manuscript will be considered as a newly arrived.

Editorial Board uses e-mail on the all of the steps of the communication with authors; authors must be very careful with the e-mail that is indicated in the manuscript, and in case of any changes immediately provide the information about changes to the Editorial Board.

After the approval of the layout of the journal, Editorial Board will send the proofreading to authors via e-mail in the PDF file.

On the step of correction, it is not possible to change the text, figures of tables. If it is necessary, this question will be considered by Editorial Board.

REFERENCES FORMAT

For books: author's surname and initials, full name of the book, the place of publishing, publishing company, year of publication, volume or issue and total number of pages.

For periodical issues: author's surname and initials, the name of the journal, year of publication, volume, number, first and last page of the paper.

Bressanelli S., Tomei L., Roussel A., et al // Proc. Natl. Acad. Sci. USA. 1999. V. 96. P.13034–13039 (if there are more than 5 authors), and all authors, if there are less than 5 authors.

References to the author's abstracts of the dissertations should contain author's surname and initials, title of the dissertation, the place where work was done, year of the dissertation's defence.

References to patents should contain surnames and initials of authors, the type of the patent's document (certificate of recognition or patent), number, the name of the country in which document was issued, index of international classification of inventions, year of the patent's issue.

The list of references should be represented on the separate page.

Tables should be represented on the separate page.

Comments to figures should be represented on the separate page.

For the communication with Editorial Board, use e-mail addresses: rpetrov@strf.ru, knorrevd@gmail.com, vsychev@strf.ru, tel.: (495) 727-38-60, (495) 930-80-05