Letter from the Editors

Dear friends,

We would like to bring you this seventh volume of our journal. As usual, the volume begins with review articles on promising aspects of life sciences. A review by V.A. Stepanov focuses on predicting one's susceptibility to various diseases based on knowledge of one's genome structure and ethnicity. Studies in this direction will establish a basis for personalized medicine, which is set to become a major component of the healthcare system in a decade. A review by A.P. Sokolenko addresses the situation around breast and ovary cancer in Russia, emphasizing the role of hereditary factors. This appears to be a major problem, since a considerable level of mortality is associated with these two types of cancers in women. Two review articles are dedicated to the molecular genetics of eukaryotes. These include an article by E.V. Dementyeva and S.M. Zakiyan on the dose-dependent compensation of genes in sex chromosomes and a review by E.P. Galimov dedicated to oxidative stress and apoptosis.

As always, the experimental articles in this volume cover a wide range of topics in both fundamental physico-chemical biology (M.V. Zagoskin *et al.*, R.V. Reshetnikov *et al.*, K.V. Tchernorizov and V.K. JЬvedas) and biomedicine (A. N. Glushkov *et al.*, A.V. Maksimenko *et al.*, L.E. Salnikova *et al.*). In our opinion, a particularly interesting article is S.S. Shishkin's Proteomics of Prostate Cancer Database, which is a description of an open database created by the authors. This database will be useful to a broad audience, including specialists in proteomics and cell biology, as well as practicing oncologists.

In the present volume, the Forum section is dedicated to the development of national technological platforms, whose launch was announced recently by the Ministry of Science and Education of the Russian Federation. The ideology here is that these platforms would bring together scientists, representatives of the business community, and the state in order to foster a coordinated innovative development in key economic areas. An article by I. Sterligov sums up the experience of the EU, where similar platforms have already been created and are now substantially contributing to technological progress. An article by M. Muravieva contains elaborate comments on the Russian Federation government's initiative and sheds light on how the platforms work. The article offers useful guidelines for the reader to find his niche within the structure of the platform. Particular examples of the integration of the efforts of science, business, and the state can be found in the article by E. Novoselova. This article describes the formation of so-called "pharmaceutical clusters," which were discussed at the International Forum "Innovative Drug Research and Development in Russia" (Moscow, November 17-18).

We are pleased to inform our contributors and readers that our journal has now evolved into its fully fledged form. We have finally succeeded in signing a contract with PubMed. The database is currently being replenished, and hopefully, our articles will soon be trackable online. The primary goal now is increasing the scientific quality of the published articles in order to increase impact in the future. This is not an easy task, but we hope that, together, we will succeed.

Good luck in 2011,
The Editorial board

Docking approaches are further improved by implementing new algorithms of the conformational search and new scoring functions (methods to estimate the free energy of ligand binding). Scoring functions may include either components of molecular mechanics force fields [2] or empirical terms, e.g. hydrogen 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 notable role in their recognition by protein-targets. We have recently shown that an explicit account of stacking in scoring functions increases the efficiency of ATP docking [5]. The aromatic interactions were identified by the mutual orientation of two cycles described by geometrical parameters: the height h and displacement d of one cycle 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 still not very well defined and usually taken as arbitrary [6, 7]. Defining it more accurately would assist in developing more efficient scoring functions and should increase the prediction quality of the spatial structures of protein-ligand complexes by molecular modeling methods. With this aim in view, we performed an analysis of the spatial structures of protein-ligand complexes determined experimentally with atomic resolution where ligands contained adenine or guanine as a substructure.

One well-known example of stacking interactions is the parallel packing of purine 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 [7, 10] and for model systems of carbon aromatic cycles (benzene and naphtalene) [11–14]. Besides, such compounds participate in cation-π interactions, where a positively charged group interacts with the negatively charged cloud of aromatic π-electrons [15–17].

Taking all that into account, we analyzed the distribution of geometrical parameters h, d, and α for contacts of adenine and guanine moieties of ligands with the aromatic side chains of receptor amino acids Phe, Tyr, Trp, and His, as well as with the positively charged guanidine 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, shown in red and green, respectively). The displacement d lies in the same range 0–3 Å for both types of contacts. Meanwhile, they clearly differ in the value of height h, which is <4.5 for parallel Å and <5.5 Å for perpendicular orientation. Similar distributions were obtained for Tyr, Trp, and His, though the data are scarcer in these cases. However, the T-shaped contact is not as typical for Tyr, Trp, and His as it is for Phe.

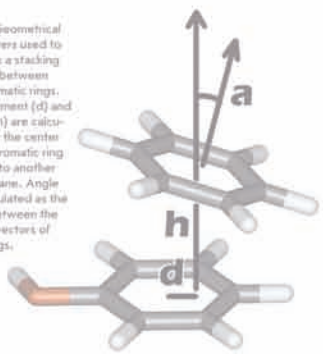
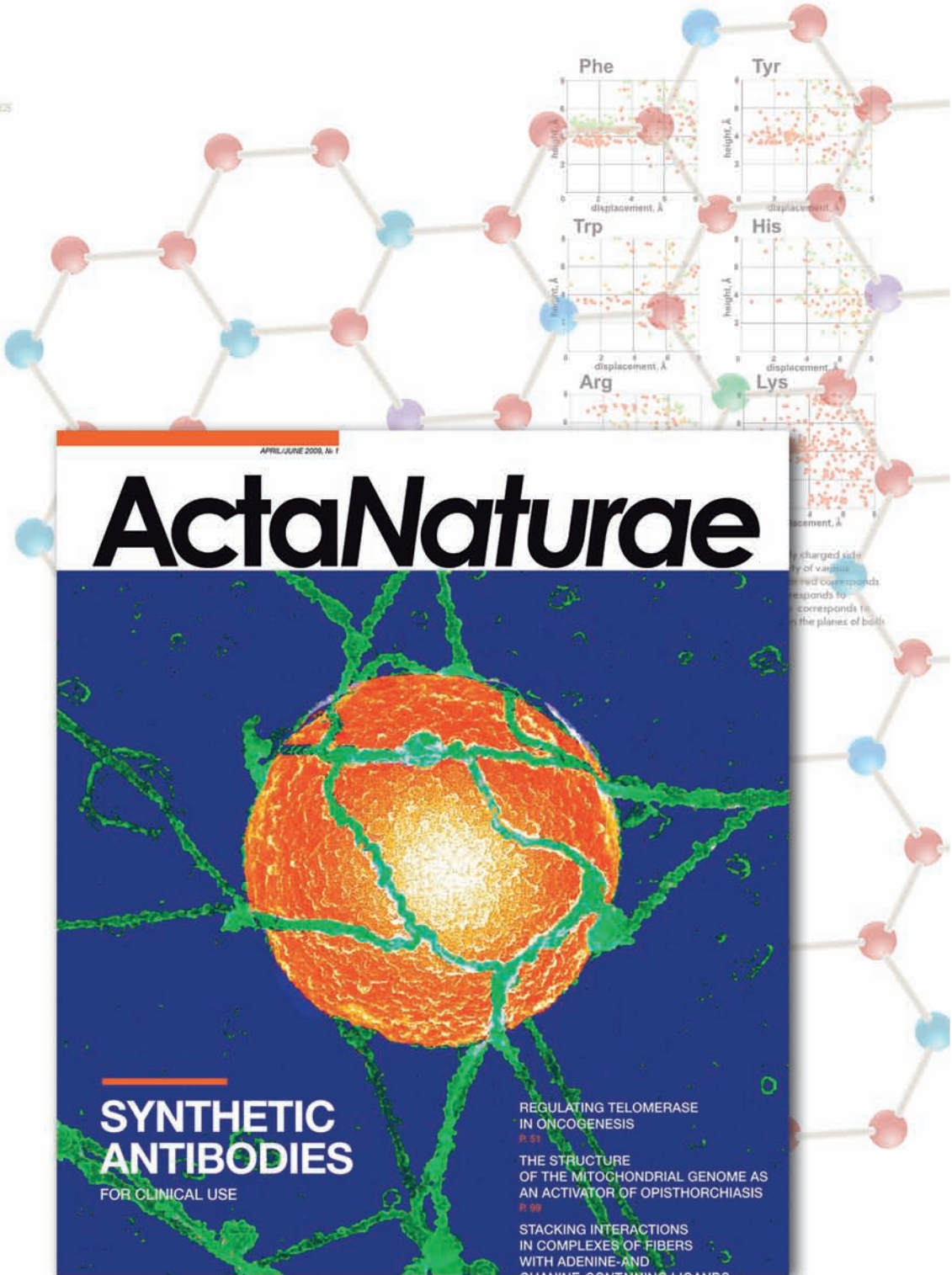
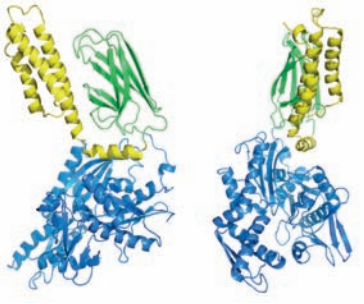


Fig. 1. Geometrical parameters used to describe a stacking contact between two aromatic rings. Displacement (d) and height (h) are calculated for the center of one aromatic ring relative to another ring's plane. Angle α is calculated as the angle between the normal vectors of both rings.



K. A. Chernorizov, V.K.Švedas

Modeling of the Full-Size 3D Structure of Human Chaperone hsp70 and Study of Its Interdomain Interactions



Model hHsp70_2p32 after molecular dynamics simulation. ATPase domain is colored blue, αSBD - yellow, βSBD – green

Hsp70 is a chaperone protein that participates in the folding of de novo synthesized proteins, protection of the hydrophobic regions of denaturated proteins, the regulation of apoptosis, the immune response, and several other cellular processes. Several probable full-size models of human Hsp70 have been constructed based on the structures of individual domains and their components from different organisms and using molecular modeling methodology. As a result of such an analysis, the most adequate model was selected. Based on the performed molecular modeling, the scheme of the mechanism triggering ATP hydrolysis and leading to the separation of ATPase and the substrate-binding domains was proposed.

Four-stranded intermolecular parallel G-quadruplex

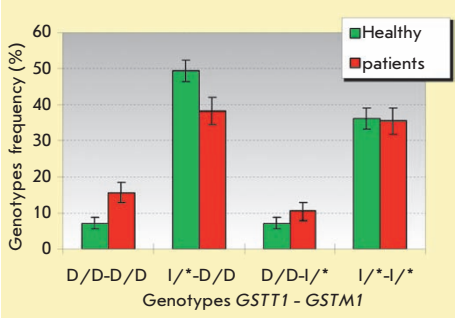
R. V. Reshetnikov, A. M. Kopylov, A. V. Golovin

Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets

The present work is devoted to the analysis of the G-quadruplex DNA structure using the bioinformatics method. The interest towards quadruplex DNAs is determined by their involvement in the functioning of telomeres and onco-promoters, as well as by the possibility to create on their basis aptamers and nanostructures. Here, we present an algorithm for a general analysis of the polymorphism of the G-quadruplex structure from the data bank PDB, using original parameters. 74 structures were grouped according to the following parameters: the number of DNA strands, the number of G-quartets, and the location and orientation of the connecting loops. Hence, the correlation between the twist angle and the tension in the structure of quadruplex DNA is revealed.

L. E. Salnikova, N. I. Zelinskaya, O. B. Belopolskaya, M. M. Aslanyan, A. V. Rubanovich

Association Study of Xenobiotic Detoxication and Repair Genes with Malignant Brain Tumors in Children



Frequencies of GSTT1-GSTM1 allele combinations among children with malignant brain tumors and in the healthy group

This study presents the results of research on DNA polymorphism in children with malignant brain tumors (172 patients, 183 in the control group). Genotyping was performed using an allele-specific tetraprimer reaction for the genes of the first (*CYP1A1* (2 sites)) and second phases of xenobiotic detoxication (*GSTM1*,*GSTT1*, *GSTP1*, *GSTM3*), DNA repair genes *XRCC1*, *XPD* (2 sites), *OGG1*, as well as *NOS1* and *MTHFR*. The increased risk of disease is associated with a minor variant of *CYP1A1* (606G) ($p = 0.009$; OR = 1.50) and a deletion variant of *GSTT1*, ($p = 0.013$, OR = 1.96). Maximum disease risk was observed in carriers of double deletions in *GSTT1-GSTM1* ($p = 0.017$, OR = 2.42).

ActaNaturae
OCTOBER–DECEMBER 2010, VOL. 2 № 4 (7)

Founders
Ministry of Education and
Science of the Russian Federation,
Lomonosov Moscow State University,
Park Media Ltd

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

Номер подписан в печать 18 декабря 2010 г.
Тираж 200 экз. Цена свободная.
Отпечатано в типографии «МЕДИА-ГРАНД»

CONTENTS

RESEARCH ARTICLES

Technology Platforms: Joining Forces, Establishing Dialogue

Marina Muravieva

The Russian government has begun creating a series of Technology Platforms (TPs). These platforms are meant to serve as a tool that should help close the gap between science and industry, encourage innovation at enterprises, and allow the government to focus its financial resources on the type of research and development that is of interest to business. The Ministry of Economic Development has issued a request for applications from initiators of TP development, and the government will determine the number of TPs based on the results of a selection process led by experts.

Nominally, the workgroup of the Governmental Commission on High Technologies and Innovations headed by **Andrey Klepach**, the deputy minister of economic development, is responsible for the formation of the federal panel of Technology Platforms. The idea of establishing a federal panel of TP was borrowed from the European Union, where such platforms have been in existence for several years. The Russian Ministry of Education and Science and the Ministry of Economic Development are responsible for the regulatory aspects of the process, and the Ministry of Industry and Trade actively participates in the preparation of the panel. The Russian government sees the TPs as the tool that will help link the efforts of the state, business, and science in addressing innovation challenges, as well as in setting and pursuing long-term scientific and technological priorities across different sectors of the economy, at the junction of different industries.

The term "Technology Platform" has, therefore, been a permanent fixture in documents of the European Commission for quite some time. Experts say that the fact that Russia has decided to create something similar should be welcomed and supported, It is necessary to bring representatives of business and researchers onto one platform in order to encourage a dialogue between them and identify what they have in common and their interests. That was the thinking behind the creation of TPs in the European Union. As a result, the thrust of many scientific research groups in Europe has been toward satisfying the needs of their domestic consumer markets.

Andrey Klepach, the deputy minister for economic development, said during a governmental commission in early August that the TPs should help solve several important tasks in Russia, the first of which is to bolster innovation at enterprises, The creation of the TP is a logical step in the development of a private-public partnership in research and development, innovation, and high technology. The TPs will gauge the interest future users (consumers) might show for the created technologies, making it easier to raise funds from private sources. Another goal, which should be taken up by the government using the TPs, is to channel government resources toward research and development, which are both in demand by business. This is extremely important, because the framework of the national innovative effort is unbalanced. The state is the main customer and the purveyor of funds for innovative development; it contributes about 70% of investments, while the private sector only contributes 30%. To reverse the situation, the government is launching a process of "necessity of innovation," requiring private companies (so far only those in which the state has a stake) to plan their innovative development. It is likely that the TPs that are in the TP federal register will be linked to these plans.

REFERENCE:

The Government assumes that "Technology Platforms" are a communicative tool for enhancing efforts in:

- the creation of promising commercial technologies, novel products/services,
- the attraction of additional resources for research and developments involving all interested participants (business, science, government, civil society)
- improving the legal framework in the field of scientific-technological and innovation development.

STATE BUDGETARY FUNDING — THROUGH PLATFORMS

The new government initiative to create TPs has sparked a great interest in the professional community. This is easy to explain; it is assumed that after a while serious budgetary funding will only be allocated through TPs.

"TPs are considered by the government to be an important element in the reform of the scientific and applied sectors," says **Alexey**

Alexey Khokhlov

Khokhlov, vice-rector in charge of innovation at Moscow State University. In particular, it is assumed that the bulk of funding to be allocated through grants will be channeled through the TPs. TPs should be regarded as a global expert platform that will determine the direction favored by business and the scientific community. What these TPs will do is help scientists formalize their projects and describe them using language that is clear to business. "If Russian TPs are made similar to European ones, then they could be the platforms that provide a certain set of documents, foresight, certain proposals, and plans. These platforms will be advisory. This is a great collective intellectual resource," says **Vladimir Popov**, director of the Bach Institute of Biochemistry, Russian Academies of Science.

There will be many TPs, and each one will compete for funding in certain fields. The state has identified certain areas of priority, whereas the remaining fields would receive funding only if self-organized communities of experts are sufficiently representative and able to convincingly lobby their interests.

"I still do not understand the work of TPs very well," said Academician **Vsevolod Tkachuk**, dean of the Department of Fundamental Medicine, MSU, "but I think success would depend on the business and research teams involved." The question is whether a TP can gain credibility and influence the competition for project funding. In previous years, calls for competition were opened by another mechanism, and it was sometimes unclear as to why the government believed one area to be more important than others. Now the government has explained that the chosen area is proposed by the TP, not an individual. And every scientist that is a TP participant is able to convince his colleagues that the call should be opened for the given field. I believe that this would be another important mechanism supporting scientific projects."

There will be a transitional period that will last about a year or two until the platforms are fully formed. For now, another particular tool will be used in the allocation of funds.

HOW TO SET PRIORITIES

Given the European experience, one can distinguish three stages in the development of a platform. A concept is to be developed in the first stage, explaining why the TP is necessary in a given field. In Europe, they are all initiated by big business. The strategic plan for research and development that needs to be developed is determined in the second stage. The methods for transforming research and development, and the probable results that could be achieved after 3-7 years of work on the projects supported by the TP, are to be determined in the third stage.

The main issue generating much controversy in the discussion of TPs for Russia is associated with the choice of priorities. Deputy Head of the Department of Priority Projects in Science and Technology at the Ministry of Education of the Russian Federation **Mikhail Puchkov** believes that it is important not to focus on solving small-scale industry problems. In economics, the boundaries between industries are increasingly blurred; therefore, the focus of the TPs should be on interdisciplinary problems. The question is: What are these problems?

"I see the technology platforms as a giant funnel involving a huge amount of resources, thus depriving them of other interesting programs,

REFERENCE:

Goals of Russian TPs:

- Technological modernization of the economy,
- Increasing the competitiveness of particular industries,
- Reducing resources consumption in primary-good production
- Solving social problems (health care, safety, ecology, education, culture),
- Stimulating the development of a new high-tech market and new ventures in these areas.

says **Alexander Smirnov**, general director of the Association of Military-Industrial Complex Manufacturers of Medical Devices and Equipment. "Could it occur that Russia, while engaged in the creation of TPs, will fall behind in many strategically important areas, only because promising developments would be out of the scope of the platforms?"

The experts and initiators of the first TPs agree that such risks always exist when setting priorities. However, they are certain that the implication of partners from the business community should at least halve the risks. Besides, expert TP groups should be carefully built – this will be the first stage in the TP creation. This process must be approached carefully. The experts have to determine which basic technologies would underline the TPs. This strictly limited list will be recorded in the passport of the platform. The competent and coordinated work of the platform experts will help avoid the duplication of studies that are funded from different sources.

FIRST INITIATIVES

Basic Principles of technology platform:

- Clear direction toward serving the interests of society, business, the state;
- Significant representation of business (a minimum of 50 percent);
- Middle- and long-term oriented projects;
- Formation of educational programs for staff training and retraining;
- Orientation on the expansion of cooperation; openness, publicity.

The Lomonosov Moscow State University made a focal point for the creation of four TPs within the limits of the development program for the innovative framework, which has been developed according to Government Regulation 219. The work at the university was launched as early as this spring. Alexey Khokhlov has stressed that MSU only plays the role of a coordinator; however, there are no "major" or "minor" participants. All participants are equal. A TP represents a voluntary association of organizations of any ownership type, governmental or nongovernmental agencies, professional associations, and professionals who share the goals and objectives of the platform. The issue of organizations joining a platform has not been formalized on legal grounds yet. Besides, the creation of a TP is a dynamic process, and the lists of participants remain open or will be created.

The first-proposed TP is "Strategic Information Technologies." Several large areas are determined in the framework of this platform. One of them is devoted to the creation of new computer architectures for exaflop generation calculators (the next generation of computing power). The second is devoted to hybrid architectures; the third, to engineering calculations, such as the design of various mechanisms. Other directions are associated with distinct industrial branches, particularly with pharmacology (development of physiologically active substances), the oil and gas industry, and materials (prediction/calculations of material characteristics based on their molecular structure).

The second TP is "Nanomaterials for Energy Efficiency." This refers to solar panels, fuel cells, new types of batteries and energy storage, "smart home" materials, and polymer nanocomposites for transport that will make electric cars or planes lighter.

Alexey Konov, Executive Director, The Bioprocess Group.

My opinion is that everybody looks at the TPs based on their own point of view: employees of the Russian Academy of Sciences (RAS) and universities understand it to be a new source of funding, like the Federal Target Program; officials in the ministries see it as a mixture between foresight and an attempt to woo business, etc. In actuality, the TP, in my opinion, is an attempt to outline a "set of possibilities," or "window" for tomorrow and provide steps for "entering it" in time. As for the TP "Post-genomic and Cell Technologies in Biology and Medicine," there are some simple things it could do:

Determine the direction of development in biotechnology on a global scale in the next 20 years and the choice of priorities for Russia;

Determine the best existing core competencies in Russia, as well as missing or poorly developed ones;

Generate proposals for supporting the best of the selected core competencies, as well as proposals for changing, enhancing, or closing the weak ones and, if necessary, developing new ones from scratch;

Generate proposals for improving the legislative and regulatory framework, and ensure that promotion of the selected competencies is free and transparent;

Help with the formation of the best competencies into distinct projects, help in linking projects to investors, support of projects after investment, and lobbying for the supported projects;

Identifying the competencies (persons, scientific groups) that cannot be packaged as projects but are advanced scientifically and can yield breakthroughs in a few years.

Assistance for such projects is possible through the formation of an environment with the most favorable conditions – financial (grants, non-performing loans, etc.) and organizational support (establishment of laboratories under project leaders at the Institutions of the RAS, RAMS, Universities).

In addition, two other platforms are associated with biology, an area that is on the rise and is growing rapidly.

The idea only emerged two months ago on creating a platform for "Post-genomic and Cell Technologies in Biology and Medicine." Since this time, much work has been done. The Action Team conducted a survey of experts who recommended participants in the TP and identified the main problems. A memorandum has been prepared, and it was signed by 20 companies. The goals of this TP are scientific-technological and innovative development of post-genomic and cell technologies for appropriate development of the Russian economy; improvement of the normative-legal regulation in this area; and consolidation of the Russian medical and biotechnological community to lobby its interests. The goal for the near future is to be included on the list of Russian TPs.

The "Industrial Biotechnologies and Bioenergetics" TP was initiated by the public corporation Ros-technologies. Its importance was stated by **Vladimir Popov in these words**: "While the rest of the world is focused on the creation of a bio economy, which is set to define the XXI century, Russia lacks any official high-level documents governing the development of biotechnologies. Biotechnologies will account for up to 3% of total GDP in developed countries, according to experts."

"Russia lost its biotechnological economy during the last 20 years, although it had been second only to the U.S. prior to the advent of Perestrojka," Popov said. Now Russia lags far behind on the international market, with a tiny market worth 2 billion dollars (which is 0.2% of the world market).

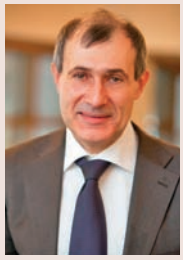
"The participants in the technology platform aim to steer development in this direction. Our aim is to consolidate the biotechnological community and lobby its interests at all levels of government," Popov says. At present, about 30 organizations are on the list of this TP. A memorandum has been developed. The next steps are expanding the list of participants, identifying possible sources of funding, preparing the conception of development, and being included in the Russian Federal TPs.

HOW MANY PLATFORMS CAN BE DEVELOPED

The federal TP list is currently under development. It is clear that the TPs that are in line with the set priorities in technological development are at the top of the list, as claimed in the document Conception of Long-term Development up to 2020 in Russia. Among them are the creation of next-generation aircraft and energy-efficient engines, the construction of safer nuclear power plants, the development of hydrogen-based energy sources and production of new motor fuels, the development of optoelectronics and micromechanics, special equipment design for the Arctic and other extreme environments, and the development of new technologies for metal processing.

The Ministry of Economic Development continued to accept proposals for the creation of platforms until November 25, 2010. The approved projects, after they have been reviewed in cabinet offices, will be sent over for approval to a Government Commission workgroup. The selected platforms have not been announced yet. Experts have opined that about 10-15 projects will be approved. ●

Alexander Gabibov, professor, corresponding member of RAS:



The creation of novel structures or new "communication tools" makes sense only if old ones operate poorly or require radical improvement. It is obvious that the new form, virtually developed rather recently in the Western world, can result in the creation of novel products, but only until *three main components*, i.e., science, production, and management, adopt equally competent approaches in solving problems. Only equal development of these three components can provide stable growth of the economy; otherwise, the idea of platforms will lose its shine. Unfortunately, prerequisites exist for such a pessimistic forecast. While the level of scientific research in Russia, which is constantly under criticism, can be evaluated in each particular case using internationally accepted benchmarks, the evaluation of the two other components is substantially more difficult. Our biotechnology business is extremely backward, and the quality of management doesn't stand scrutiny at all. The results of scientific research can be chosen quite consciously. However, those who moved into the biotechnological business are mostly former academic misfits who lack advanced training and are poorly equipped to work with staff. Unfortunately, this is obvious at various scientific and practical conferences, seminars, forums, and exhibitions. People are involved in promoting well-known products which can hardly be characterized as innovative. Only competent staff able to solve problems relating to the training of production managers, biotechnology engineers, and biotechnological production managers competent in manufacturing can make the TP program a reality.

National Technology Platforms: The European Experience

Ivan Sterligov

The Ministry of Education of the Russian Federation has announced the start of the formation of national technology platforms to ensure coordinated development of innovation in key economic areas. The mechanism of the platforms is borrowed from that of the EU.

Coordinating the work of members of the scientific community and technological progress is a fundamental problem. The life of a society improves or diversifies only when scientific ideas are transformed into technology and have an application. Normally, science, development, and technology can be likened to a swan, a pike, and a crab, respectively: the odds that all will move in the same direction are very small. Scientists, engineers, and managers pursue different goals and use different ways to measure their accomplishments. To ensure progress, it is necessary to coordinate the work of all three groups.

Modern economic theory describes such coordination using the concept of *National Innovation System* (NIS) developed in the late 1980s by an Englishman, Christopher Freeman, and a Dutchman, Bengt-Åke Lundvall. NIS theory is, in fact, the underlying component of the entire development strategy of the European Union, and Freeman and Lundvall themselves were among the masterminds of the famous Lisbon EU strategy.

According to Freeman's definition, NIS *is the network of institutions in the public and private sectors whose activities and interactions initiate, import, modify and diffuse new technologies*. Since most EU economies are of a mixed, socialist-capitalist type, the role of government in the European model of NIS is particularly big. The EU is a key customer and consumer of this research and development, financing such research through their framework programs.

The current Seventh Framework Program (FP7) is designed for the years 2007-2013. The ten areas of thematic priority are highlighted in it as the most general tool for coordination, just as in the Russian programs. Most of the money is allocated to IT, health, transport, and nanotechnology.

The main instrument of coordination amongst all the players in the field of R&D has become the *European Technology Platforms*. Formally, they are not included in FP7, but they are closely linked to it. There are 36 such platforms, the first of which was created in 2002.

Each platform is designed around a specific group of commercially and socially important technologies, such as Photovoltaics, Water Supply and Sanitation Technologies, Industrial Safety, and Textiles and Clothing of the Future. The EU has adopted the opinion that the platform is formed from the bottom up, "But, in fact, business, investment and finance, research and community organizations put them into close contact with government agencies and services."

Such technology platforms do not have a legal status; they are open organization networks that formalize the industrial NIS. Their existence is subject to three phases:

1. Interested participants form a common vision for the development of subject fields, in meetings and discussions;
2. Jointly, but under the industry's leadership, a *Strategic research plan* is formed. In this regard, the needs for both medium- and long-term research and development are laid out;
3. The implementation of the strategic plan is carried out, involving private and public investors (example, through FP7 and national ministries and foundations).

One of the main objectives of these platforms is to help EU officials shape the subject contest in the FP7. At the same time, the EU is only funding the work of the platform secretariat, and the basic organizational costs are borne by the participants. Additionally, the EU ensures that the platform concepts are not eroded. The corresponding status of the "European Platform" is assigned only to the cohesive and motivated associations that have emerged around breakthrough technology trends.

For example, let us turn to the strategic plan for the technology platform "European Research Council in the Field of Road Transport." Here are some points of the plan:

In the years 2020-2025, working trials will be held to test automated traffic control, simultaneous braking and acceleration of vehicles, as well as keeping the distance between vehicles.
In 2010-2015, experiments will be done to evaluate the possibility of direct measurement for the friction of tires.
In 2010-2015, a full-fledged system of networked communications between vehicles, as well as vehicles and infrastructure, will be developed.

Long before the formal adoption, the plan will be posted on the platform's website, where it is freely available for discussion.

The platform includes more than a dozen core business associations (the Asphalt Association, Association of Automotive Components Suppliers, amongst others), several universities and nonprofit foundations, countries – members of the EU, the European Commission as a whole and its individual committees. A special role is played by such corporations as Bosch, Renault, Volvo, and others.

Over all, an executive board of five members manages the platform, led by Wolfgang Steiger, the director of new technologies for the Volkswagen Group. The Council meets about once a month, and the more illustrative meetings are linked to specialized exhibitions and conferences.

These Platforms are at different stages of development. A separate group is composed of the most advanced initiative associations, demanding particularly complex and expensive research. Joint technology initiatives (JTI) are specifically developed for them in FP7. To date, there are five such initiatives, and they all work in partnership with the "parent" platforms: Fuel Cells and Hydrogen Energy, Nanotechnology, Innovative Medicines, Embedded Electronic Systems, and Aeronautics and Air Transport.

Legally, these are public-private partnerships, each of which works between the European Commission, the countries concerned, and representatives of private business. The EU allocates 1-2 billion euro to each JTI on average for the period ending in the year 2013, a matching amount is provided by business. An open competition is used to select projects for funding, bringing together research centers, small businesses, and corporations. The main criterion for selection is scientific excellence. The first projects were selected in late 2008, but so far it is too early to judge the effectiveness of JTI.

However, there are sufficient monitoring results for the conventional technology platforms. In 2008, a survey of 950 organizations showed that in general there is greater coordination and harmonization of policies in organizations participating in such platforms. Does participation in the platform give access to financial resources? The respondents consent, but the degree of optimism on this matter is much higher for civil servants than for corporate managers, university professors, and especially, owners of small innovative firms.

There are problems in integration with FP7. The representatives of some platforms are happy with how their proposals are presented in the program competitions, while others believe that their proposals are being completely ignored. There are complaints that the efforts to establish and promote a strategic research plan do not correspond to success in FP7, and that contests are won by the same small group of applicants as always.

EU experts were faced with great difficulties when trying to obtain information about the activities in individual platforms. Sometimes their secretariats worked poorly, and the members rarely met to discuss.

Finally, the expectation that the platform will build and develop vocational education was not met. In this area there has been practically no activity. But in general, 93% of respondents reported that they would not have changed their decision to join the platform had they known in advance about their development. A study in 2009 confirmed the fact of the usefulness of such platforms, but it emphasized the low degree of participants' involvement in their work. Due to the informal nature of the platforms, the specific economic and statistical effects of their existence are not recorded.

The moderate success of technology platforms is combined with the overall modest achievements of European research and innovation policies. A key objective of the Lisbon strategy was to bring spending on science up to 3% of GDP by 2010, and this goal remains unfulfilled. Now experts are creating a new European Commission strategy. Technology platforms will be maintained, but they are encouraged to merge into a "technological innovation platform," organized in the cluster form. In addition to the strategic plans, the participants are now instructed to develop and present "Plans for Innovative Actions" to the European Commission, and conduct studies consistent with the EC forecasts. They will have to assess themselves joining high-risk projects and implement their own "program innovation." Only time will tell how the strategy of such enhancement of the platforms was justified. The EU is not planning on giving up, and European officials insist that this mechanism has enormous potential, referring to the platforms as the "Flagship of Europe." ●

Pharmaceutical Clusters: Remedy for Regional Economies

Elena Novoselova

The idea of pharmaceutical clusters is being actively pursued in Russia. Several projects have already been implemented, while others are in the pipeline. There are optimistic reports about regions setting up manufacturing, improving the level of education, and creating jobs, and about pharmaceutical companies receiving infrastructure and tax incentives. Is there reason to expect innovation with pharmaceutical clusters as well? That issue, along with others, was taken up at the International Forum "Innovative Drug Research and Development in Russia" organized by the Adam Smith Institute on November 17–18, 2010, in Moscow.

A cluster is a voluntary regional industrial association of businesses in active collaboration with research and social institutions and local governments, aimed at making their products more competitive and promoting economic development in the region.

More and more such clusters are currently being created in Russia in mechanical engineering, metal works, aerospace engineering, shipbuilding, IT, optoelectronics, instrument engineering, wood processing, agriculture and the agro industry, pharmaceuticals, etc. The creation of pharmaceutical clusters has been announced in St. Petersburg, Moscow and the Moscow region, Yaroslavl, Kaluga and the Sverdlovsk regions, Siberia, Stavropol Territory, and Tatarstan.

The effort continues to face a number of issues on this path: are the regions ready for pharmaceutical clusters? Will they be able to find companies willing to settle in the clusters? What kind of production should be developed (own brands of drugs, generics, or simply drug packaging)? How to lead innovative research and educational programmes? Will the pharmaceutical clusters help replace imported drugs? What is the legal framework? How will the efficiency of the cluster be assessed? And how many pharmaceutical clusters does Russia need?

The idea is that the entire chain of drug development and production will be implemented on the territory of a pharmaceutical cluster. Only few regions can claim to have created full-cycle clusters, including personnel training, research and development (R&D), and manufacturing. In Russia, it is rather more accurate to speak of the existence of special economic zones, which attract drugs manufacturers by the attractiveness of their pharmaceutical markets (which continued to grow even during the recent crisis) and active government support (the Pharma-2020 Strategy).

Apparently, Russia does not have enough companies to populate the mooted pharmaceutical clusters. In addition, attracting foreign companies is important for technology transfer and personnel training in drug manufacturing, as well as R&D. However, it is encouraging that big international pharmaceutical companies have joined the pharmaceutical clusters project. The Yaroslavl cluster for the pharmaceutical industry and innovative medicine meets all the criteria for a pharmaceutical cluster, and it has already attracted some serious players. One of the residents of that pharmaceutical cluster is the Swiss drug maker Novartis Pharma. On 7 September 2010, Hari Swen Krishnan, Director of Novartis Pharma Russia, and Sergey Vakhrukov, governor of Yaroslavl region, signed a Memorandum of Understanding between the Government of Yaroslavl region and Novartis Pharma for cooperation in healthcare and medical science. According to the Memorandum, the two main thrust of the cooperation will be to provide additional personnel development in healthcare and carry out research in innovative drugs, particularly including staff training for carrying out more clinical tests. In 2011, Novartis Pharma plans to create a Centre for Clinical Studies at Yaroslavl State Medical Academy. The collaboration in clinical research has been ongoing in the Yaroslavl region for 7 years; leading academic institutions and hospitals in the region, in partnership with Novartis Pharma, have participated in 17 international clinical studies of 14 parameters.

An inter-institutional research and educational innovative center connecting leading educational institutions is under development in the region. It will provide specialists for innovative R&D in pharmacology and medicine. In addition, special personnel training programmes for pharmaceutical companies are being set up at educational institutions. For instance, Novartis Pharma has started a joint educational programme with Yaroslavl State Medical Academy comprising special personnel training modules in medicine and pharmaceuticals with internships at the company's plants, as well as teaching staff training modules. Nycomed, together with Yaroslavl State Medical Academy and Yaroslavl Technical University, based at the Chemical Engineering school, has launched a technical personnel-training programme for the

pharmaceutical industry. On September 1, 2010, the first group of 25 students began studies in 3 professional programmes.

In June 2010, Nycomed also started construction of a plant in the Yaroslavl region for manufacturing drugs for the treatment and prevention of neurological and cardiological diseases. The company has invested 75 million euro in the construction and plans to begin production as early as 2014. R-Pharm and ChemRar high-tech Centers have announced the construction of a plant for producing drug active ingredients in Yaroslavl region. The project, known as Pharmoslavl, is aimed at making healthcare in Russia less dependent not only on the import of drugs, but also on the import of drug ingredients.

The first production of nanotechnological flu vaccines in Russia will begin at the Yaroslavl pharmaceutical cluster, with Rusnano and NT Pharma Ltd. as investors. The creation of the new entity, NT Pharma, instead of investment in the existing Immapharma, helped streamline financial audit and ensure Rusnano's 49% participation in the project. Rusnano has invested 1.5 billion rubles initially, and additional funding will be obtained by reinvesting part of the profits from swine flu nanovaccine sales.

The Kaluga pharmaceutical cluster is another example of a regional association of educational, research, and industrial organisations in pharmaceutics.

In 2006, Hemofarm, a member of the German STADA group, built a plant in the city of Obninsk (Kaluga region) with maximum annual capacity of 2.5 billion tablets, with 32 million euro invested in construction. Novo Nordisk has announced the construction of an insulin plant in the region. According to the Memorandum of Intent between Novo Nordisk and the government of Kaluga region, $80–100 million will be invested in construction. The agreement sets the partners' responsibilities for creating the necessary infrastructure (power, gas, water supply, roads, etc.) and provides general guidelines for land lot purchase. Berlin-Chemie AG, a member of the Menarini Group, has also signed an agreement on establishing production in Kaluga region. In the first stage of the project, with an investment of 30 million Euro, a packaging line will begin operating in Kaluga in 2013, and drug production will begin in 2014.

In September 2010, two Russian pharmaceutical companies, PharmSynthesis and NIARMEDIC PLUS, announced plans to establish production in Kaluga region. PharmSynthesis will build a full-cycle production and packaging facility for oncological drugs and diagnostic products made from proprietary ingredients. The project is scheduled for completion in summer 2011. The company will invest about 10 million Euro by 2012. NIARMEDIC PLUS will create a full-cycle production facility for the company's two key products: an antiviral interferon inducer and a collagen material used in surgery, trauma treatment, and orthopaedic care.

The Russian venture company Seed Fund is invested in shaping the pharmaceutical cluster of the Kaluga region. The Fund has committed its first investment into local projects proposed by the regional enterprises: the Cardiomarker myocardial infarction diagnostic system (Obninsk Pharmaceutical Company Ltd.) and a renal carcinoma drug (OncoMax Ltd.).

For the move from industrial parks to regional clusters to be successful, designing educational programs, training personnel, localizing R&D in the cluster region, and creating clinical test centres are very important. Thus, a few years ago, in the framework of the Kaluga region pharmaceutical cluster, a Department of Medicine was established at Obninsk State Technical University of Nuclear Energy (subsidiary of the National Research Nuclear University MEPHI).

A personnel training centre for the pharmaceutical industry has been created in the Kaluga region in collaboration with the Berlin Centre of Professional Training, a center that specializes in the pharmaceutical and chemical industries, as well as with People's Friendship University of Russia and the International Pharmaceutical Federation. On September 1, 2011, 300 students will begin studying in more than 20 educational programmes.

In his speech at the "**Innovative Drug Research and Development in Russia**" forum, Ruslan A. Zalivatskiy, the minister of economic development of Kaluga region, said that the Kaluga region government had created all necessary conditions for the success of investment projects. Thus, the industrial parks created provide companies with a complete engineering and logistic infrastructure; there are land plots for lease and sale, as well as ready-to-use production facilities available for long-term lease. This philosophy of the regional government has already yielded 56 active investment projects, with 5 more that were to be launched by the end of 2010.

On April 22, 2010, the St. Petersburg region government approved the concept of the St. Petersburg pharmaceutical cluster. St. Petersburg has all that is needed for a successful development of an innovative pharmaceutical industry: scientific and production staff, research institutes, drug developers, 11 academic institutions specialized in chemistry and medicine, and a broad clinical base. Currently, a dozen large and middle-size enterprises manufacture a wide range of drugs in St. Petersburg. Valentina Matvienko, the mayor of the St. Petersburg region, says that pharmaceutical projects will become a priority for the city's economy and will be supported by the government. An inter-institutional research centre incorporating the St. Petersburg State University of Technology, Pavlov St. Peters-

burg Medical University, and the St. Petersburg State Chemical-Pharmaceutical Academy was created to coordinate implementation of the concept.

The shaping of the St. Petersburg pharmaceutical cluster began with the creation of the Neudorf special economic zone; the Novoorlovskaya special economic zone will follow, and then the Pushkinskaya industrial zone. The first enterprises to join the new cluster were the Russian companies Geropharm Ltd. (injection drugs, up to 55 million vials, 1.3 billion rubles invested), Biocad (original and copied oncological, neurological, urological, and gynaecological drugs, 1.07 billion rubles invested), Neon Ltd. (medical and pharmaceutical equipment, pharmaceuticals, 910 million rubles invested), and Samson-Med Ltd. (drugs based on natural active ingredients, 1.5 billion rubles invested).

On October 27, 2010, the St. Petersburg Legislature approved a draft law On Amendments to the St. Petersburg Law On Tax Exemptions. The profit tax rate for investors will be reduced to 13.5%, and investors will be exempt from property tax. Until now, the profit tax rate, which varied depending on the invested amount, could reach up to 15.5%. In addition, the new law sets the uniform threshold investment for an organization to qualify for tax exemptions at 800 million rubles. Under existing legislation, only organizations investing 3 billion rubles and more can enjoy the 13.5% profit tax rate. The new law also extends the exemption term from 3 to 5 years; at the same time, the investor will now have to invest funds during 3 years, instead of 1, as is the case now. The law also sets some requirements for investors in the high-tech sector. For example, profit tax deductions start at a 50-million-ruble investment.

These are just a few examples of the implementation of the pharmaceutical cluster concept. Will the clusters meet the expectations of business, the state, and regional players, and how will one be able to assess the efficiency of a cluster? Should one rely on the number of participants (plants and research centres), the number of registered and produced drugs, the amount of money invested, or annual production volumes?

Regional governments are striving to create a favourable investment climate in the clusters, not only for manufacturing but also to spur innovation. The federal government supports the idea of creating pharmaceutical clusters, since the clusters are seen both as a tool for an innovative economy and for addressing social issues such as unemployment and education. Naturally, the legal landscape needs adjustment to facilitate the trend. The companies involved in pharmaceutical clusters argue for organizing an Association of Russian Pharmaceutical Clusters, in order to lobby the interests of those in the clusters. In October 2010, the Committee on Entrepreneurship in the Healthcare Industry at the Chamber of Commerce and Industry proposed the idea of preparing an independent Development Concept of pharmaceutical clusters in Russia. Following the initiative, a workgroup will be created with the participation of representatives of the ministries of Economic Development, Industrial Trade, Healthcare and Social Development, the Russian Patent Bureau, as well as other agencies, regional governments, the scientific community, entrepreneurs associations, and the business community. The Committee also recommended that the Chamber propose that the Federal Government create a special chapter called Development of Pharmaceutical Clusters in Russia in the Federal target programme Development of the Pharmaceutical and Medical industries in Russia until 2020. ●

COMMENT

Victor Dmitriev, Director General, Association of Russian Pharmaceuticals Manufacturers

Many regions have announced the creation of pharmaceutical clusters. However, we have no legal definition of what a cluster is.

Regarding the Association of Russian Pharmaceutical Clusters, the Association of Russian Pharmaceuticals Manufacturers is not against the idea, but the goal should be made clear. Associations are formed in order to resolve certain issues and lobby interests.

Moreover, some do not understand what the purpose of building a cluster in the region is. Is it to create jobs, develop the pharmaceutical industry, and provide people with affordable drugs? Otherwise, to promote scientific development and create innovative drugs? Or perhaps to further someone's political ambition? Indeed, the purposes of a pharmaceutical cluster should be in line with government contracts. This means that the pharmaceutical clusters should lead to lower diseases and death rate, a higher quality of life—these are the main indices of health care.

I can point to a couple of promising projects with clear goals, scientific potential, a manufacturing base, qualified personnel, good infrastructure, investors, etc. With such inputs a pharmaceutical cluster will really be able to function properly. Thus, St. Petersburg is perfectly positioned to make a pharmaceutical cluster a success. They have scientific potential, educational facilities, manufacturing capabilities, companies willing to invest in pharmaceutical production, and investors. The level of scientific development in St. Petersburg makes me hopeful that the cluster will pursue innovation and not follow in someone else's steps. Everything is ready for the creation of new drugs. Yaroslavl and Kaluga are other good examples.

Genomes, Populations and Diseases: Ethnic Genomics and Personalized Medicine

V. A. Stepanov
Research Institute for Medical Genetics, Siberian Branch, Russian Academy of Medical Sciences
E-mail: vadim.stepanov@medgenetics.ru
Received 19.09.2010

SUMMARY **This review discusses the progress of ethnic genetics, the genetics of common diseases, and the concepts of personalized medicine. We show the relationship between the structure of genetic diversity in human populations and the varying frequencies of Mendelian and multifactor diseases. We also examine the population basis of pharmacogenetics and evaluate the effectiveness of pharmacotherapy, along with a review of new achievements and prospects in personalized genomics.**

ABBREVIATIONS **ADD – autosomal dominant diseases, ARD – autosomal recessive diseases, HVSI – hypervariable segment I, mtDNA – mitochondrial DNA, CD – common diseases, RFLP – restriction fragments length polymorphism, IHD – ischemic heart disease, COLD – chronic obstructive lung disease, SNP – single nucleotide polymorphism, CNV – copy number variation, STR – short tandem repeats, HapMap – Haplotype Map of the Human Genome, CEU – population of Central European origin, YRI – population of Yoruba from Ibadan, Nigeria, CHB – Chinese population from Beijing, JPT – Japanese population from Tokyo, CMT1 – Charcot-Marie-Tooth disease, type 1, GWAS – genome-wide associations search, OR – odds ratio, OMIM – on-line Mendelian Inheritance in Man catalogue.**

PERSONALIZED MEDICINE

The concept of personalized medicine, which puts the individual patient, with all of his specific peculiarities, into the center of attention, is not new. The 19th-century Russian physicians M.Y. Mudrov and N.I. Pirogov were well aware of this principle. "A doctor treats the patient, not the disease... Each patient needs special treatment depending on his physical constitution, even if the disease is the same," wrote Mudrov. The great medical practitioners of the past also acknowledged the prophylactic value of personalized medicine – "Healthy people should be kept in hand... they should be advised on keeping to a healthy lifestyle" [1] and "a disease is easier to prevent than to treat" [2]. A new appreciation and the real potential for personalized medicine have re-appeared in the age of molecular genetics. By the end of the 1990s, a new concept of genomic medicine had started to take form [3, 4], and this concept involved "the routine use of genotyping methods, usually in the form of DNA-testing, for improving the quality of healthcare" [3]. The modern understanding of personalized medicine is based on the principles of preventive medicine, which were postulated by Nobel-prize laureate Jean Dausset [5]. The outlines of this concept are eloquently described by the 4P medicine principle or system medicine, which was suggested by Leroy Hood [6–8]. This principle states that "reactive" medicine (which reacts to a disease and fights its symptoms) must turn into predictive, preventative, personalized, and participatory medicine; namely, medicine that will be aimed at predicting the disease before it manifests itself through symptoms, take into account any individual (mainly genetic) traits of the patient, as well as involve the active help of the patient in identifying his or her genetic traits and in determining preventive measures.

In Russia, ideas related to personalized medicine based on the advances in molecular genetics are being actively pursued at several genetic schools. One school, headed by RAMS full member V.P. Puzyrev [9-11], is developing the concept of genomic medicine. Another, headed by RAMS corresponding member V.S. Baranov [12-14], is developing the concept of genetic passports.

ETHNIC GENETICS

The geographical region, ethnic group, and population largely determine the genetic traits of an individual. Setting aside the issues of the substance, the terminological and meaningful differences in the geographic, ethnic and population levels of gene-pool organization, we shall review these terms in the context of personalized medicine in which they are essentially synonymous – they reflect the individual traits of a human that are dependent on his genetic origins.

It is agreed that a detailed understanding of genetic diversity in human populations is crucial for determining the genetic basis of most common diseases [15].

Approaches aimed at identifying genetic similarities between various ethnic groups and populations, and which involve the study of polymorphic genetic markers, have been used in evolutionary and population genetics since the middle of the 1950s. Initially, protein polymorphisms played the role of genetic markers [16–18]. However, as molecular genetic techniques improved, population studies were reoriented toward various classes of DNA markers, with non-recombinant lineages of mtDNA and Y-chromosomes being the most widely used [19–21, etc.]. These studies allowed researchers to form an understanding of the main stages of population spread and of the ethnic divergence of modern humans. This also resulted in the appearance of a new scientific field: ethnogenetics. According to Balanovsky and Rychkov [22], ethnogenetics is a branch of population genetics which "pays special attention to the ethnic structure of populations and attempts to identify the genetic results of the ethno-historic and ecological development of human populations."

Currently, the most useful tools for describing genetic variability in various ethnic groups and populations are genome-wide sets of single nucleotide polymorphisms (SNPs), which are sometimes complemented by copy number variability (CNV) data [23–26, etc.]. A promising method that is sure to be used in years to come is the re-sequencing of complete genomes in representative cohorts from different populations.

Studies in ethnogenetics are among the most productive areas of genetic science in Russia, and such studies are being actively pursued in a number of research facilities [27–31].

The importance of population genetics for personalized medicine also derives from the fact that knowledge of the role of genetic variability in the pathogenesis of common diseases can only be obtained by a detailed analysis of the associations between genetic markers and diseases on large cohorts of patients and healthy people from various populations. Specifically, one of the most productive approaches for association analysis, the so-called Genome-wide Association Study (GWAS), requires the testing of hundreds, if not thousands, of individuals and replication of the discovered associations in other populations.

The cooperative development of ethnic genetics, the genetics of common diseases, and the concept of personalized medicine spawns a number of key questions. Answers to these questions will determine to what use and how quickly genetics will be adopted in predictive medicine:

- How marked are interethnic differences in disease incidence and disease susceptibility gene frequencies?
- What are the evolutional mechanisms behind the differences in disease gene frequencies?
- Do racial, ethnic or geographical origins influence the impact of distinct genetic variants on the course of a disease?
- To what degree does genetic diversity account for the differences in the spread and outcome of diseases between racial and ethnic groups?
- Is there any need for information on the racial/ethnic origin of patients for medical research?

Drawing up a picture of the current understanding of the answers to these questions is the main aim of this article.

STRUCTURE OF GENETIC VARIABILITY IN HUMAN POPULATIONS

How different are human populations genetically? Human population genetics give a precise answer to this question – interpopulation differences in a global sense (comparing the populations of different continents) are responsible for 10–15% of the genetic variability in humans (Table 1). In other words, the Wright Fixation Index (F_{st}) is 0.10-0.15

Table 1. Genetic Differentiation of Human Populations

Marker type	Populations	F_{ST}	Reference
Classical markers			
Blood types	World	0.16	[40]
Protein polymorphism	World	0.11	[40]
DNA-markers			
RFLP	-"-	0.11	[21]
dinucleotide	-"-	0.11	[41]
trinucleotide	-"-	0.04	[42]
tetranucleotide	-"-	0.04	[21, 43]
Microsattelites and RFLP	-"-	0.15	[44]
Alu-repeats	-"-	0.12	[45]
Alu-repeats	-"-	0.10	[46]
mtDNA HVSI	-"-	0.14	[47]
Y-chromosome, haplogroups	North Eurasia	0.19	[28]
Y-chromosome STR	North Eurasia	0.19	[28]
Genome-wide marker sets			
600K SNP	YRI, CEU, JPT, CHB	0.12	[48]
1 million SNP	YRI, CEU, JPT, CHB	0.10	[48]
1 million	YRI, CEU, JPT, CHB	0.13	[49]
440K SNP	World	0.05	[26]
50K SNP	Asia	0.06	[25]
2.8 million	YRI, CEU, JPT, CHB	0.11	[50]
244K SNP	World	0.12	[51]
200K SNP	World	0.13	[33]
67 CNV	World	0.11	[52]

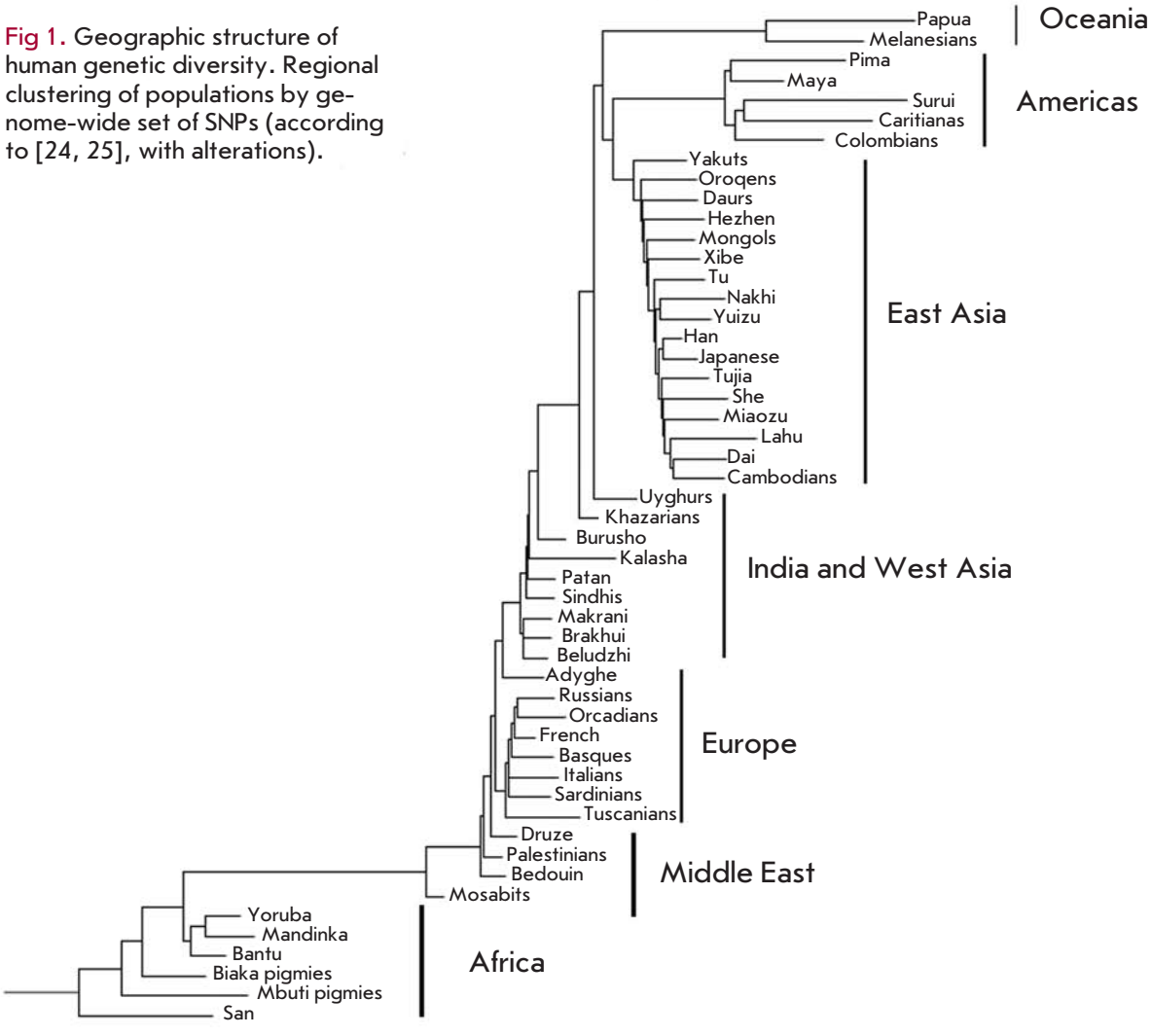


Fig 1. Geographic structure of human genetic diversity. Regional clustering of populations by genome-wide set of SNPs (according to [24, 25], with alterations).

when estimating the global level of genetic differentiation in human populations. This interval includes values obtained for most systems of genetic markers in classical and molecular population genetics of humans -- blood type, protein polymorphism, RFLP, Alu-repeats, hypervariable segments of mtDNA [28, 32]. Exceptions are the highly mutable microsatellites (STR), whose level of genetic differentiation is much lower (4–5%), and the Y-chromosome, whose variants differ (20-30%) between populations much more than other marker systems. These two types of markers are so distinctly different because of specific evolutional, population, and social mechanisms we will not discuss in this work (see [28]).

A relatively low level of genetic subdivision in human populations can be observed in the most representative and complete sets of markers – on large and random datasets of autosomal polymorphisms, including genome-wide sets of hundreds of thousands of SNPs. Li *et al.* [24] analyzed data for 650,000 SNPs in 51 populations obtained from the Human Genome Diversity Project (HGDP) and found that interpopulation differences accounted for 11% of overall genetic diversity. Recent work by us yielded another estimate of the genetic differentiation in 36 populations (32 Eurasian populations and 4 HapMap populations) for 200,000 SNPs, the result being 13.4% [33]. Somewhat smaller genetic differences were observed during the analysis of a lower number of continental groups. The level of genetic differentiation between populations in Asia is 5.9% according to data from the Panasian SNP Consortium [25], while the population differentiation in East Asia, South Asia, Europe, and Mexico is 5.2% [26].

The low level of genetic differentiation in human populations as compared to related species (chimpanzees (F_{ST} = 0.32) [34] and gorillas (F_{ST} = 0.38) [35]), despite the much larger population area, indicates that the human population originated relatively recently from a small number of ancestors.

The most general distribution pattern of human population diversity is its strict geographical struc-

ture, namely the clustering of geographically adjacent populations. On a worldwide scale, populations can be grouped into racial-continental groups for any set of markers. These groups are African Negroids, Caucasoids (which are divided into the Middle Eastern, European, and Indian sub-clusters), Asian Mongoloids, Austronesians, and American Indians [24, 25, 36] (Fig. 1). This pattern can also be observed on a smaller scale - for continental and subcontinental groups of populations [23, 37]. The projection of genetic differences between representative population datasets onto a space of major components or factors always yields a geographical map at first approximation. The cause of such a distribution is the evolutionary history of genetic diversity, which resulted mainly from migration and genetic drift during the spread of modern humans.

The population of Russia is not exempt from this pattern. Russian populations cluster into several large ethnogeographical groups: Slavs, Northern Caucasus populations, Finno-Ugric peoples of the North European and Volga-Ural regions, the populations of South Siberia and Central Asia, and the populations of Eastern Siberia and North Asia [28, 33]. The geographical structure of the Russian gene pool can be observed for all of the genetic markers – lineages of mtDNA, *Y*-chromosomes, *X*-chromosomes, and autosomal markers, including complete genomic sets of SNPs.

It is probable that the only major exception to the "geographical pattern" is the Indian subcontinent, in which the genetic diversity is better correlated to the language group, rather than the geographical origin. This is due to the complex ethnic and social structure of the population, caste hierarchy of large ethnic groups, and the presence of numerous small clan/tribe groups [38, 39].

How different are human populations in terms of disease incidence and disease gene frequency?

If we assume that the question of general interpopulation genetic diversity in humans has been answered, then the following questions arise: To what degree is genetic variation correlated with phenotypic variation, especially for clinical phenotypes (diseases)? To what degree are the differences in disease gene frequencies responsible for the interethnic and interpopulation differences in disease incidence? The first question can be answered using data on genetic epidemiology and medical statistics, while the second question needs special approaches.

Ethnic component of monogenic diseases

Genetic epidemiology has collected a large set of data on the frequency of Mendelian (monogenic) diseases in various populations. The overall load of hereditary diseases (HD) (the summed frequencies of autosomal dominant, autosomal recessive and X-linked diseases) in stable populations is relatively low and varies in a narrow range from 1.5 to 3.5 cases per 1,000 (Fig. 2). For

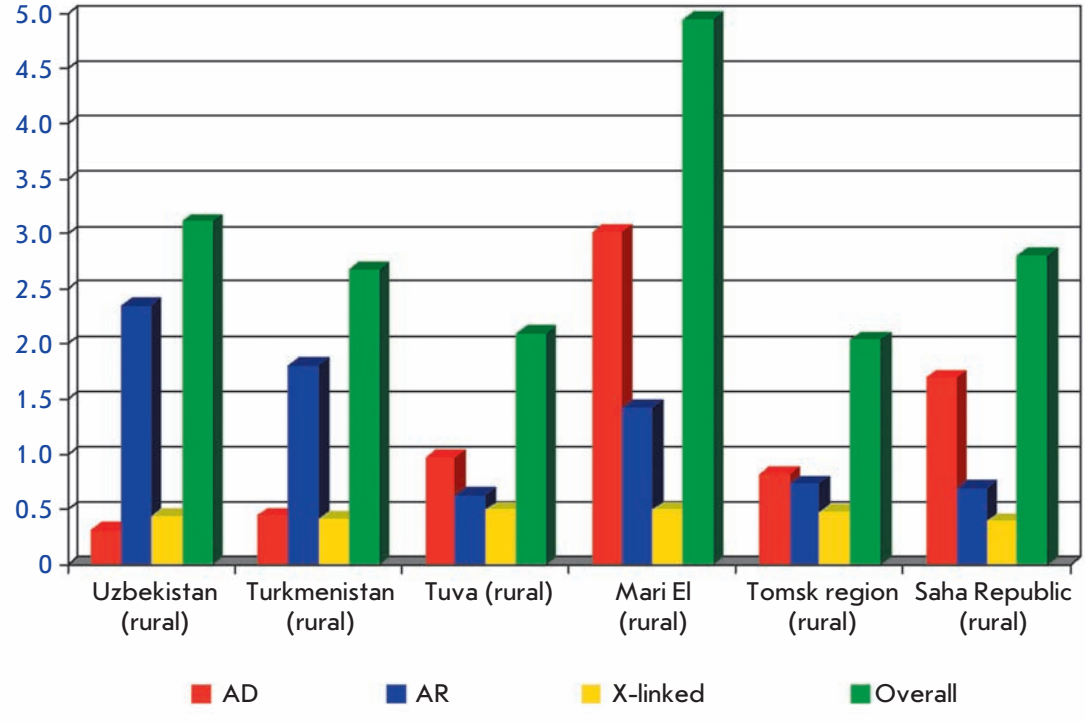


Fig. 2. Cumulative frequency of monogenic diseases in several populations in Russia and neighboring countries.

instance, 10 Russian populations including Slav, Finno-Ugric, and North Caucasus populaces vary in their summed HD load from 1.59 per 1,000 in the cities of the Kirovsk region to 3.5 per 1,000 in rural Mari populations [53]. A similar variability of HD loads is observed in the native populations of Siberia [54, 55]. However, some forms of HD can vary in much wider intervals. For instance, the incidence of cystic fibrosis can vary up to 10-fold in different regions of Siberia [56].

In the case of HD, differences in disease incidence are directly linked to differences in the allele frequencies in the population. The main factors behind population dynamics, which form the overall picture of interpopulation differences in an HD load, are genetic drift and the founder effect [53]. Drift plays a leading role even when the size of the population is stable, and its effect is compounded by rapid changes in the effective population size (population waves). Overall, the role of natural selection in the HD gene differentiation of populations is small, since mutations that lead to HD lower the fitness of individuals, irrespective of ethnicity or geographical origins.

However, there are several interesting exceptions to this rule. The most well-known is the high number of individuals that are heterozygous for sickle-cell anemia and β-thalassemia in subtropical and tropical regions, such as the Mediterranean. Regions where these erythrocyte diseases occur frequently are virtually identical to those where malaria incidence is high [57]. Heterozygous carriers of the mutant alleles have a selective advantage due to their higher level of resistance to malaria, and the high frequency of heterozygotes is supported by balancing selection.

Another common hereditary disease, cystic fibrosis, occurs frequently among Europeans and is much less common in other geographical regions. The wide spread of the main mutation (ΔF508) across all of Europe and its rarity in other parts of the world indicate that this mutation appeared a long time ago, sometime after the migration of modern humans from Africa. Direct estimates of the mutation's age using various methods suggest the opposite: namely that the mutation appeared relatively recently, about 10 thousand years ago [58, 59]. The likely reason for the wide spread of this mutation in Europe, just as for the erythrocyte diseases, is the lower susceptibility of ΔF508 heterozygotes to dehydration during typhoid and cholera, which remained a common menace in Europe until recently.

Local adaptation of populations to dietary products can also lead to differentiation for certain diseases. Thus, the high incidence of the celiac disease in Northern Europe, especially in Scandinavian countries, and low incidence in Southern Europe are likely connected to the longer history of agriculture, specifically cereal plant cultivation, in the south of Europe and the prevalence of game as the main source of food for the Scandinavian population, which would mean that there were virtually no cereals in their diet.

The role of genetic drift and the founder effect in the spread of HD can be illustrated well by the accumulation of certain forms of hereditary pathologies in some populations. Well-known examples of such populations are the Finns, Ashkenazi Jews, French Canadians and the Amish peoples. More than 20 so-called "Finnish" diseases have been documented - these are HD (mostly autosomal recessive), whose incidence among Finns is much higher than among other populations [60, 61]. The phenomenon of HD accumulation in the Finnish population is due to effective drift, long-term genetic isolation, and high occurrence of inbreeding. The same population mechanisms were probably the reason behind the accumulation of certain HD genes among the Ashkenazi Jews. They were also observed to have an extremely high incidence of more than 20 diseases (the most common among these being the Tay–Sachs disease and Type I Gaucher's disease) [62].

In Russian populations, the phenomenon of ethnospecific diseases is most often seen in Yakuts (Table 2). As many as 6 diseases can be termed "Yakut," since the frequency of these diseases is above mean global frequencies by as much as several ten-fold. These diseases include two types of dwarfism which have been documented only recently [63, 64]. The population mechanism behind this accumulation of HD in Yakut groups is the founder effect, which coincided with some of the waves of expansion and migration of the Yakut people.

Overall, some Mendelian HD can exhibit considerable interethnic frequency differences, which are due to differences in the frequency and type of mutations. These traits must, of course, be taken (and are taken) into account during medico-genetic counseling, DNA-diagnostics, and screening programs. In this context, DNA-diagnostics of HD can be considered as the first real use of personalized genomic medicine.

Genetic diversity and complex diseases. GWAS

Interpopulation comparisons of the frequencies of common multi-factor diseases (MFD), also known as complex diseases, are complicated by the absence of homogenous medical statistics for global populations and the considerable clinical and genetic heterogeneity of MFD. However, there is a considerable amount of data on the interpopulation differences in the occurrence of MFDs, such as cardio-vascular diseases [68, 69], diabetes [70], some types of cancer [71], glaucoma [72], and nephropathies [73].

The U.S. population can act as a good model for comparing the incidence of complex diseases: it is a multira-

Table 2. Ethno-specific diseases in Yakuts

Diseases (OMIM number)	World prevalence (1 per 100,000)	Prevalence in Yakuts (1 per 100,000)	References
Spinocerebellar ataxia type 1 (164400)	1.0	38.6	[65]
Myotonic dystrophy (160900)	4.0-5.0	21.3	[66]
Inherited enzymopenic methaemoglobinaemia (250800)	1.0	5.7	[67]
Oculopharingeal muscular dystrophy (164300)	1.0	11.1	[63, 64]
3M syndrome (Yakut short stature syndrome) (273750)	25 cases	12.72	[64]
Syndrome of short stature with cone dysfunction, optic atrophy, and Pelger-Huet anomaly (SCOP) (not present in OMIM)	Not described	9.95	[64]

Table 3. Mortality rates in 4 major racial groups in U.S.[1]

	Relative share in overall mortality inside the racial group				Mortality relative to White Americans[2]		
Cause of Death	Whites	African Americans	Latino Americans	Asians	African Americans	Latino Americans	Asians
Heart disease	27	26.6	25.3	25.5	0.98	0.94	0.94
Coronary heart disease	17.6	139	16.4	16.2	0.79	0.93	0.92
Stroke	52	6	5.7	8.6	1.15	1.09	1.65
Chronic obstructtive pulmonary diseases	4.9	26	2.5	2.8	0.53	0.51	0.57
Cancer	268	23.3	22.8	28.3	0.87	0.85	1.05
Pneumo-nia / Influenza	2.8	2.5	2.9	3.9	0.89	1.03	1.39
Liver diseases / Cirrhosis	1.6	1.1	3.5	0.9	0.69	2.18	0.56
Diabetes	2.7	4.2	5.5	3.3	1.55	2.03	1.22
HIV infection	0.6	3	1.8	0.3	5	3	0.5
External causes	10.4	9.9	13.4	9.2	0.95	1.29	0.88
All causes (per 100000)	450.4	690.9	332.8	264.6			

Notes.
[1]According to National Center for Health Statistics [74].
[2]Mortality rate in White Americans is taken as 1.

cial country with a highly developed healthcare system and thorough medical statistics. Table 3 shows mortality data from the National Center of Statistics in Healthcare for the four main ethno-racial groups of the U.S. population - white Americans, African Americans, Hispanics from Latin America, and Asian Americans [74]. Since the overall mortality per 100,000 varies considerably (this value is twice higher for African Americans as compared to Latino Americans and Asians, while the mortality of the white population is intermediate), we recalculated these data as fractions of the overall value for each cause of death in each ethnic group. The right-hand side of the table lists ratios between the mortalities from various causes in three minor ethnic groups and white Americans. These data allow us to conclude that the two main causes of death in the U.S., namely cardiovascular and oncological diseases, which are responsible for half of the mortality rate, do not display any significant interracial differences. Other disease groups sometimes display considerable racial differences. Thus, the relative mortality due to diabetes is about 1.5 times higher for African Americans as compared to white Americans, while the mortality due to IHD (ischemic heart disease), chronic lung, and kid-

ney diseases is much lower for African Americans. The relative mortality due to diabetes and kidney diseases among Hispanic Americans is more than twice as high compared with white Americans, while the mortality due to chronic lung diseases is twice as low. Asians die from strokes and pneumonia much more often than do white Americans; however, mortality due to COPD (Chronic Obstructive Pulmonary Disease) and kidney diseases is half as common in Asians as it is in white Americans.

To what extent can such differences be attributed to interethnic genetic differentiation? Important information on this subject can be obtained by analyzing the associations of genetic markers with complex diseases, including genome-wide association searches.

Ioannides *et al.* [75] compared the frequencies of genetic markers and their effects on diseases in the European, Asian, and African populations. They conducted a meta-analysis of 135 gene-disease associations, 45 of which proved to be statistically significant either on the level of a general meta-analysis (32 associations), or at least at the level of a single racial group (11 associations). The data of 45 meaningful meta-analyses encompassed 697 individual association studies with a combined cohort size of 300,000 individuals. The authors detected a statistically significant heterogeneity of the disease-associated allele frequencies (i.e. meaningful interpopulation differences) for 58% of the gene-disease associations. Significant differences in OR (odds ratio, a measure of the genetic risk of disease incidence) were detected only in 14% of the meta-analyses. Notably, interracial comparisons did not yield any significant associations with opposite effects in different populations.

These data indicate that the differences in susceptibility gene frequencies may be one of the causes behind the interethnic differences in MFD incidence. However, the biological effect of the associated alleles is unidirectional, irrespective of the racial/ethnic origins, even though the relative share of the marker in the disease or susceptibility can vary. This is most probably due to the genotype (haplotype) surroundings, as well as gene-gene and gene-environment interactions.

In recent years, the main source of new data concerning MFD susceptibility genes has been genome-wide association studies (GWAS). GWAS requires high-throughput analysis, which is achieved by using large cohorts (several hundreds or thousands) representative of the population, and a large number (hundreds of thousands) of tested polymorphisms, which are representative of the genomic diversity. A catalog of published GWAS is supported by the U.S. National Institute of Genomic Research and includes GWAS which were performed under very strict criteria: no fewer than 100,000 SNP must be analyzed, and the level of significance of a SNP-trait association must be no lower than 0.00001 [76]. As of the end of March 2010, the catalog contains 527 published studies and 2,516 SNP that are reliably associated with complex phenotypes.

A major part of these GWAS have been conducted on European populations, and no reliable estimations of the interethnic differentiation of disease-associated genome regions can be made using GWAS data alone. Adeyemo and Rotimi [77] managed to skirt this problem by analyzing the genetic heterogeneity of markers chosen from the GWAS catalog in populations from the HapMap project. The HapMap project (a map of human haplotypes) currently contains data on the polymorphism of several million SNP and the level of linkage disequilibrium across the whole genome for 11 populations of various ethnic origins, which are representative of the world population (Europeans, Asians, Africans, Indians, and Latin Americans). Adeyemo and Rotimi chose 621 SNPs from the GWAS catalog, which were associated with 26 complex diseases, including Alzheimer's disease, hypertension, obesity, schizophrenia, type I and type II diabetes, rheumatoid arthritis, certain types of cancer, etc. The allelic frequencies of the chosen SNPs varied across the populations in a relatively wide range - differences of up to 20 to 40-fold were observed between some pairs of population groups. The interpopulation genetic diversity ratio (F_{st}) also varied considerably from marker to marker (such as, from 0.02 to 0.2 for type II diabetes or from 0.006 to 0.52 for the level of lipids). The mean level of interpopulation difference was 10.5%; i.e., it did not significantly differ from the differentiation level observed for conditionally neutral or genome-wide datasets of markers.

These data suggest that the level of interpopulation and interethnic differences for genes associated with MFD does not differ from the general level of differentiation in the gene-pool, and that the risk of developing a disease associated with a genetic marker can vary significantly between population groups, depending on the frequencies of the associated marker and the modulating effects of other genes and environmental factors.

The role of gene-gene and gene-environment interactions is usually hard to differentiate; however, their overall effect in the modification of disease risk associated with a certain gene or marker can be very significant and has population specificity. As an example, we can examine data for the role of the ε4 allele of the *APOE* gene in Alzheimer's disease (AD). Reliable association of this marker with AD has been observed for all of the tested race groups; however, the frequency of the allele differs considerably (9% among the Japanese population, 14% among white Americans, 19% among

African Americans). Homozygozity for the ε4 allele increases the risk of AD 33-fold for Japanese, 15-fold for white Americans, and only 6-fold for African Americans. For heterozygous individuals these values are correspondingly 5.6–3–1.1.

Thus, even if the risk of complex disease development is reliably associated with distinct genetic markers in all populations and these markers have unidirectional effects, the magnitude of the effect or the severity of the risk still greatly depends on ethnicity- and population-specific factors of genetic and probably non-genetic nature. So, data on ethnic origins can provide additional information in making personalized medical prognoses.

ETHNOGENETICS AND PHARMACOGENOMICS

One of the main advantages of personalized medicine is the individualization of drug therapy. Response to a drug, choice of the optimal drug class, dosage, and usage schedules can, at least partially, be determined using genetic factors. Information on individual genetic markers can help a clinician select the appropriate drug strategy. Based on these principles, pharmacogeneticists attempt to identify the genes and gene variants that influence the efficiency of drug therapy and lower the risk of side effects. It has been shown that the most widely used drugs are only effective in 25-60% of patients, and there have been 2 million cases of side effects per year in the U.S. alone, including more than a 100,000 deaths [78].

Pharmocogenomics studies have collected considerable data on the association of genetic markers with the effectiveness of certain drugs. During the last 20 years there have been approximately 2,000 studies on this subject, and hundreds of genes associated with drug therapy efficiency have been identified [79]. Most of these pharmacogenetics studies concern cardiovascular, oncological, and neurological diseases.

The FDA (Food and Drug Administration) of the U.S. Ministry of Healthcare has approved the addition of genetic marker information into the annotations of about 30 drugs, including warfarin, abacavir, imatinib, atorvastatin, etc. [80]. The list of biomarkers which were appended into these annotations includes genes that encode cytochromes, a low density lipoprotein receptor, N-acetyltransferases, epidermal growth factor receptor, etc.. The effects of drug therapy that depend on the genotypes for these markers include the clinical response to therapy, risk of side effects, choice of optimal dosage, sensitivity or resistance towards the drug, and polymorphism of drug targets.

Most of the relevant pharmacogenetic data has been collected on Caucasoids - more than 80% of the published research has been conducted on the European and U.S. populations [80], which is why there is little information on the interethnic differences of drug efficiency and on the role of genetic factors. As an example, we can note the decreased effectiveness of enalapril, weakened effect of the vasodilator sodium nitroprusside (an antihypertnesion vasodilator), and decreased effect of propranolol and atenolol (adrenoreceptor blockers) during hypertension therapy for African Americans as compared to Caucasians [81]. In some cases, the interethnic differences can be associated with frequency differences for a specific marker. Thus, the difference in propranolol and atenolol efficiency is due to the higher frequency of one of the missense mutations of the β1-adrenoergic receptor in white Americans (72%) as compared to African Americans (57%).

We can surmise that interracial and interethnic differences in the effectiveness of drug therapy can be as common as interpopulation differences in MFD frequencies, since the genetic variability of genes that metabolize therapeutic drugs is the same as that of complex disease susceptibility genes [82, 83]. For instance, there is a 10-fold difference in the frequency of slow metabolizing variants of cytochrome CYP2D6 between Caucasians and Asians (10% for Caucasians and 1% for Japanese). This enzyme is involved in the metabolism of over 40 drugs, such as the widely used β-blockers and tricyclic antidepressants. The frequency of extremely fast metabolizing alleles of this enzyme exhibits a 10-fold difference even inside Europe - about 1-2% in Spain and up to 10% in Sweden. Our data also indicate a considerable variability within Russian populations in terms of drug metabolizing genes. For instance, the frequency of the CYP2C9*2 allele of one of the cytochrome genes is 12% for Russians, which is within the variability interval observed for Europeans (10–17%), while the allele is not present in Eastern Asian populations and occurs in the native populations of Siberia with a frequency of 1 to 6% [83]. The overall level of genetic differentiation for cytochrome genes is relatively low (F_{st} = 0.021) in Russian populations; however, it is tightly correlated with the geographical layout, as are most other marker systems (Fig.3).

FROM THE HUMAN GENOME TO THE INDIVIDUAL GENOME

Re-sequencing of complete individual genomes yields new data on the genetic variability of humans and can help create individual health prognoses in the future. The first personal genome to be sequenced was the genome of Craig Venter, one of the key figures in the study of the human genome. Venter's genome sequence was completed in October 2007 [84]. Data on the second sequenced genome (of Nobel Laureate James Watson) were published half a year later [85]. Currently (from

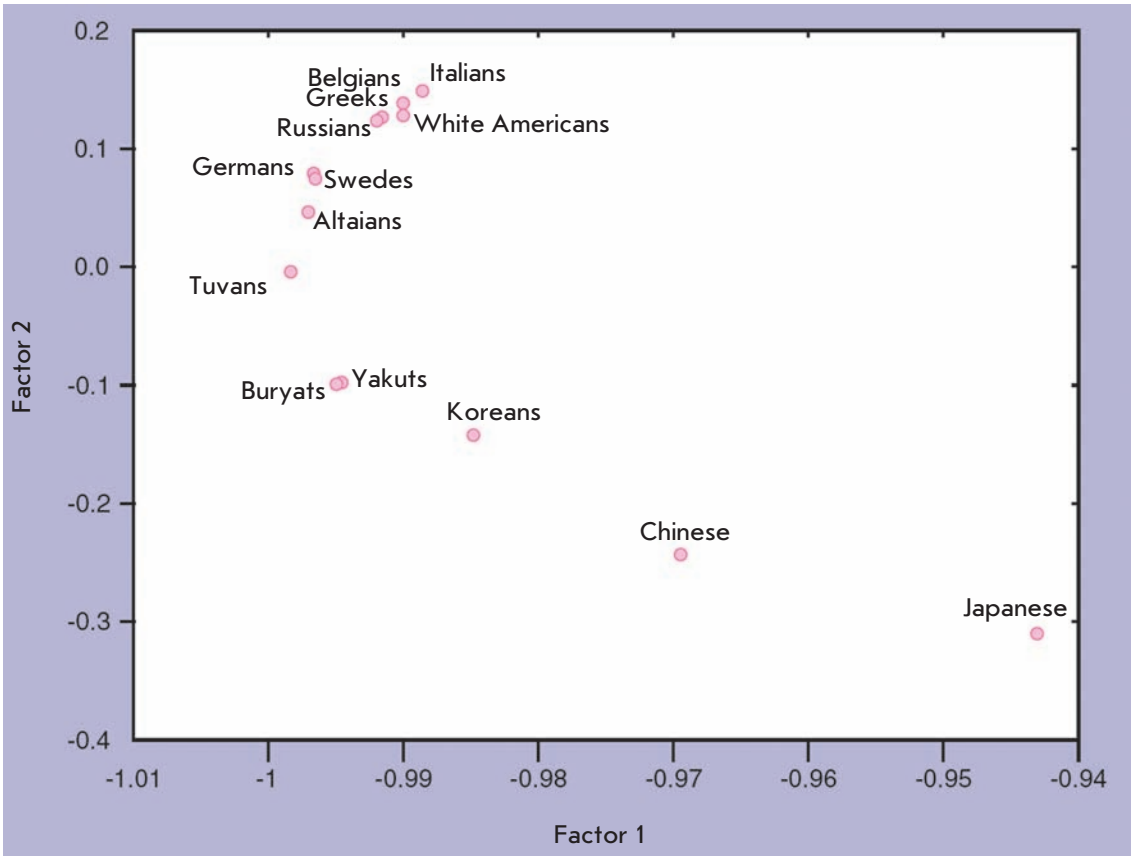


Fig. 3. Geographic pattern of Russian and worldwide populations in the space of two first principal components of CYP gene frequencies (according to [83], with alterations).

the middle of 2010), there are data on 20 re-sequenced genomes (Table 4), which include those of Craig Venter, James Watson, archbishop Desmond Tutu, the twice-sequenced genome of a Yoruba individual, a Chinese genome, two Koreans, several Europeans, an ancient Eskimo, a Russian, and an Indian [84–99].

Apart from the complete genomes of unrelated healthy people of various descents and renown, researchers have also obtained complete sequences of patients with monogenic diseases: these include four genomes of a family quartet in which the children suffer from Miller syndrome (primary ciliary diskinesia in a lung form, which is phenotypically similar to cystic fibrosis) [98], a as well as a genome of a type I Charcot–Marie–Tooth disease patient [97]. No less than 7 complete tumor cell genomes have been published for cancers of various localizations: acute myeloid leucosis, malignant melanoma, glioblastoma, etc. [86, 100–105].

Progress in sequencing technology has been colossal in terms of speed and cost reduction. The sequencing of Venter's genome, which was conducted on first generation sequencers, cost his company about 2 million U.S. dollars. Genomes which are sequenced on second-generation machines (Illumina Genome Analyzer and Applied Biosystems Solid System) cost about 200–500,000, while the most recent studies involving the re-sequencing of complete genomes cost no more than 1,500 U.S. dollars. So, Francis Colins' [106] prognosis in 2001that the cost of sequencing a genome would fall to 1,000 dollars by 2030 has come to pass 20 years ahead of time!

The "1,000 genomes" project is aimed at obtaining complete and accurate genome sequences of 2,000 individuals from different populations of the main geographical regions of the world (Africa, Europe, Asia, and the Americas) [107]. During the first phase of the project, 180 samples from four HapMap populations (CEU, YRI, CHB, and JPT) were sequenced, albeit with only a few repetitions (2-8).

What new knowledge do personal genomes add to the understanding of human genetic variability and what is their use in personalized medicine? First of all, researchers discover new genetic variants – each genome has about 2.5–4 million SNPs, several thousands of insertions/deletions and several hundreds or thousands of CNVs (Table 2). Most of these variants are being described for the first time. Re-sequencing data from complete genomes confirms the level of individual variability observed for the genomes sequenced in the "Human Genome" project - on average 3 million SNP

Table 4. Personal genomes chronology

	Date of publication (submission)	Person	Journal	Institution (country)	Platform	Cove-rage	Number of SNPs (millions)	Notes
1.	2007, October (9 May 2007)	Craig Venter (m)	PloS Biology, [83]	J Craig Venter Institute (U.S.)	ABI 3730xl	7.5	3.47	$10 mln
2.	2008, April (3 December 2007)	James Due Watson (m)	Nature [84]	Baylor College of Medicine / 454 Life Sciences (U.S.)	Roche / 454	7.4	3.32	$2 mln
3.	2008, November (28 May 2008)	Caucasian American with acute myeloid leukemia (normal fibroblasts) (f)	Nature [85]	University of Washington (U.S.)	Illumina	13.9	2.92	$1 mln
4.	2008, November (24 June 2008)	Yoruba (NA18507) (m)	Nature [86]	Illumina / University of Cambridge (UK)	Illumina	41	3.45	$250,000
5.	2008, November (21 August 2008)	Chinese (Yanhuang, YH) (m)	Nature [87]	Beijing Genomic Institute (China)	Illumina	36	3.07	$500,000
6.	2009, May (3 February 2009)	Korean Seong-Jin Kim, SJK (m)	Genome Research [89]	Gachon University of Medicine and Science (Rep. of Korea)	Illumina	29	3.44	
7.	2009, June (1 February 2009)	Yoruba (NA18507) (m)	Genome Research [90]	Life Technologies (U.S.)	ABI SOLiD	17.9	3.87	
8.	2009, August (6 March 2009)	Korean, AK1 (m)	Nature [91]	Seoul National University (Rep. of Korea)	Illumina	27.8	3.45	
9.	2009, August (10 June 2009)	Steven Quake, Caucasoid USA, P0 (m)	Nature Biotechnol [92]	Stanford University (U.S.)	Helicos SMS Heliscope	28	2.81	Sequence from 1 molecule $48,000
10.	2009, December	Russian with kidney cancer (m)	Acta Naturae, [93]	RNC Kurchatov's Institute (Russia)	Illumina/ABI SOLiD	16	?	
11.	2010, January (3 September 2009)	Caucasoid (NA07022) (m)	Science [94]	Complete Genomics (USA)	Complete Genomics DNA nanoarray	87	3.08	$8,000
12.	2010, January (3 September 2009)	Yoruba (NA19240) (f)	Science [94]	Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	63	4.04	$3,400
13.	2010, January (3 September 2009)	Caucasoid (NA20431) (f)	Science [94]	Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	45	2.91	$1,700
14.	2010, February (30 November 2009)	Paleo-Eskimo, Saqqaq (м)	Nature [95]	University of Copenhagen / Beijing Genomic Institute (Denmark / China)	Illumina	20	2.19	Ancient DNA (4000 years)
15.	2010, February (11 August 2009)	Khoisan, KB1 (m)	Nature [96]	Pennsylvania State University (U.S.)	Roche / 454 / Illumina	33.4	4.05	
16.	2010, February (11 August 2009)	Bantu, ABT Archbishop Desmond Tutu (m)	Nature [96]	Pennsylvania State University)	ABI SOLiD / Illumina	37.2	3.62	
17.	2010, March	Caucasoid, CMT1 patient (m)	New Eng J Med [97]	Baylor College of Medicine (U.S)	ABI SOLiD	29.9	3.42	
18.	2010, March (7 January 2010)	Pedigree #1, mother (Caucasoid, f)	Science [98]	Institute for System Biology / Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	51	2.87	
19.	2010, March (7 January 2010)	Pedigree #1, father (Caucasoid, m)	Science [98]	Institute for System Biology / Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	88	3.16	
20.	2010, March (7 January 2010)	Pedigree #1, daughter (Caucasoid, f, patient with Miller syndrome)	Science [98]	Institute for System Biology / Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	54	3.23	
21.	2010, March (7 January 2010)	Pedigree #1, son (Caucasoid, m, patient with Miller syndrome)	Science [98]	Institute for System Biology / Complete Genomics (U.S.)	Complete Genomics DNA nanoarray	52	3.23	
22	2010, September (10 April 2010)	Irish	Genome Biol [99]	University College Dublin (Ireland)	Illumina	11	3.12	

Table 5. Characteristics of six individual genomes

Genome	Ethnic origin, Country	Number of SNPs	Overlapping with other genomes by common (%)					
			HuRef	Watson	NA18507	YH	SJK	Saqqaq
HuRef (Venter)	White American, U.S.	3075858	100	55.8	52.9	51.6	56.4	34.2
Watson	White American, U.S.	3321942	51.6	100	50.8	49.9	54.9	36
NA18507	Yoruba, Nigeria	4189457	38.8	40.3	100	42.1	45.8	27
YH	Chinese, China	3074097	51.6	54	57.3	100	67.3	38.2
SJK	Korean, South Korea	3439107	50.5	53	55.8	60.1	100	39
Saqqaq	Paleo-Eskimo, Greenland	2193396	47.9	33.9	32.5	53.5	61.1	100

from a 3 billion base genome yield differences in 1 nucleotide in a thousand.

The genomes of two individuals overlap for about half of the SNPs (Table 5). The degree of relation between the genomes is correlated with the genetic differentiation between the populations to which these genomes belong. The Yoruba genome is the farthest removed from others (38–45% overlapping SNPs), while the most related genomes are those of a Chinese and Korean (60–67% overlapping SNPs).

"Overlaying" the variable positions in complete genomes onto the data obtained in large-scale population projects (such as HapMap) allows us to glimpse at the genetic origins of an individual. For instance, by comparing SNPs in the Venter, Watson, and YH Chinese genomes with four HapMap populations (Fig. 4), based on the marker distributions in populations CEU, YRI, CHB, and JPT, researchers can estimate the level of cross-breeding between the main racial and ethnic components observed in these genomes. Thus, the shares of Caucasoid, Negroid, and Mongoloid components in Venter's genome were estimated at 93.3, 5.1, and 1.6% respectively. The genome of the Nobel-prize winner Watson had a lower share of Caucasoid SNP (73.0%) due to the increased amount of Negroid markers (25.6%).

A personal genome sequence provides complete information on whether the individual carries any alleles associated with clinical phenotypes and is thus extremely valuable for individual health prognoses and for estimating MFD risks. The precision and relevance of the genetic health prognosis are thus unlimited by the technical possibilities of genome study but depend on the amount of knowledge on a phenome and its genetic determinants.

Data on complete genomes are insufficient at the moment in order to systematize the layout of interpopulation differences on a genome-wide scale; however, the existence of interindividual and interracial differences in the number of MFD-associated markers is obvious. Thus, Venter's genome has about 50 SNPs which are associated with alcoholism. The sequenced Yoruba genome has 30, Watson - 25, and the Mongoloids (Chinese, Korean, and ancient Eskimo) have less than 20. Venter is also a record-holder in terms of SNPs associated with tobacco addiction (about 40 SNPs). The Chinese and ancient Greenlander have about 20 of these markers, while Watson's genome, as well as the Korean and Yoruba genomes, has none.

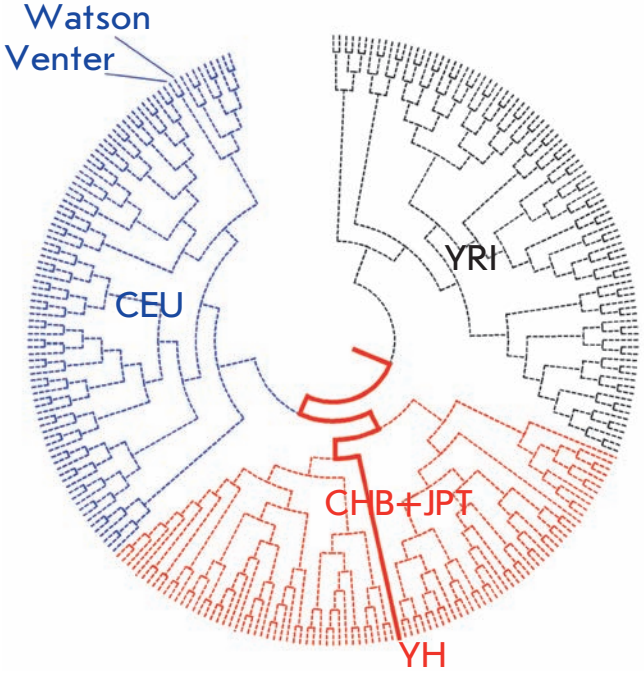


Fig 4. Individual genomes of Venter, Watson, and a Chinese individual (YH) on the tree of HapMap individuals (according to [88], with alterations).

Table 6. Individual calculation of myocardial infarction risk based on genomic data (according to [108])

Gene	SNP	Genotype	OR	Risk, % Pretest Post-test
LPA	rs3798220	CT	1.86	3.7
THBS2	rs8089	AC	1.09	4.0
LDLR	rs14158	GG	2.88	10.6
LIPC	rs11630220	AG	1.15	12.0
ESR2	rs1271572	CC	0.73	9.1
ESR2	rs35410698	GG	1.03	9.4
FXN	rs3793456	AA	0.94	8.9

The most complete approach to using genome-scale information for making individual health prognoses was demonstrated in the recent work of Ashley *et al.* [108]. In order to estimate the risk of a certain disease, they proposed the use of a so-called pretest risk as a baseline. This pretest risk is an epidemiological risk estimate, which can in its simplest form be the incidence of this disease in the appropriate ethnic and age group. The individual genome was then tested for SNPs, which are reliably associated with the disease according to GWAS data. Based on these data, the researchers calculated post-test risk, which views each marker as an independent risk factor. An example calculation of the individual risk of myocardial infarction is shown in Table 6. The risk was estimated for a healthy 40-year-old white American, whose genome was sequenced in study [92]. The DNA donor had a family history of cardio-vascular diseases and some cases of sudden death among relatives. His biochemical and electrocardiogram parameters were normal. The pretest risk was estimated at 2%. The individual's genome contained 7 SNPs which were reliably associated with myocardial infarction according to GWAS data. The OR (genetic risk estimate) values of the individual's genotypes varied from 0.75 to 2.86. The overall risk of myocardial infarction (product of the pretest risk and OR of each marker) was 8.9%. In this case, the genetic composition of the tested individual increased the overall risk by 4.5-fold as compared with the pretest risk.

CONCLUSION

Humans display a relatively low level of genetic variability (both at the level of population differentiation and at the level of individual genomes), which is set on a background of high phenotypic variability and strict geographical structure of the genetic variation, which is manifested in the clustering of geographically adjacent populations. The spatial nature of the genetic variability distribution of modern humans can be observed at different levels of population structure and in various groups of markers, including the genes associated with MFD development. Genetic differences between human populations are responsible for only 10-15% of the genetic variability in humans. However, these differences prove significant in the field of personalized medicine in terms of diagnosing monogenic diseases, estimating the susceptibility to common diseases, and making health prognoses, and the efficiency of drug therapy.

Returning to the issues stated in the introduction, let's sum up our current knowledge. Genetic differentiation in populations for disease genes is as large as the overall level of interpopulation diversity on a genome-wide scale. Observed interethnic differences in the incidence of human diseases can be almost completely (for Mendelian diseases) or to a significant degree (for MFD) explained by the different frequencies of disease-associated genes. The general "geographical pattern" of genetic diversity took shape during the spread of modern humans through migrations, genetic drift, and sudden changes in the effective size of populations. However, natural selection could have played a significant role in the development of the variability of some regions of the genome both on a global scale (such as the immune response or skin pigmentation gene) or at the level of population adaptations to the local environment (such as genes for metabolizing compounds present in the diet). The biological effects of distinct genetic variants (mutations of polymorphisms) in relation to disease are usually stable and do not depend on the racial, ethnic, or geographic context. However, the magnitude of the effects (the relative role of the marker in the disease itself or in the risk of disease development) can exhibit strong variation at the population and individual levels due to the different genetic (haplotypic) environment and modifying gene-gene and gene-environment interactions.

Undoubtedly, the "ethnic factor" must be taken into account during medical studies, especially those concerning personalized genomic medicine. The long history of discussions on this issue in medical and genetic literature continues (see [36, 109–111]). However, the professional community of researchers who work in the field of personal genomics devotes considerable attention to population aspects both at the stage of data collection (such as the GWAS or pharmacogenomic studies) and at the data-interpretation stage of individual genetic testing or genetic screening at the population level [110].

Integration of genomics and phenomics in the framework of system biology, novel powerful instruments for describing and analyzing genetic diversity - sequencing individual genomes and genome-wide analysis of SNP on microarrays, the HapMap and "1000 genomes" projects - all of these new happenings give hope as to fast progress in the categorization of genetic diversity associated with the risk of common diseases, the efficiency of drug therapy, and the establishment of a tight link between basic scientific results and proven recommendations for personalized medicine. ●

This work was partially supported by the Russian Foundation for Basic Research (grants 09-04-00143, 09-04-99028-r_ofi) and the Federal Targeted Program "Scientific and Scientific Education Staff of Innovative Russia in 2009–2013" (GC № P321 and 02.740.11.0284).

REFERENCES

1. Mudrov M.Y. Izbrannie Proizvedenia (Selected Works)/ Edited by Gukasyan A.G. M.: Izdatelstvo AMN SSSR, 1949. 296 pages.
2. Sochinenia N.I. Pirogova (Essays of N.I. Pirogova). V. 1, 2. Kiev: Izdatelstvo Pirogovskogo Tovarishestva, 1910. 682 pages.
3. Beaudet A. // Am. J. Hum. Genet. 1999. V. 64. P. 1–13.
4. Bloom B. // Nature. 1999. *V.* 402. № 6761 Suppl. P. C63–64.
5. Dausset J. // Proc. Indiannatn. Sci. Acad. 1994. V. B60. P. 449–454.
6. Hood L., Heath J.R., Phelps M.E., Lin B. // Science. 2004. V. 306. P. 640–643.
7. Weston A.D., Hood L. // J. Proteome Res. 2004. V. 3. P. 179–196.
8. Auffray C., Zhu Chen, Hood L. // Genome Medicine. 2009. V. 1. P. 2.1–2.11.
9. Puzyrev V.P., Stepanov V.A. Patalogicheskaya anatomia Cheloveka (Pathological Anatomy). Tomsk: STT, 1997. 248 c.
10. Puzyrev V.P. Volnosti Genoma i Meditsinskaya Patogenetika (Liberties of the Genome and Medical Pathogenetics). Seriya «Nasledstvennost i Zdorovye» ("Heredity and Health"). Issue 3. Tomsk: Pechatnaya manufactura, 2001. 44 pages.
11. Puzyrev V.P. Genomnaya meditsina - nastoyashee i budushee. Molecularno-biologicheskie tekhnologii v meditsniskoy praktike. (Genomic Medicine - Present and Future. Molecular-Biological Technology in Medical Practice). Issue 3. Novosibirsk: Alfa Vista, 2003. p. 3–26.
12. Baranov V.S., Aseev M.V., Baranova E.V. // Priroda (Nature). 1999. № 3. p. 17–27.
13. Baranov V.S.// Vestnik RAMN (Herald of the Russian Academy of Medical Sciences). 2000. № 10. p. 27–37.
14. Geneticheskiy pasport – osnova individualnoy i prediktivnoy meditsini (The Genetic Passport - the Basis of Individual and Predictive Medicine) / Edited by Baranov V.S. Saint-Petersburg: Izdatelstvo Nauchnoy Literaturi, 2009. 528 p.
15. Cavalli-Sforza L.L. // TIG. 1998. V. 14. № 2. P. 60–65.
16. Mourant A.E. The distribution of the human blood groups. Oxford: Blackwell Scientific, 1954.
17. Cavalli-Sforza L.L., Bodmer W.F. The genetics of human populations. San Francisco: W.H. Freeman, 1971.
18. Mourant A.E., Kopec A.C., Domaniewska-Sobczak K. The distribution of the human blood groups and other polymorphisms. London: Oxford University Press, 1976.
19. Cann R.L., Stoneking M., Wilson A.C. // Nature. 1987. V. 325. P. 31–36.
20. Underhill P.A., Shen P., Lin A.A., Jin L., Passarino G., Yang W.H., Kauffman E., Bonné-Tamir B., Bertranpetit J., Francalacci P., Ibrahim M., Jenkins T., Kidd J.R., Mehdi S.Q., Seielstad M.T., Wells R.S., Piazza A., Davis R.W., Feldman M.W., Cavalli-Sforza L.L., Oefner P.J. // Nat. Genet. 2000. V. 26. P. 358–361.
21. Jorde L.B., Bamshad M.J., Watkins W.S., Zenger R., Fraley A.E., Krakowiak P.A., Carpenter K.D., Soodyall H., Jenkins T., Rogers A.R. // Am. J. Hum. Genet. 1995. V. 57. P. 523–538.
22. Balanovskaya E.V., Richkov Yu.G. // Genetika (Genetics). 1990. V. 26. P. 114–121.
23. Novembre J., Johnson T., Bryc K., Kutalik Z., Boyko A.R., Auton A., Indap A., King K.S., Bergmann S., Nelson M.R., Stephens M., Bustamante C.D. // Nature. 2008. V. 456. P. 98–101.
24. Li J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M., Barsh G.S., Feldman M., Cavalli-Sforza L.L., Myers R.M. // Science. 2008. V. 319. P. 1100–1104.
25. HUGO Pan-Asian SNP consortium. // Science. 2009. V. 326. P. 1541–1545.
26. Auton A., Bryc K., Boyko A.R., Lohmueller K.E., Novembre J., Reynolds A., Indap A., Wright MH., Degenhardt J.D., Gutenkunst R.N., King K.S., Nelson M.R., Bustamante C.D. // Genome Res. 2009. V. 19. P. 795–803.
27. Genofond i genogeografiya narodonaseleniya (Gene Pool and Population Genogeography). V. I. Genofond naseleniya Rossii i sopredelnih stran (The Gene Pool of the Population of Russia and Adjacent Countries). Saint-Petersburg.: Nauka, 2000. 612 p.
28. Stepanov V.A. Etnogenomica naseleniya Severnoy Evrazii (Ethnogenomics of Population of Northern Eurasia). Tomsk: Pechatnaya manufaktura, 2002. 242 p.
29. Limborskaya S.A., Husnutdinova E.K., Balanovskaya E.V. Etnogenomica i genogeografiya narodov Vostochnoy Evropi (Ethnogenomics and Genogeography of Population of Eastern Europe). Moscow.: Nauka, 2002. 264 p.
30. Fedorova S.A. Geneticheskie portreti narodov Respubliki Saha (Yakutia): analiz liniy mitohondrialnoy DNK i Y-chromosomi (Genetic Portraits of the Populations of the Sakha Republic (Yakutia): Analysis of Lines of Mitochondrial DNA and Y-chromosome). Yakutsk: Izdatelstvo YaNTs SO RAN, 2008. 235 p.
31. Balanovskaya E.V., Balanovskiy O.P. Russkiy genofond na Russkoy ravnine (Russian Gene Pool of the Russian Plain). Moscow: Luch, 2007. 416 p.
32. Jorde L.B., Wooding S.P. // Nat. Genet. 2004. V. 36. P. S28–S33.
33. Stepanov V.A., Prohorchuk E.B., Husnutdinova E.K., Mazur A.M., Tesluk A., Chekanov N.N., Hrameeva E., Kharkov V.N., Gubina M.A., Kutuev I.A., Husainova R.I., Marusin A.V., Spiridonova M.G., Akhmetova V.L., Hidiyatova I.M., Litvinov S.S., Buligina E., Puzyrev V.P., Skryabin K.G. Materiali VI Rossiyskogo syezda obshestva meditsinskih genetikov (Proceedings of the VIII Congress of Russian Society of Medical Geneticists). Rostov-na-Donu. May 14–18. 2010//

Meditsinskaya genetika (Medical Genetics), 2010. P. 171.
34. The Chimpanzee Sequencingand Analysis Consortium // Nature. 2005. V. 437. P. 69–87.
35. Thalmann O., Fischer A., Lankester F., Paabo S., Vigilant L. // Mol. Biol. Evol. 2007. V. 24. P. 146–158.
36. Riesch N., Burchardt E., Ziv E., Tang H. // Genome Biology. 2002. V. 3. № 7. P. 2007.1–2007.12.
37. Tian C., Plenge R.M., Ransom M., Lee A., Villoslada P., Selmi C., Klareskog L., Pulver A.E., Qi L., Gregersen P.K., Seldin M.F. // PLoS Genetics. 2008. V. 4. № 1. e4.
38. Indian Genome Variation Consortium. // J. Genet. 2008. V. 87. № 1. P. 3–20.
39. Reich D., Thangaraj K., Patterson N., Price A.L., Singh L. // Nature. 2009. V. 461. P. 489–494.
40. Nei M., Roychoudhury A.K. Human polymorphic genes: world distribution. N.Y.: Oxford University Press, 1988.
41. Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R., Cavalli-Sforza L.L. // Nature. 1994. V. 386. P. 455–457.
42. Watkins W.S., Bamshad M., Jorde L.B. // Hum. Mol. Genet. 1995. V. 4. P. 1485–1491.
43. Jorde L.B., Rogers A.R., Bamshad M., Watkins W.S., Krakowiak P., Sung S., Kere J., Harpending H.C. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. P. 3100–3103.
44. Barbuhani G., Magagni A., Minch E., Cavalli-Sforza L.L. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. P. 4516–4519.
45. Stoneking M., Fontius J.J., Clifford S.L., Soodyall H., Arcot S.S., Saha N., Jenkins T., Tahir M.A., Deininger P.L., Batzer M.A. // Genome Res. 1997. V. 7. P. 1061–1071.
46. Watkins W.S., Rogers A.R., Ostler C.T., Wooding S., Bamshad M.J., Brassington A.M., Carroll M.L., Nguyen S.V., Walker J.A., Prasad B.V., Reddy P.G., Das P.K., Batzer M.A., Jorde L.B. // Genome Res. 2003. V. 13. P. 1607–1618.
47. Jorde L.B., Bamshad M., Rogers A.R. // BioEssays. 1998. V. 20. P. 126–136.
48. Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M., Hill W.G. // Genome Res. 2005. V. 15. P. 1468–1476.
49. International HapMap Consortium // Nature. 2005. V. 437. P. 1299–1320.
50. Barreiro L.B., Laval G., Quach H., Patin E., Quintana-Murci L. // Nat. Genet. 2008. V. 40. P. 340–345.
51. Xing J., Watkins W.S., Witherspoon D.J., Zhang Y., Guthery S.L., Thara R., Mowry B.J., Bulayeva K., Weiss R.B., Jorde L.B. // Genome Res. 2009. V. 19. P. 815–825.
52. Redon R., Ishikawa S., Fitch K.R., Feuk L., Perry G.H., Andrews T.D., Fiegler H., Shapero M.H., Carson A.R., Chen W., Cho E.K., Dallaire S., Freeman J.L., González J.R., Gratacòs M., Huang J., Kalaitzopoulos D., Komura D., MacDonald J.R., Marshall C.R., Mei R., Montgomery L., Nishimura K., Okamura K., Shen F., Somerville M.J., Tchinda J., Valsesia A., Woodwark C., Yang F., Zhang J., Zerjal T., Zhang J., Armengol L., Conrad D.F., Estivill X., Tyler-Smith C., Carter N.P., Aburatani H., Lee C., Jones K.W., Scherer S.W., Hurles M.E. // Nature. 2006. V. 444. P. 444–454.
53. Geneticheskaya struktura i nasledstvennie bolezni chuvashskoy populyatsii (Genetic Structure and Genetic Disease of the Chuvash Population) / Edited by Ginter E.K. Cheboksari: Izdatelskiy Dom “Pegas”, 2006. 232 p.
54. Puzyrev V.P., Nazarenko L.P. Genetiko-epidemiologicheskoe issledovanie nasledstvennoy patologii v Zapadnoi Sibiri (Genetic-Epidemiological Study of Hereditary Diseases in Western Siberia). Tomsk: STT, 2000. 192 p.
55. Salyukova O.A., Minaycheva L.I., Nazarenko L.P. // Genetika cheloveka i patologiya (Human Genetics and Pathology). 2007. I. 8. P. 272–274.
56. Odinokova O.N., Nazarenko L.P.// Genetika cheloveka i patologiya (Human Genetics and Pathology). 2007. I. 8. P. 173–178.
57. Jobling M.A., Hurles M.E., Tyler-Smith C. Human evolutionary genetics. Origins, peoples and diseases. N.Y.: Garland Publ., 2004. 524 p.
58. Slatkin M., Bertorelle G. // Genetics. 2001. V. 158. P. 449–454.
59. Wiuf C. // Genet. Res. 2001. V. 78. P. 41–47.
60. de la Chapelle A. // Med. Genet. 1993. V. 30. P. 857–865.
61. Peltonen L., Pekkarinen P., Aaltonen J. // Biol. Chem. Hoppe-Zeyler. 1995. V. 376. P. 697–704.
62. Kedar-Barnes I., Rosen P. // Familial Cancer. 2004. V. 3. P. 193–199.
63. Maksimova N.R., Nikolaeva I.A., Korotov M.N., Ikeuchi T., Onodera O., Nishizava M., Stepanova S.K., Kurtanov H.A., Suhomyasova A.L., Nogovitsina A.N., Gurinova E.E., Stepanov V.A., Puzyrev V.P. // Zhurnal nevrologii i psikhiatrii im. S.S. Koraskova (Journal of Neurology and Psychiatry named after S.S. Korsakov). 2008. V. 108. № 6. P. 52–60.
64. Maksimova N.R. Kliniko-genealogicheskaya i molekularno-geneticheskaya kharakteristika etnospecificheskih form nasledstvennoy patologii u yakutov (Clinical and Genealogical and Molecular Genetic Characteristics of Ethno-specific Forms of Hereditary Pathology in the Yakut). Dissertation by a doctor of medical sciences. Tomsk, 2009. 436 p.
65. Platonov F.A., Illarioshkin S.N., Kononova S.K., Gogolev M.P., Ivano-Smolenskaya I.A. // Med. Genetika (Medical Genetics). 2004. V. 5. P. 242–248.
66. Suhomyasova A.L. Autosomno-dominantnaya miotonicheskaya distrofiya v Respublike Saha (Yakutia) (Autosomal Dominant Myotonic Dystrophy in the Republic of Sakha (Yakutia)). Abstract of a dissertation of a Ph.D. in medical sciences. Tomsk, 2005. 22 p.
67. Banshikova E.S. Osobennosti klinicheskogo techeniya i morfofunktsionalnoe sostoyanie eritrocitov u detey s nasledstvennoy enzimopenicheskoy metgemoglobinemiey (Clinical Features and Morphological Functional State of Erythrocytes in Children with Hereditary Enzymopenic Methemoglobinemia). Abstract of a dissertation of a Ph.D. in medical sciences. Tomsk, 2002. 25 p.
68. Voevoda M.I., Stepanov V.A., Romaschenko A.G., Maximov V.N. // Bulleten SO RAMN. 2006. № 2. P. 63–72.
69. Wassel C.L., Pankow J.S., Peralta C.A., Choudhry S., Seldin M.F., Arnett D.K. // Cardiovasc. Genet. 2009. V. 2. P. 629–636.
70. Karter A.J., Ferrara A., Liu J.Y., Moffet H.H., Ackerson L.M., Selby J.V. // JAMA. 2002. V. 287. P. 2519–2527.
71. Shibata A., Whittemore A.S. // Prostate. 1997. V. 32. P. 65–72.
72. Tielsch J.M., Sommer A., Katz J., Royall R.M., Quigley H.A., Javitt J. // JAMA. 1991. V. 266. P. 369–374.
73. Tarver-Carr M.E., Powe N.R., Eberhardt M.S., LaVeist T.A., Kington R.S., Coresh J., Brancati F.L. // Jam. Soc. Nephrol. 2002. V. 13. P. 2363–2370.
74. MacKay A.P., Fingerhut L.A., Duran C.R. Health, United States. With adolescent health chartbook. Hayatsville, Maryland: National Center for Health Statistics, 2000. 456 p.
75. Ioannides J.P., Ntzani E.E., Trikalinos T.A. // Nat. Genet. 2004. V. 36. P. 1312–1318.
76. Hindorff L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S., Manolio T.A. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. P. 9362–9367.
77. Adeyemo A., Rotimi C. // Public Health Genomics. 2010. V. 12. P. 72—79.

78. Wilkinson G.R. // N. Eng. J. Med. 2005. V. 352. P. 2211–2221.
79. FDA (Food and Drug Administration), 2010. Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm
80. Holmes M.V., Shah T., Vickery C., Smeeth L., Hingorani A.D., Casas J.P. // PloS ONE. 2009. V. 4. № 12. e7960.
81. Tate S.K., Goldstein D.B. // Nat. Genet. 2004. V. 36. P. S21–S27.
82. Wilson J.E., Weale M., Smith A.C. // Nat. Genet. 2001. V. 29. P. 265–269.
83. Makeeva O.A., Stepanov V.A., Puzyrev V.P., Grossman A., Goldstein D.B. // Pharmacogenetics. 2008. V. 9. P. 847–868.
84. Levy S., Sutton G., Ng P.C., Feuk L., Halpern A.L., Walenz B.P., Axelrod N., Huang J., Kirkness E.F., Denisov G., Lin Y., MacDonald J.R., Pang A.W., Shago M., Stockwell T.B., Tsiamouri A., Bafna V., Bansal V., Kravitz S.A., Busam D.A., Beeson K.Y., McIntosh T.C., Remington K.A., Abril J.F., Gill J., Borman J., Rogers Y.H., Frazier M.E., Scherer S.W., Strausberg R.L., Venter J.C. // Plos Biol. 2007. V. 5. № 10. e254.
85. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X.Z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. // Nature. 2008. V. 452. P. 872–877.
86. Ley T.J., Mardis E.R., Ding L., Fulton B., McLellan M.D., Chen K., Dooling D., Dunford-Shore B.H., McGrath S., Hickenbotham M., Cook L., Abbott R., Larson D.E., Koboldt D.C., Pohl C., Smith S., Hawkins A., Abbott S., Locke D., Hillier L.W., Miner T., Fulton L., Magrini V., Wylie T., Glasscock J., Conyers J., Sander N., Shi X., Osborne J.R., Minx P., Gordon D., Chinwalla A., Zhao Y., Ries R.E., Payton J.E., Westervelt P., Tomasson M.H., Watson M., Baty J., Ivanovich J., Heath S., Shannon W.D., Nagarajan R., Walter M.J., Link D.C., Graubert T.A., DiPersio J.F., Wilson R.K. // Nature. 2008. V. 456. P. 66–72.
87. Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell H.R., Boutell J.M., Bryant J., Carter R.J., Keira Cheetham R., Cox A.J., Ellis D.J., Flatbush M.R., Gormley N.A., Humphray S.J., Irving L.J., Karbelashvili M.S., Kirk S.M., Li H., Liu X., Maisinger K.S., Murray L.J., Obradovic B., Ost T., Parkinson M.L., Pratt M.R., Rasolonjatovo I.M., Reed M.T., Rigatti R., Rodighiero C., Ross M.T., Sabot A., Sankar S.V., Scally A., Schroth G.P., Smith M.E., Smith V.P., Spiridou A., Torrance P.E., Tzonev S.S., Vermaas E.H., Walter K., Wu X., Zhang L., Alam M.D., Anastasi C., Aniebo I.C., Bailey D.M., Bancarz I.R., Banerjee S., Barbour S.G., Baybayan P.A., Benoit V.A., Benson K.F., Bevis C., Black P.J., Boodhun A., Brennan J.S., Bridgham J.A., Brown R.C., Brown A.A., Buermann D.H., Bundu A.A., Burrows J.C., Carter N.P., Castillo N., Chiara E. Catenazzi M., Chang S., Neil Cooley R., Crake N.R., Dada O.O., Diakoumakos K.D., Dominguez-Fernandez B., Earnshaw D.J., Egbujor U.C., Elmore D.W., Etchin S.S., Ewan M.R., Fedurco M., Fraser L.J., Fuentes Fajardo K.V., Scott Furey W., George D., Gietzen K.J., Goddard C.P., Golda G.S., Granieri P.A., Green D.E., Gustafson D.L., Hansen N.F., Harnish K., Haudenschild C.D., Heyer N.I., Hims M.M., Ho J.T., Horgan A.M., Hoschler K., Hurwitz S., Ivanov D.V., Johnson M.Q., James T., Huw Jones T.A., Kang G.D., Kerelska T.H., Kersey A.D., Khrebtukova I., Kindwall A.P., Kingsbury Z., Kokko-Gonzales P.I., Kumar A., Laurent M.A., Lawley C.T., Lee S.E., Lee X., Liao A.K., Loch J.A., Lok M., Luo S., Mammen R.M., Martin J.W., McCauley P.G., McNitt P., Mehta P., Moon K.W., Mullens J.W., Newington T., Ning Z., Ling Ng B., Novo S.M., O'Neill M.J., Osborne M.A., Osnowski A., Ostadan O., Paraschos L.L., Pickering L., Pike A.C., Pike A.C., Chris Pinkard D., Pliskin D.P., Podhasky J., Quijano V.J., Raczy C., Rae V.H., Rawlings S.R., Chiva Rodriguez A., Roe P.M., Rogers J., Rogert Bacigalupo M.C., Romanov N., Romieu A., Roth R.K., Rourke N.J., Ruediger S.T., Rusman E., Sanches-Kuiper R.M., Schenker M.R., Seoane J.M., Shaw R.J., Shiver M.K., Short S.W., Sizto N.L, Sluis J.P., Smith M.A., Ernest Sohna Sohna J., Spence E.J., Stevens K., Sutton N., Szajkowski L., Tregidgo C.L., Turcatti G., Vandevondele S., Verhovsky Y., Virk S.M., Wakelin S., Walcott G.C., Wang J., Worsley G.J., Yan J., Yau L., Zuerlein M., Rogers J., Mullikin J.C., Hurles M.E., McCooke N.J., West J.S., Oaks F.L., Lundberg P.L., Klenerman D., Durbin R., Smith A.J. // Nature. 2008. V. 456. P. 53–59.
88. Wang J., Wang W., Li R., Li Y., Tian G., Goodman L., Fan W., Zhang J., Li J., Zhang J., Guo Y., Feng B., Li H., Lu Y., Fang X., Liang H., Du Z., Li D., Zhao Y., Hu Y., Yang Z., Zheng H., Hellmann I., Inouye M., Pool J., Yi X., Zhao J., Duan J., Zhou Y., Qin J., Ma L., Li G., Yang Z., Zhang G., Yang B., Yu C., Liang F., Li W., Li S., Li D., Ni P., Ruan J., Li Q., Zhu H., Liu D., Lu Z., Li N., Guo G., Zhang J., Ye J., Fang L., Hao Q., Chen Q., Liang Y., Su Y., San A., Ping C., Yang S., Chen F., Li L., Zhou K., Zheng H., Ren Y., Yang L., Gao Y., Yang G., Li Z., Feng X., Kristiansen K., Wong G.K., Nielsen R., Durbin R., Bolund L., Zhang X., Li S., Yang H., Wang J. // Nature. 2008. V. 456. P. 60–65.
89. Ahn S.M., Kim T.H., Lee S,. Kim D., Ghang H., Kim D.S., Kim B.C., Kim S.Y., Kim W.Y., Kim C., Park D., Lee Y.S., Kim S., Reja R., Jho S., Kim C.G., Cha J.Y., Kim K.H., Lee B., Bhak J., Kim S.J. // Genome Res. 2009. V. 19. № 9. P. 1622–1629.
90. McKernan K.J., Peckham H.E., Costa G.L., McLaughlin S.F., Fu Y., Tsung E.F., Clouser C.R., Duncan C., Ichikawa J.K., Lee C.C., Zhang Z., Ranade S.S., Dimalanta E.T., Hyland F.C., Sokolsky T.D., Zhang L., Sheridan A., Fu H., Hendrickson C.L., Li B., Kotler L., Stuart J.R., Malek J.A., Manning J.M., Antipova A.A., Perez D.S., Moore M.P., Hayashibara K.C., Lyons M.R., Beaudoin R.E., Coleman B.E., Laptewicz M.W., Sannicandro A.E., Rhodes M.D., Gottimukkala R.K., Yang S., Bafna V., Bashir A., MacBride A., Alkan C., Kidd J.M., Eichler E.E., Reese M.G., De La Vega F.M., Blanchard A.P. // Genome Res. 2009. V. 19. P. 1527–1541.
91. Kim J.I., Ju Y.S., Park H., Kim S., Lee S., Yi J.H., Mudge J., Miller N.A., Hong D., Bell C.J., Kim H.S., Chung I.S., Lee W.C., Lee J.S., Seo S.H., Yun J.Y., Woo H.N., Lee H., Suh D., Lee S., Kim H.J., Yavartanoo M., Kwak M., Zheng Y., Lee M.K., Park H., Kim J.Y., Gokcumen O., Mills R.E., Zaranek A.W., Thakuria J., Wu X., Kim R.W., Huntley J.J., Luo S., Schroth G.P., Wu T.D., Kim H., Yang K.S., Park W.Y., Kim H., Church G.M., Lee C., Kingsmore S.F., Seo J.S. // Nature. 2009. V. 460. P. 1011–1015.
92. Pushkarev D., Neff N., Quake S.R. // Nat. Biotechnol. 2009. V. 27. P. 847–850.
93. Skryabin K.G., Prokhorchuk E.B., Mazur A.M., Buligigna E.S., Tsigankova S.V., Nedoluzhko A.V., Rastorguev S.M., Matveev V.B., Chekanov N.N., Goranskaya D.A., Tesluk A.B., Gruzdeva N.M., Velihov V.E., Zardize D.G., Kavalchuk M.V. // Acta Naturae. 2009. V. 1. № 3. P. 113–119.
94. Drmanac R., Sparks A.B., Callow M.J., Halpern A.L., Burns N.L., Kermani B.G., Carnevali P., Nazarenko I., Nilsen G.B., Yeung G., Dahl F., Fernandez A., Staker B., Pant K.P., Baccash J., Borcherding A.P., Brownley A., Cedeno R., Chen L., Chernikoff D., Cheung A., Chirita R., Curson B., Ebert J.C., Hacker C.R., Hartlage R., Hauser B., Huang S., Jiang Y.,

Karpinchyk V., Koenig M., Kong C., Landers T., Le C., Liu J., McBride C.E., Morenzoni M., Morey R.E., Mutch K., Perazich H., Perry K., Peters B.A., Peterson J., Pethiyagoda C.L., Pothuraju K., Richter C., Rosenbaum A.M., Roy S., Shafto J., Sharanhovich U., Shannon K.W., Sheppy C.G., Sun M., Thakuria J.V., Tran A., Vu D., Zaranek A.W., Wu X., Drmanac S., Oliphant A.R., Banyai W.C., Martin B., Ballinger D.G., Church G.M., Reid C.A. // Science. 2010. V. 327. P. 78–81.
95. Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke I., Metspalu M., Metspalu E., Kivisild T., Gupta R., Bertalan M., Nielsen K., Gilbert M.T., Wang Y., Raghavan M., Campos P.F., Kamp H.M., Wilson A.S., Gledhill A., Tridico S., Bunce M., Lorenzen E.D., Binladen J., Guo X., Zhao J., Zhang X., Zhang H., Li Z., Chen M., Orlando L., Kristiansen K., Bak M., Tommerup N., Bendixen C., Pierre T.L., Grønnow B., Meldgaard M., Andreasen C., Fedorova S.A., Osipova L.P., Higham T.F., Ramsey C.B., Hansen T.V., Nielsen F.C., Crawford M.H., Brunak S., Sicheritz-Pontén T., Villems R., Nielsen R., Krogh A., Wang J., Willerslev E. // Nature. 2010. V. 463. P. 757–762.
96. Schuster S.C., Miller W., Ratan A., Tomsho L.P, Giardine B., Kasson L.R., Harris R.S., Petersen D.C., Zhao F., Qi J., Alkan C., Kidd J.M., Sun Y., Drautz D.I., Bouffard P., Muzny D.M., Reid J.G., Nazareth L.V., Wang Q., Burhans R., Riemer C., Wittekindt N.E., Moorjani P., Tindall E.A., Danko C.G., Teo W.S., Buboltz A.M., Zhang Z., Ma Q., Oosthuysen A., Steenkamp A.W., Oostuisen H., Venter P., Gajewski J., Zhang Y., Pugh B.F., Makova K.D., Nekrutenko A., Mardis E.R., Patterson N., Pringle T.H., Chiaromonte F., Mullikin J.C., Eichler E.E., Hardison R.C., Gibbs R.A., Harkins T.T., Hayes V.M. // Nature. 2010. V. 463. P. 943–947.
97. Lupski J.R., Reid J.G., Gonzaga-Jauregui C., Rio Deiros D., Chen D.C., Nazareth L., Bainbridge M., Dinh H., Jing C., Wheeler D.A., McGuire A.L., Zhang F., Stankiewicz P., Halperin J.J., Yang C., Gehman C., Guo D., Irikat R.K., Tom W., Fantin N.J., Muzny D.M., Gibbs R.A. // N. Eng. J. Med. 2010. V. 362. P. 1181–1191.
98. Roach J.C., Glusman G., Smit A.F., Huff C.D., Hubley R., Shannon P.T., Rowen L., Pant K.P., Goodman N., Bamshad M., Shendure J., Drmanac R., Jorde L.B., Hood L., Galas D.J. // Science. 2010. V. 328. P. 636–639.
99. Pin Tong, Prendergast J.G.D., Lohan A.J., Farrington S.M., Cronin S., Friel N., Bradley D.G., Hardiman O., Evans A., Wilson J.F., Loftus B. // Genome Biology. 2010. V. 11. R91.
100. Mardis E.R., Ding L., Dooling D.J., Larson D.E., McLellan M.D., Chen K., Koboldt D.C., Fulton R.S., Delehaunty K.D., McGrath S.D., Fulton L.A., Locke D.P., Magrini V.J., Abbott R.M., Vickery T.L., Reed J.S., Robinson J.S., Wylie T., Smith S.M., Carmichael L., Eldred J.M., Harris C.C., Walker J., Peck J.B., Du F., Dukes A.F., Sanderson G.E., Brummett A.M., Clark E., McMichael J.F., Meyer R.J., Schindler J.K., Pohl C.S., Wallis J.W., Shi X., Lin L., Schmidt H., Tang Y., Haipek C., Wiechert M.E., Ivy J.V., Kalicki J., Elliott G., Ries R.E., Payton J.E., Westervelt P., Tomasson M.H., Watson M.A., Baty J., Heath S., Shannon W.D., Nagarajan R., Link D.C., Walter M.J., Graubert T.A., DiPersio J.F., Wilson R.K., Ley T.J. // N. Engl. J. Med. 2009. V. 361. P. 1058–1066.
101. Pleasance E.D., Cheetham R.K., Stephens P.J., McBride D.J., Humphray S.J., Greenman C.D., Varela I., Lin M.L., Ordóñez G.R., Bignell G.R., Ye K., Alipaz J., Bauer M.J., Beare D., Butler A., Carter R.J., Chen L., Cox A.J., Edkins S., Kokko-Gonzales P.I., Gormley N.A., Grocock R.J., Haudenschild C.D., Hims M.M., James T., Jia M., Kingsbury Z., Leroy C., Marshall J., Menzies A., Mudie L.J., Ning Z., Royce T., Schulz-Trieglaff O.B., Spiridou A., Stebbings L.A., Szajkowski L., Teague J., Williamson D., Chin L., Ross M.T., Campbell P.J., Bentley D.R., Futreal P.A., Stratton M.R. // Nature. 2010. V. 463. P. 191–196.
102. Pleasance E.D., Stephens P.J., O'Meara S., McBride D.J., Meynert A., Jones D., Lin M.L., Beare D., Lau K.W., Greenman C., Varela I., Nik-Zainal S., Davies H.R., Ordoñez G.R., Mudie L.J., Latimer C., Edkins S., Stebbings L., Chen L., Jia M., Leroy C., Marshall J., Menzies A., Butler A., Teague J.W., Mangion J., Sun Y.A., McLaughlin S.F., Peckham H.E., Tsung E.F., Costa G.L., Lee C.C., Minna J.D., Gazdar A., Birney E., Rhodes M.D., McKernan K.J., Stratton M.R., Futreal P.A., Campbell P.J. // Nature. 2010. V. 463. P. 184–190.
103. Shah S.P., Morin R.D., Khattra J., Prentice L., Pugh T., Burleigh A., Delaney A., Gelmon K., Guliany R., Senz J., Steidl C., Holt R.A., Jones S., Sun M., Leung G., Moore R., Severson T., Taylor G.A., Teschendorff A.E., Tse K., Turashvili G., Varhol R., Warren R.L., Watson P., Zhao Y., Caldas C., Huntsman D., Hirst M., Marra M.A., Aparicio S. // Nature. 2009. V. 461. P. 809–813.
104. Ding L., Ellis M.J., Li S., Larson D.E., Chen K., Wallis J.W., Harris C.C., McLellan M.D., Fulton R.S., Fulton L.L., Abbott R.M., Hoog J., Dooling D.J., Koboldt D.C., Schmidt H., Kalicki J., Zhang Q., Chen L., Lin L., Wendl M.C., McMichael J.F., Magrini V.J., Cook L., McGrath S.D., Vickery T.L., Appelbaum E., Deschryver K., Davies S., Guintoli T., Lin L., Crowder R., Tao Y., Snider J.E., Smith S.M., Dukes A.F., Sanderson G.E., Pohl C.S., Delehaunty K.D., Fronick C.C., Pape K.A., Reed J.S., Robinson J.S., Hodges J.S., Schierding W., Dees N.D., Shen D., Locke D.P., Wiechert M.E., Eldred J.M., Peck J.B., Oberkfell B.J., Lolofie J.T., Du F., Hawkins A.E., O'Laughlin M.D., Bernard K.E., Cunningham M., Elliott G., Mason M.D., Thompson D.M., Jr., Ivanovich J.L., Goodfellow P.J., Perou C.M., Weinstock G.M., Aft R., Watson M., Ley T.J., Wilson R.K., Mardis E.R. // Nature. 2010. V. 464. P. 999–1005.
105. Clark M.J., Homer N., O'Connor B.D., Chen Z., Eskin A., Lee H., Merriman B., Nelson S.F. // PLoS Genet. 2010. V. 6. e1000832.
106. International Human Genome Sequencing Consortium // Nature. 2001. V. 409. P. 860–921.
107. 1000 Genomes Meeting Report: A workshop to plan a deep catalog of human genetic variation. http://www.1000genomes.org/page.php
108. Ashley E.A., Butte A.J., Wheeler M.T., Chen R., Klein T.E., Dewey F.E., Dudley J.T., Ormond K.E., Pavlovic A., Morgan A.A., Pushkarev D., Neff N.F., Hudgins L., Gong L., Hodges L.M., Berlin D.S., Thorn C.F., Sangkuhl K., Hebert J.M., Woon M., Sagreiya H., Whaley R., Knowles J.W., Chou M.F., Thakuria J.V., Rosenbaum A.M., Zaranek A.W., Church G.M., Greely H.T., Quake S.R., Altman R.B. // Lancet. 2010. V. 375. P. 1525–1535.
109. Cooper R.S. // Int. J. Epidemiol. 2002. V. 32. P. 23–25.
110. Khoury M.J., McBride C.M., Schully S.D., Ioannidis J.P., Feero W.G., Janssens A.C., Gwinn M., Simons-Morton D.G., Bernhardt J.M., Cargill M., Chanock S.J., Church G.M., Coates R.J., Collins F.S., Croyle R.T., Davis B.R., Downing G.J., Duross A., Friedman S., Gail M.H., Ginsburg G.S., Green R.C., Greene M.H., Greenland P., Gulcher J.R., Hsu A., Hudson K.L., Kardia S.L., Kimmel P.L., Lauer M.S., Miller A.M., Offit K., Ransohoff D.F., Roberts J.S., Rasooly R.S., Stefansson K., Terry S.F., Teutsch S.M., Trepanier A., Wanke K.L., Witte J.S., Xu J., Centers for Disease Control and Prevention // Genetics in Medicine. 2009. V. 11. P. 559–567.
111. Genetics of the human race // Nat. Genet. 2004. V. 36. № 11. Suppl. Special Issue.

Hereditary Breast-Ovarian Cancer Syndrome in Russia

A. P. Sokolenko[1,2], A. G. Iyevleva[1,2], N. V. Mitiushkina[1], E. N. Suspitsin[1,2], E.`V. Preobrazhenskaya[1], E. Sh. Kuligina[1], D. A. Voskresenskiy[2], O. S. Lobeiko[1], N. Yu. Krylova[1], T. V. Gorodnova[1], K. G. Buslov[2], E. M. Bit-Sava[1], G. D. Dolmatov[4], N. V. Porhanova[5], I. S. Polyakov[5], S. N. Abysheva[1], A. S. Katanugina[1], D. V. Baholdin[1], G. A. Yanus[1,2], A. V. Togo[1], V. M. Moiseyenko[1,3], S. Ya. Maximov[1], V. F. Semiglazov[1], E. N. Imyanitov[1,2,3]*

[1]Petrov Institute of Oncology, St. Petersburg
[2]State Pediatric Medical Academy, St. Petersburg
[3]Medical Academy for Postgraduate Studies, St. Petersburg
[4]City Oncological Hospital, St. Petersburg
[5]Kuban State Medical University, Krasnodar
*E-mail: evgeny@imyanitov.spb.ru
Received 20.09.2010

ABSTRACT **Hereditary breast-ovarian cancer syndrome contributes to as much as 5–7% of breast cancer (BC) and 10–15% of ovarian cancer (OC) incidence. Mutations in the "canonical" genes *BRCA1* and *BRCA2* occur in 20–30% of affected pedigrees. In addition to *BRCA1* and *BRCA2* mutations, germ-line lesions in the *CHEK2*, *NBS1*, and *PALB2* genes also contribute to familial BC clustering. The epidemiology of hereditary breast-ovarian cancer in Russia has some specific features. The impact of the "founder" effect is surprisingly remarkable: a single mutation, *BRCA1* 5382insC, accounts for the vast majority of *BRCA1* defects across the country. In addition, there are two other recurrent *BRCA1* alleles: *BRCA1* 4153delA and *BRCA1* 185delAG. Besides *BRCA1*, in Russia breast cancer is often caused by germ-line alterations in the *CHEK2* and *NBS1* genes. In contrast to *BRCA1* and *BRCA2*, the *CHEK2* and *NBS1* heterozygosity does not significantly increase the OC risk. Several Russian breast cancer clinics recently started to investigate the efficacy of cisplatin in the therapy of *BRCA1*-related cancers; initial results show a unique sensitivity of *BRCA1*-associated tumours to this compound.**
KEYWORDS **breast cancer, ovarian cancer, hereditary cancer syndromes, *BRCA1*, *CHEK2*, NBS.**

INTRODUCTION

Breast cancer (BC) and ovarian cancer (OC) contribute significantly to cancer incidence and mortality. BC is the most frequent malignant pathology in women, with the lifetime risk reaching approximately 10%. In some cases, BC can be easily diagnosed at early stages and ultimately cured. Unfortunately, even with the implementation of total population screening, the BC related mortality rate has not decreased significantly, due to insufficient sensitivity of available diagnostic methods, as well as the high metastatic potential of some BC forms [1]. OC is a much rarer disease than BC, being found only in 1.5% of women around the world; however, it is almost always diagnosed at late (incurable) stages. Early ovarian tumours do not cause symptoms and are often missed by ultrasound examination and CA-125 biomarker assay [2]. Both BC and OC are diseases of the reproductive system; therefore, their hormonal, metabolic, and behavioural risk factors are common to a certain extent. Interestingly, these two diseases are the main components of the most frequent genetic disease – hereditary breast-ovarian cancer (HBOC) syndrome [3].

HBOC has been intensively studied since the early 1990s. In 1994, the first gene associated with this syndrome was discovered and named *BRCA1* (BReast CAncer 1) [4], and the second gene, *BRCA2*, was discovered a year later [5]. Although *BRCA1* and *BRCA2* code for different proteins, their products play a key role in preserving genome integrity by participating in DNA repair [6]. Notably, *BRCA1* or *BRCA2* mutations occur only in 20–30% of familial BC/OC cases. There has been an active search for other hereditary BC/OC genes. The effort has already helped to identify several new relevant genes, e.g. *CHEK2*, *NBS1*, *PALB2* etc. [7].

The first studies on the contribution of the *BRCA1* and *BRCA2* genes in BC and OC incidence were performed on European and North American women. The mutations in these genes are very diverse [8], which complicates *BRCA* diagnostics. Indeed, to perform the

complete analysis of *BRCA1* and *BRCA2*, one needs not only to perform full sequencing of these long genes, but also to find rearrangements using the MPLA method (multiplex ligation-dependent probe amplification). In the mid-1990s, it was established that the so-called "founder effect" was present in some small isolated ethnic groups. For example, in females Ashkenazi Jew nearly all *BRCA1* and *BRCA2* mutations are represented by only 3 recurrent alleles, i.e. *BRCA1* 185delAG, *BRCA1* 5382insC, *BRCA2* 6174delT; *BRCA2* 999del5 is a prevailing mutation in Icelandic females [9, 10]. Population-specific distribution of hereditary cancer mutations may significantly affect the design of genetic studies. In countries without a strong founder effect, only cancer cases with a high probability of detecting the mutation are usually taken into the analysis; they include oncological patients with a proven cancer history in their family and/or patients with multiple primary tumours and/or young women with BC or OC. The presence of the founder effect greatly simplifies the DNA testing procedure, enabling comprehensive studies, such as revealing the influence of hereditary cancer gene mutations on the overall BC/OC incidence rate, as well as analyzing the gene mutations in healthy women [11].

EPIDEMIOLOGY OF THE *BRCA1*, *BRCA2*, *CHEK2* AND *NBS1* MUTATIONS IN RUSSIA

In Russia, the studies of the HBOC syndrome were initiated later than in the U.S. and Europe but they produced rather unexpected results. The first paper published in 1997 reported on the results obtained in patients with familial OC living in Moscow and several other regions of the former Soviet Union [12], the main result being the extremely high frequency of the *BRCA1* 5382insC mutation. As was mentioned above, this mutation had been first found in Jewish women; therefore, it had been for many years considered in the context of that particular ethnic group [13]. However, it appeared that the *BRCA1* 5382insC mutation was not of Jewish origin. This mutation is found not only in females living in various regions of Russia, but also in native populations of Poland, Lithuania, Latvia, and Belarus [14–17]. It is perhaps more accurate to say that the *BRCA1* 5382insC mutation is of Slavic origin, and that the relatively high frequency of this mutation in Ashkenazi Jews observed mostly in Eastern Europe is likely due to the long coexistence of the Slavic and Jewish peoples in the Baltic region and adjacent territories.

The epidemiology of the *BRCA1* 5382insC mutation is surprising, to say the least, since it contradicts the stereotype of the multinational culture in the Russian Empire and the Soviet Union. *BRCA1* 5382insC accounts for up to 90% of all *BRCA1* mutations in women living in distant regions of Russia, ranging from Moscow to St. Petersburg, Krasnodar, Tomsk, etc. [12, 18–20, 22–24, 26]. Moreover, this mutation is dominant in neighboring countries with a mostly Slavic population such as Poland, Belarus, Latvia, and Lithuania [14–17]. Notably, the relative genetic homogeneity of the Slavs is in accordance with the results of general population studies on the genetic diversity of people living in Russia [31]. The *BRCA1* 5382insC allele frequency in healthy women is approximately 0.1%. This variant accounts for approximately 2–5% of total BC cancer incidence. Among the high-risk patients (familial cancers, bilateral breast tumours, or early onset cancers), this mutation is observed in 10% of patients. The *BRCA1* 5382insC contribution to the OC rate is even bigger: this mutation is found in 10–15% of patients (Table). It is important to note that in contrast to the BC, the *BRCA1* 5382insC distribution in women with OC is independent of age, family history, and the number of primary tumours [26]. Therefore, while DNA testing of BC patients can be restricted by high-risk cases, all OC patients have to undergo BRCA testing.

In the pioneering report [12], the relatively frequent *BRCA1* 4153delA (4154delA) mutation was described. The mutation was found not only in Russian patients, but also in those from other neighboring Slavic countries [14–17]. The *BRCA1* 4153delA frequency in Russian patients, however, is an order of magnitude lower than that of the *BRCA1* 5382insC mutation, which complicates the study of the *BRCA1* 4153delA epidemiology. Polish scientists had reported on the preferential association of *BRCA1* 4153delA with OC [14, 32]; however, their observations could not be confirmed in later studies [21].

A number of Russian studies indicate that there is a relatively high frequency of the "Jewish" *BRCA1* 185delAG allele in Russian patients [20, 23, 24, 26]. In contrast to the *BRCA1* 5382insC mutation, however, this mutation is not dominant and could be better explained by interethnic marriages.

The *BRCA1* gene mutations in familial BC/OC patients have been repeatedly analyzed by sequencing of all coding sections, with similar results obtained in Moscow, St. Petersburg, and Tomsk. It has been shown that non-founder mutations are much rarer in Russia than in Europe and North America [12, 18–20, 23, 33]. Given the rapidly falling costs of DNA analysis, it is logical to expect an increase in the use of *BRCA1* sequencing even for the patients with low probability of cancer genetic predisposition. So far, there has been only one study on gross rearrangements of the *BRCA1* gene, and the data indicate a low frequency of such mutations in Russian patients with hereditary BC/OC [33].

Table. Hereditary breast-ovarian cancer genes in Russia

Gene	Major mutations	Frequency in healthy subjects	Frequency in "high-risk" (familial and/or bilateral and/or young-onset) breast cancer patients	Frequency in non-selected breast cancer patients	Frequency in ovarian cancer patients	References
BRCA1	5382insC, 4153delA, 185delAG	~ 0.1%	~ 10%	2–4%	> 10%	[18–26]
CHEK2	1100delC, IVS2+1G>A	< 1%	~ 5%	~ 2%	<1%	[26–29]
NBS1	657del5	0.5%	~ 1%	0.7%	<1%	[29, 30]

While the *BRCA1* gene has been systematically analyzed, data on the *BRCA2* mutations in Russia are scarce. In Siberia, there have been several reported cases of *BRCA2* being inactivated in hereditary BC/OC patients [18]; at the same time, studies performed in Moscow revealed no connection between this gene and hereditary BC/OR in the European part of Russia [23]. Polish scientists performed comprehensive studies showing that the *BRCA2* mutations contributed very little to BC and OC aetiology in Slavs [34].

Another interesting feature of Russian patients is the frequent occurrence of *CHEK2* mutations. This gene, as *BRCA1* and *BRCA2*, participates in the maintaining of genomic integrity. Heterozygous *CHEK2* mutations are frequent in Finland, the Netherlands, Poland, and several other countries [14, 35]. In Russia, *CHEK2* mutations are found in fewer than 2% of "random" BC patients, and in up to 5% of hereditary cancer patients [27]. In contrast to the situation with *BRCA1* and *BRCA2*, heterozygous inactivation of *CHEK2* does not increase the risk of OC [21, 26].

Another important gene for Russia is *NBS1* (*NBN*). Homozygous defects of this gene were found in patients with serious immunodeficiency, the so-called Nijmegen breakage syndrome [36]. Heterozygous *NBS1* mutations are observed mostly in Slavs, and they are associated with an increased BC risk [30, 37, 38]. No increased frequency of this gene defect is observed in OC patients [26]. Nevertheless, in the only reported case of combined germ-line heterozygosity for *BRCA1* and *NBS1* genes in ovarian tumor, there was somatic inactivation of the *NBS1* gene, whereas the *BRCA1* gene remained intact [39]. This observation may be an argument speaking in favor of the involvement of the *NBS1* gene in the degree of OC risk.

MEDICAL ASPECTS OF HEREDITARY BC AND OC IN RUSSIA

The main goal of hereditary cancer syndrome diagnostics is to find healthy women with corresponding mutations. It is believed that timely detection of the genetic defect can help to avoid fatal cancer outcome. Women with heterozygous *BRCA1* and *BRCA2* genes are under regular observation for early BC/OC diagnostics. In addition, preventive surgical removal of target tissues is recommended [40] to women with a BRC A mutation [40].

Healthy carriers of hereditary cancer genes are usually found during the examination of relatives of BC or OC patients with the genetic defect. According to current ethical standards, the patient herself should encourage her relatives to undergo DNA analysis. Our experience shows that very few relatives of the *BRCA* mutation carriers undergo DNA testing. This could be because either hereditary cancer patients conceal their condition to their relatives or the healthy women avoid medical procedures aimed at determining cancer risk. Even more surprising is the fact that the majority of healthy women with *BRCA1* mutations monitored by us are extremely careless when it comes to undergoing preventive screening. Preventive surgery presents the biggest challenge. While it has become a routine clinical practice in the U.S., Canada, Western and Eastern Europe, Israel, Australia, South Africa, Japan, Korea and other countries, yet in Russia the discussion of such an option is suppressed or distorted not only by ordinary people but even by the medical community.

While preventive measures for BRCA carriers are frequently neglected, many doctors are enthusiastic to try novel therapeutic schemes for HBOC patients. In 2009, Polish scientists published the results of clinical studies showing the unique sensitivity of *BRCA1*-associated tumors to cisplatin [41]. This is possible because of unique therapeutic window. In tumors of the *BRCA1* mutation carriers, complete inactivation of this gene is observed. It causes a homologous recombination defect. *BRCA1*-deficient cells are extremely vulnerable to cisplatin, a well-known DNA crosslinking compound causing double-strand breaks. It is important that normal tissues, in contrast to neoplasms, retain heterozygous *BRCA1* status, the presence of a single func-

tional copy of the gene being sufficient for performing its functions. Russian scientists were the first to provide independent confirmation of the results of Byrski *et al.* [42]. Cisplatin is now commonly used for the therapy of *BRCA1*-associated tumors in several Russian clinics.●

This work was supported by the Ministry of Education and Science (contract № 02.740.11.0780), the Russian Foundation for Basic Research (grants № 08-04-00369, 09-04-90402, 10-04-92110, 10-04-92601), and the Government of Moscow (project 15/10-Gene-M).

REFERENCES

1. Benson J.R., Jatoi I., Keisch M., Esteva F.J., Makris A., Jordan V.C. // Lancet. 2009. V. 373. P. 1463–1479.
2. Kristensen G.B., Tropé C. // Lancet. 1997. V. 349. P. 113–117.
3. Narod S.A., Foulkes W.D. // Nat. Rev. Cancer. 2004. V. 4. P. 665–676.
4. Miki Y., Swensen J., Shattuck-Eidens D., Futreal P.A., Harshman K., Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R., Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hattier T., Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z., Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P., Ward J., Tonin P., Narod S., Bristow P.K., Norris F.H., Helvering L., Morrison P., Rosteck P., Lai M., Barrett J.C., Lewis C., Neuhausen S., Cannon-Albright L., Goldgar D., Wiseman R., Kamb A., Skolnick M.H. // Science. 1994. V. 266. P. 66–71.
5. Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J., Collins N., Gregory S., Gumbs C., Micklem G. // Nature. 1995. V. 378. P. 789–792.
6. Gudmundsdottir K., Ashworth A. // Oncogene. 2006. V. 25. P. 5864–5874.
7. Oldenburg R.A., Meijers-Heijboer H., Cornelisse C.J., Devilee P. // Crit. Rev. Oncol. Hematol. 2007. V. 63. P. 125–149.
8. Fackenthal J.D., Olopade O.I. // Nat. Rev. Cancer. 2007. V. 7. P. 937–948.
9. Neuhausen S.L. // Breast Cancer Res. 2000. V. 2. P. 77–81.
10. Ferla R., Calò V., Cascio S., Rinaldi G., Badalamenti G., Carreca I., Surmacz E., Colucci G., Bazan V., Russo A. // Ann. Oncol. 2007. V. 18. Suppl 6. P. vi93–vi98.
11. Abbott A. // Nature. 2000. V. 406. P. 340–342.
12. Gayther S.A., Harrington P., Russell P., Kharkevich G., Garkavtseva R.F., Ponder B.A. // Am. J. Hum. Genet. 1997. V. 60. P. 1239–1242.
13. Roa B.B., Boyd A.A., Volcik K., Richards C.S. // Nat. Genet. 1996. V. 14. P. 185–187.
14. Górski B., Cybulski C., Huzarski T., Byrski T., Gronwald J., Jakubowska A., Stawicka M., Gozdecka-Grodecka S., Szwiec M., Urbański K., Mituś J., Marczyk E., Dziuba J., Wandzel P., Surdyka D., Haus O., Janiszewska H., Debniak T., Tołoczko-Grabarek A., Medrek K., Masojć B., Mierzejewski M., Kowalska E., Narod S.A., Lubiński J. // Breast Cancer Res. Treat. 2005. V. 92. P. 19–24.
15. Tikhomirova L., Sinicka O., Smite D., Eglitis J., Hodgson S.V., Stengrevics A. // Fam. Cancer. 2005. V. 4. P. 77–84.
16. Elsakov P., Kurtinaitis J., Petraitis S., Ostapenko V., Razumas M., Razumas T., Meskauskas R., Petrulis K., Luksite A., Lubiński J., Górski B., Narod S.A., Gronwald J. // Clin. Genet. 2010. V. 78. P. 373-376.
17. Uglanitsa N., Oszurek O., Uglanitsa K., Savonievich E., Lubiński J., Cybulski C., Debniak T., Narod S.A., Gronwald J. // Clin. Genet. 2010 (in press).
18. Tereschenko I.V., Basham V.M., Ponder B.A., Pharoah P.D. // Hum. Mutat. 2002. V. 19. P. 184.
19. Grudinina N.A., Golubkov V.I., Tikhomirova O.S., Brezhneva T.V., Hanson K.P., Vasilyev V.B., Mandelshtam M.Y. // Russ. J. Genet. 2005. V. 41. P. 318–322.
20. Loginova A.N., Pospekhova N.I., Lyubchenko L.N., Budilov A.V., Zakhar'ev V.M., Gar'kavtseva R.F., Ginter E.K., Karpukhin A.V. // Bull. Exp. Biol. Med. 2003. V. 136. P. 276–278.
21. Krylova N.Y., Lobeiko O.S., Sokolenko A.P., Iyevleva A.G., Rozanov M.E., Mitiushkina N.V., Gergova M.M., Porhanova T.V., Urmancheyeva A.F., Maximov S.Y., Togo A.V., Imyanitov E.N. // Hered. Cancer. Clin. Pract. 2006. V. 4. P. 193–196.
22. Sokolenko A.P., Mitiushkina N.V., Buslov K.G., Bit-Sava E.M., Iyevleva A.G., Chekmariova E.V., Kuligina E.Sh., Ulibina Y.M., Rozanov M.E., Suspitsin E.N., Matsko D.E., Chagunava O.L., Trofimov D.Y., Devilee P., Cornelisse C., Togo A.V., Semiglazov V.F., Imyanitov E.N. // Eur. J. Cancer. 2006. V. 42. P. 1380–1384.
23. Smirnova T.Y., Pospekhova N.I., Lyubchenko L.N., Tjulandin S.A., Gar'kavtseva R.F., Ginter E.K., Karpukhin A.V. // Bull. Exp. Biol. Med. 2007. V. 144. P. 83–85.
24. Sokolenko A.P., Rozanov M.E., Mitiushkina N.V., Sherina N.Y., Iyevleva A.G., Chekmariova E.V., Buslov K.G., Shilov E.S., Togo A.V., Bit-Sava E.M., Voskresenskiy D.A., Chagunava O.L., Devilee P., Cornelisse C., Semiglazov V.F., Imyanitov E.N. // Fam. Cancer. 2007. V. 6. P. 281–286.
25. Mitrofanov D.V., Chasovnikova O.B., Kovalenko S.P., Liakhovich V.V. // Mol. Biol. (Mosk). 2009. V. 43. P. 930–936.
26. Suspitsin E.N., Sherina N.Y., Ponomariova D.N., Sokolenko A.P., Iyevleva A.G., Gorodnova T.V., Zaitseva O.A., Yatsuk O.S., Togo A.V., Tkachenko N.N., Shiyanov G.A., Lobeiko O.S., Krylova N.Y., Matsko D.E., Maximov S.Y., Urmancheyeva A.F., Porhanova N.V., Imyanitov E.N. // Hered. Cancer Clin. Pract. 2009. V. 7. P. 5.
27. Chekmariova E.V., Sokolenko A.P., Buslov K.G., Iyevleva A.G., Ulibina Y.M., Rozanov M.E., Mitiushkina N.V., Togo A.V., Matsko D.E., Voskresenskiy D.A., Chagunava O.L., Devilee P., Cornelisse C., Semiglazov V.F., Imyanitov E.N. // Breast Cancer Res. Treat. 2006. V. 100. P. 99–102.
28. Fedorova O.E., Liubchenko L.N., Paiadini Iu.G., Kazubskaia T.P., Amosenko F.A., Gar'kavtseva R.F., Zasedatelev A.S., Nasedkina T.V. // Mol. Biol. (Mosk). 2007. V. 41. P. 32–36.
29. Krylova N.Y., Ponomariova D.N., Sherina N.Y., Ogorodnikova N.Y., Logvinov D.A., Porhanova N.V., Lobeiko O.S., Urmancheyeva A.F., Maximov S.Y., Togo A.V., Suspitsin E.N., Imyanitov E.N. // Hered. Cancer. Clin. Pract. 2007. V. 5. P. 153–156.
30. Buslov K.G., Iyevleva A.G., Chekmariova E.V., Suspitsin E.N., Togo A.V., Kuligina E.Sh., Sokolenko A.P., Matsko D.E., Turkevich E.A., Lazareva Y.R., Chagunava O.L., Bit-Sava E.M., Semiglazov V.F., Devilee P., Cornelisse C., Hanson K.P., Imyanitov E.N. // Int. J. Cancer. 2005. V. 114. P. 585–589.
31. Balanovsky O., Rootsi S., Pshenichnov A., Kivisild T., Churnosov M., Evseeva I., Pocheshkhova E., Boldyreva M., Yankovsky N., Balanovska E., Villems R. // Am. J. Hum. Genet. 2008. V. 82. P. 236–250.

32. Menkiszak J., Gronwald J., Górski B., Jakubowska A., Huzarski T., Byrski T., Foszczyńska-Kłoda M., Haus O., Janiszewska H., Perkowska M., Brozek I., Grzybowska E., Zientek H., Góźdź S., Kozak-Klonowska B., Urbański K., Miturski R., Kowalczyk J., Pluzańska A., Niepsuj S., Koc J., Szwiec M., Drosik K., Mackiewicz A., Lamperska K., Strózyk E., Godlewski D., Stawicka M., Waśko B., Bebenek M., Rozmiarek A., Rzepka-Górska I., Narod S.A., Lubiński J. // Int. J. Cancer. 2003. V. 106. P. 942–945.
33. Iyevleva A.G., Suspitsin E.N., Kroeze K., Gorodnova T.V., Sokolenko A.P., Buslov K.G., Voskresenskiy D.A., Togo A.V., Kovalenko S.P., van der Stoep N., Devilee P., Imyanitov E.N. // Cancer Lett. 2010. V. 298. P. 258–263.
34. Górski B., Byrski T., Huzarski T., Jakubowska A., Menkiszak J., Gronwald J., Pluzańska A., Bebenek M., Fischer-Maliszewska L., Grzybowska E., Narod S.A., Lubiński J. // Am. J. Hum. Genet. 2000. V. 66. P. 1963–1968.
35. *CHEK2* Breast Cancer Case-Control Consortium. // Am. J. Hum. Genet. 2004. V. 74. 1175–1182.
36. Varon R., Seemanova E., Chrzanowska K., Hnateyko O., Piekutowska-Abramczuk D., Krajewska-Walasek M., Sykut-Cegielska J., Sperling K., Reis A. // Eur. J. Hum. Genet. 2000. V. 8. P. 900–902.
37. Steffen J., Nowakowska D., Niwińska A., Czapczak D., Kluska A., Piatkowska M., Wiśniewska A., Paszko Z. // Int. J. Cancer. 2006. V. 119. P. 472–475.
38. Bogdanova N., Feshchenko S., Schürmann P., Waltes R., Wieland B., Hillemanns P., Rogov Y.I., Dammann O., Bremer M., Karstens J.H., Sohn C., Varon R., Dörk T. // Int. J. Cancer. 2008. V. 122. P. 802–806.
39. Porhanova N.V., Sokolenko A.P., Sherina N.Y., Ponomariova D.N., Tkachenko N.N., Matsko D.E., Imyanitov E.N. // Cancer Genet. Cytogenet. 2008. V. 186. P. 122–124.
40. Dent R., Warner E. // Semin. Oncol. 2007. V. 34. P. 392–400.
41. Byrski T., Huzarski T., Dent R., Gronwald J., Zuziak D., Cybulski C., Kladny J., Gorski B., Lubinski J., Narod S.A. // Breast Cancer Res. Treat. 2009. V. 115. P. 359–363.
42. Moiseyenko V.M., Protsenko S.A., Brezhnev N.V., Maximov S.Y., Gershveld E.D., Hudyakova M.A., Lobeiko O.S., Gergova M.M., Krzhivitskiy P.I., Semionov I.I., Matsko D.E., Iyevleva A.G., Sokolenko A.P., Sherina N.Y., Kuligina E.Sh., Suspitsin E.N., Togo A.V., Imyanitov E.N. // Cancer Genet. Cytogenet. 2010. V. 197. P. 91–94.

Dosage Compensation of Sex Chromosome Genes in Eukaryotes

E. V. Dementyeva[1,2], S. M. Zakian[1,2,3]*
[1]Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences
[2]Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, Russian Academy of Sciences
[3]Research Center of Clinical and Experimental Medicine, Siberian Branch, Russian Academy of Medical Sciences
*E-mail: zakian@bionet.nsc.ru
Received 24.09.2010

ABSTRACT **Sex chromosome evolution is accompanied by significant divergence in morphology and gene content and results in most genes of one of the sex chromosomes being present in two dosages in one sex and in one dosage in the other. To eliminate the difference in the expression levels of these genes between sexes and to restore equal expression levels of the genes between sex chromosomes and autosomes, mechanisms of dosage compensation have appeared. Studies of three classical objects, *Drosophila melanogaster*, *Caenorhabditis elegans*, and mammals, have shown that dosage compensation of X-linked genes can be achieved through completely different chromosome-wide mechanisms. New data on sex chromosome gene expression demonstrating that many sex chromosome genes can be expressed at different levels in males and females were recently obtained from birds and butterflies. In this review, dosage compensation mechanisms in *D. melanogaster*, *C. elegans*, and mammals are considered and the data on sex chromosome gene expression in birds and butterflies, and their influence on our view of dosage compensation, are discussed.**
KEYWORDS **dosage compensation, sex chromosomes, gene expression, X-chromosome inactivation**

COEVOLUTION OF SEX CHROMOSOMES AND DOSAGE COMPENSATION MECHANISMS

In a variety of organisms, sex correlates with a distinct sex chromosome set. In particular, in *Drosophila melanogaster*, as well as in most mammals, females have two X-chromosomes, while males are heterogametic with two different sex chromosomes, X and Y. Nonetheless, the systems determining sex in *D. melanogaster* and mammals are completely different. In *D. melanogaster*, sex depends on the ratio between doses of X-linked and autosomal genes [1], whereas in mammals the presence of the Y-chromosome, rather the *Sry* gene responsible for male sex determination, is crucial [2]. In contrast, in birds, butterflies, and some reptiles, females are heterogametic (chromosomes Z and W), while males have two Z-chromosomes. The sex chromosomes X and Y, as well as Z and W, considerably differ from each other in size, morphology, and gene content (Fig. 1). The chromosomes Y and W are heterochromatinized and mainly composed of tandem DNA repeats, and their gene content is poor in comparison with that of X- and Z-chromosomes.

It is thought that X- and Y-chromosomes appeared independently in different taxa and originated from a pair of homologous autosomes. The first step in sex chromosome evolution was the development of a genetic system of sex determination in a population of hermaphrodites or individuals whose sex is determined by temperature. The most consistent is an order of events with initial mutation leading to the appearance of a recessive gene of male sterility on the future X-chromosome, followed by the appearance of a dominant gene of female sterility on the future Y-chromosome. This resulted in the suppression of recombination between the X- and Y-chromosomes at the loci, which enabled the linkage of the genes responsible for male or female sex determination. The following step was the accumulation of genes beneficial to males (but decreasing the female's fitness) on the Y-chromosome. The necessity of a tight linkage between these genes

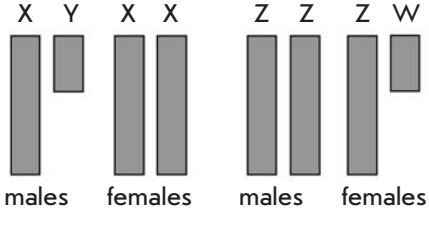


Fig. 1. XY and ZW sex chromosome systems.

and the Y-chromosome resulted in the suppression of recombination between the X- and Y-chromosomes in new loci and gradual expansion of the nonrecombining region. Suppression of recombination led to the accumulation of mutations and deletions in Y-linked genes, which are not associated with the formation of male features, thereby resulting in their degradation. Finally, the entire Y-chromosome could be lost, which probably occurred in *Caenorhabditis elegans* males that only possess the X-chromosome. A similar process is likely to have resulted in the divergence of Z- and W-chromosomes [3, 4].

To compensate for such an essential loss of genes on the Y-chromosome, natural selection might have favored mechanisms that elevated the expression of X-linked genes in males [5]. Up-regulation of the genes localized on the single X-chromosome in males has been known for a long time and is well-studied in *D. melanogaster* [6]. A similar path of restoring the X-linked gene expression level (dosage compensation) was proposed for mammals and *C. elegans*, but convincing arguments took a long time to emerge. A hypothesis on X-chromosome up-regulation in mammal and *C. elegans* males was recently confirmed thanks to the development of microarray techniques. This method allowed to determine the mean expression level of autosomal and X-linked genes, which was found to be equal in males of both mammals and *C. elegans* [7–9].

An increase in the transcription level of X-linked genes might result in an excess of their products in females. However, studies on gene expression using microarray have shown that X-linked genes are expressed in females of *D. melanogaster*, *C. elegans*, and mammals at the same level as autosomal genes [7–9]. Hence, females should also possess the mechanism(s) supporting the transcription balance between the X-linked and autosomal genes, as well as an equal expression level of the X-linked genes in both sexes. Despite the fact that it has a similar origin, dosage compensation of X-linked genes occurs in different ways in *D. melanogaster*, *C. elegans*, and mammals (Fig.2). In *D. melanogaster*, dosage compensation only occurs in males, while in females the expression levels of genes localized on the autosomes and both X-chromosomes are equal [7]. In *C. elegans*, the single X-chromosome in males and both X-chromosomes in hermaphrodites are up-regulated. The restoration of the transcription balance in hermaphrodites is achieved by a specific mechanism that partially down-regulates gene expression on both X-chromosomes [7]. In mammals, gene expression is up-regulated on the X-chromosome in males and one of the two X-chromosomes in females. On the second X-chromosome, transcription of most genes is completely repressed; i. e., this X-chromosome undergoes

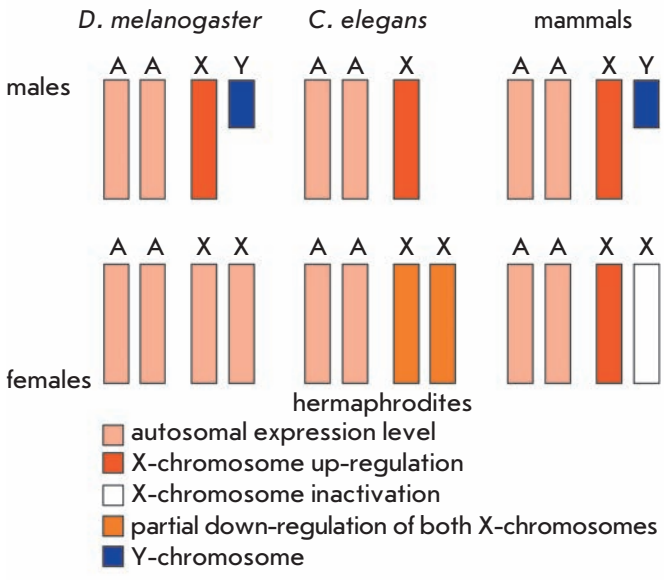


Fig. 2. Diversity of X-linked gene dosage compensation systems A – autosomal set, X and Y – sex chromosomes.

inactivation [7, 8]. The mechanisms underlying these processes deserve a more detailed examination.

DOSAGE COMPENSATION OF X-LINKED GENES IN *D. melanogaster*

The elevated level of expression of X-linked genes in *D. melanogaster* males is supported by a complex composed of six proteins: MSL1 (male-specific lethal 1), MSL2, MSL3, MOF (males absent on the first), MLE (maleless), and JIL1 (Janus kinase 1), and two noncoding RNAs: *roX1* and *roX2* (RNA on the X). The key element for the assembly of this complex is MSL2, which is only synthesized in males. The MSL2 protein is absent in females, so other components cannot form the dosage compensation complex. According to the generally accepted model, MSL2 stabilizes MSL1 by direct interaction, thus forming a platform for further assembly of the dosage compensation complex [10]. The proteins MOF and JIL1 are responsible for the activation of X-linked gene transcription in males. The MOF protein acetylates histone H4 at Lys16 (H4K16). This modification is characteristic of the transcriptionally active chromatin and specific for the male X-chromosome [11, 12]. However, recent data suggest that MOF can acetylate H4K16 not only on the X-chromosome, but also on the autosomes of both sexes [13]. MOF has been found to interact not only with the MSL complex, but also with the so-called NSL (nonspecific lethal) complex that binds to promotors of transcriptionally active autosomal genes in males as well as autosomal and sex-

chromosomal genes in females. Hence, MOF, via interaction with different protein complexes, is implicated in two processes: dosage compensation of X-linked genes in males and general regulation of gene transcription in *D. melanogaster* [14, 15]. Moreover, homologues of MOF and the NSL complex exist in mammals, in which they play the same role in the histone H4 acetylation [16, 17]. These facts suggest that the mechanism of dosage compensation in *D. melanogaster* did not appear *de novo* but was formed on the basis of existing proteins which could retain their initial functions. JIL1 kinase is freely associated with the MSL complex and phosphorylates histone H3 at Ser10. This modification is also implicated in the formation of transcriptionally active chromatin, likely counteracting the binding of the heterochromatin protein HP1 [18, 19]. Thus, up-regulation of X-linked genes in *D. melanogaster* is achieved through the creation of an "open" incompact chromatin structure accessible to transcription factors [20]. RNA-DNA-helicase MLE is thought to promote the integration of *roX1* and *roX2* RNAs into the dosage compensation complex [10]. These RNAs are interchangeable and essential for the binding of the dosage compensation complex with the X-chromosome [21]. Interestingly, the human homologue of the MSL complex does not contain the *roX1* and *roX2* RNAs. So, the recruitment of these noncoding RNAs into the MSL complex might be the turning point in the formation of the dosage compensation mechanism in *D. melanogaster* [22].

The X-chromosome of *D. melanogaster* has no less than 150 specific sites, called chromatin entry sites, which contain MSL recognition elements for the binding of the dosage compensation complex. Following the binding of these sites, the MSL complex spreads along the X-chromosome and interacts with actively transcribed genes [23]. Epigenetic features, such as trimethylated H3K36 (a characteristic of transcribed genes), rather than nucleotide sequences, are most likely significant at this stage [24]. Nevertheless, not all transcriptionally active genes of the *D. melanogaster* male X-chromosome bind the dosage compensation complex. Moreover, binding of the MSL complex not always leads to exactly a twofold increase in the X-linked gene expression level. In some cases, the level of transcription remains virtually unchanged [25–27]. Therefore, the mechanism controlling the expression level of individual X-linked genes in *D. melanogaster* males has yet to be identified.

The mechanism underlying up-regulation of X-linked genes in mammalian and *C. elegans* males remains unknown. It is likely to be supported by epigenetic mechanisms as in *D. melanogaster*. However, it is worth noting that no significant difference has been found between the chromatin structures of the X-chromosome and autosomes. So, the X-chromosome up-regulation in males may be a result of alterations in the nucleotide sequences of the gene regulatory regions that took shape during evolution [8, 28]. Besides, there is another possible way of enhancing X-linked gene expression in mammals. The fact is that the genes of the active and inactive X-chromosomes differ in methylation patterns. The alleles of the inactive X-chromosome are hypermethylated at the CpG-dinucleotides of promotor regions, which matches their inactivation. At the same time, the alleles of the active X-chromosome in females and genes of the X-chromosome in males are hypermethylated at the CpG-dinucleotides of gene bodies [29]. However, it remains absolutely unclear how methylation of gene bodies can lead to elevated expression of X-linked genes in mammals.

DOSAGE COMPENSATION OF X-LINKED GENES IN *C. elegans*

As mentioned above, dosage compensation of the *C. elegans* X-linked genes includes two processes: X-linked gene up-regulation in males and partial repression of the genes localized on both X-chromosomes in hermaphrodites. While the mechanism of the first process is absolutely unknown, the complex of dosage compensation composed of nine proteins SDC-1, SDC-2, SDC-3, DPY-21, DPY-26, DPY-27, DPY-28, DPY-30, and MIX1 has been characterized in *C. elegans* hermaphrodites [30]. Three proteins (DPY-26, DPY-27, and DPY-28) closely resemble the proteins of the 13S condensin complex responsible for chromosome compaction in mitosis and meiosis not only in *C. elegans*, but also in other eukaryotes. Another protein, MIX1 (mitosis and X-associated protein 1), is common to both complexes [31–34]. However, not only MIX1 has a dual function. The protein DPY-28 controls the number and distribution of crossovers between homologous chromosomes in meiosis [35]. DPY-30 is part of a complex that is homologous to the yeast complex Set1/COMPASS methylating histone H3. DPY-30 is likely implicated in both dosage compensation and the general regulation of gene transcription in *C. elegans* males and hermaphrodites [36, 37]. An important role in the assembly and function of the dosage compensation complex is played by the protein SDC-2 (sex determination and dosage compensation 2). Unlike other proteins, SDC-2 is only expressed in hermaphrodites and seems to be responsible for the specific impact of the dosage compensation complex on the X-chromosome, because it can bind to the X-chromosome independently of other components of the complex [38]. The complex assembly begins with the interaction between SDC-2, SDC-3, and DPY-30, which creates a platform for the binding of all other

proteins to the X-chromosome [39–41]. Interestingly, the same complex (except for DPY-21) is implicated in a 20-fold transcription repression of the autosomal gene *her-1* (hermaphrodization of X0 animals), which is responsible for male sex determination [38, 41]; i.e. this complex participates not only in the dosage compensation of X-linked genes, but also in the sex determination system.

To bind to the dosage compensation complex there are specific nucleotide sequences on the *C. elegans* X-chromosome, but their density is significantly lower than that on the *D. melanogaster* X-chromosome (~40 and 150, respectively). These sequences are divided into two types: the rex- and dox-sites. Rex (recruitment elements on X)-sites can bind to the dosage compensation complex regardless of whether they are localized on the X-chromosome or autosomes and are most likely responsible for the primary recognition of the complex. Dox (dependent on X)-sites only interact with the dosage compensation complex when localized on the X-chromosome, and they are mainly implicated in the spreading of the complex along the X-chromosomes of *C. elegans* hermaphrodites [42].

The mechanism of dosage compensation complex-dependent partial repression of X-linked gene expression in hermaphrodites is not yet known; however, a certain similitude between the *C. elegans* dosage compensation and the 13S condensin complexes allows to assume that the same principle underlies both the transcription repression of X-linked genes and the chromosome condensation in mitosis and meiosis [37]. The similarity to the 13S condensin complex and the dual function of some proteins of the dosage compensation complex suggest that in both *C. elegans* and *D. melanogaster* dosage compensation appeared due to the acquisition of novel functions by existing proteins, rather than the development of an absolutely new mechanism. Which X-linked genes are subjects to dosage compensation in *C. elegans* hermaphrodites (and to what extent) remains unknown.

DOSAGE COMPENSATION OF X-LINKED GENES IN MAMMALS

Like *C. elegans*, mammals demonstrate up-regulation of X-linked genes in both sexes. The restoration of the gene transcription balance in females is achieved by transcription repression (inactivation) of the majority of genes localized on one of the two X-chromosomes [43]. X-inactivation may be either random or imprinted [28]. When X-inactivation is imprinted, the paternally inherited X-chromosome is predominantly inactivated. This variant of inactivation occurs in marsupials as well as in the extraembryonic tissues of some eutherians. When the inactivation is random, the chances of the paternal and maternal X-chromosomes being inactivated are equal. This type of X-inactivation takes place in somatic tissues of eutherians.

The eutherian X-chromosome has a specific locus called the X-inactivation center. One gene from this locus, *Xist*, is the key gene in the initiation of the X-inactivation process. It encodes a noncoding RNA, which then spreads along the further inactive X-chromosome, which leads to a series of epigenetic changes [28, 44–46]. As a result, RNA polymerase II is excluded from the inactive X-chromosome, and chromatin-modifying complexes appear. Consequently, the inactive X-chromosome loses modifications that are characteristic of transcriptionally active chromatin, such as histone H3 dimethylated at Lys4 (H3K4) and acetylated histones H3 and H4. Instead, the inactive X-chromosome gains modifications characteristic of transcriptionally inactive chromatin, such as histone H3 trimethylated at Lys27 (H3K27), histone H2A ubiquitinylated at Lys119 (uH2A), histone H3 dimethylated at Lys9 (H3K9), and histone H4 monomethylated at Lys20 (H4K20). In addition, the inactive X-chromosome becomes late-replicating and associates with a histone H2A variant (macroH2A) containing a nonhistone domain. The last epigenetic event in the X-inactivation process is methylation at X-linked gene promotor regions, which allows to maintain the stability of the inactive state of the X-chromosome. Complexes of polycomb proteins are implicated in the establishment of the inactive state of the X-chromosome in mammalian females. PRC1 (polycomb repressor complex 1) is responsible for the ubiquitinylation of histone H2A [47, 48], while PRC2 is responsible for H3K27 trimethylation [49, 50]. However, these complexes are not specific to females, as they are also implicated in the repression of both X-linked and autosomal genes [22]. The enzymes fulfilling H3K9 dimethylation and H4K20 monomethylation have not been known precisely. It is believed that they are methyltransferase G9a and PR-Set7, respectively [51, 52]. It is worth noting that noncoding RNAs and chromatin-modifying complexes are involved in the dosage compensation of X-linked genes both in mammals and *D. melanogaster*, but their effects on gene expression are diametrically opposite. The question of how *Xist* RNA interacts with chromatin-modifying factors still remains open. Moreover, the gene *Xist* was not found in marsupials [53], despite the fact that their chromatin modification patterns on the inactive X-chromosome closely resemble those in eutherians. It becomes obvious that both the X-inactivation center and random X-inactivation only developed in eutherians [53, 54], and that the X-inactivation process in marsupials differs from that in placental mammals.

In its mechanism, X-inactivation is similar to the imprinting of autosomal genes. In both cases, noncoding RNAs are involved whose expression leads to the establishment of the same chromatin modifications: hypomethylated H3K4, hypoacetylated H3K9, trimethylated H3K27, uH2A, dimethylated H3K9, and DNA methylation [55, 56]. The final result of both processes is the transcription repression of one of two alleles. Hence, the method of gene transcription repression during X-inactivation in mammalian females is not unique; the same mechanism is also at play in the establishment of monoallele expression for certain autosomal genes.

It is worth noting that not all genes of the inactive X-chromosome undergo inactivation. Studies on the expression status of human X-linked genes have shown that 15% of genes always escape X-inactivation, while 10% of genes have heterogeneous expression; i.e., they undergo X-inactivation in some women and escape X-inactivation in others [57]. Besides, genes escaping X-inactivation were found in mice and some other mammals [58, 59]. However, the reason why some X-linked genes escape X-inactivation is as yet unknown. In some cases, this may be explained by the presence of a Y-chromosomal homologue of an X-linked gene. In this case, escaping X-inactivation enables the restoration of equal X-linked gene expression between the sexes. Nonetheless, many X-linked genes escaping X-inactivation have no Y-homologues. It is possible that a higher expression level of these genes in females is associated with the formation of female-specific features [60, 61]. Interestingly, the expression level of many genes escaping X-inactivation on the inactive X-chromosome is much lower than that on the active X-chromosome [8, 9, 57]. This suggests that a higher expression level of these genes in females is of little importance. It is also possible that an imbalance of X-linked genes can be evened out after transcription [60, 61].

It has been suggested that special elements are necessary for the effective spreading of the inactive state along the X-chromosome. The most likely candidates are long, interspersed nuclear elements (LINEs) [62]. This hypothesis is supported by the fact that murine and human X-chromosomes are twice richer in LINEs as compared to autosomes. It is worth noting that the distribution of LINEs on the human X-chromosome correlates with gene expression status. The highest density of LINEs is observed in the X-inactivation center and in regions of gene inactivation. Conversely, the density of LINEs in regions that escape X-inactivation is lower [57, 63, 64]. Yet, what are the sequences necessary for effective spreading of the inactive state along the X-chromosome and what are the mechanisms of their action remains unknown.

COMMON FEATURES OF X-LINKED GENE DOSAGE COMPENSATION SYSTEMS

Examination of three model objects (*D. melanogaster*, *C. elegans*, and mammals) demonstrates that X-linked gene dosage compensation can occur via a variety of mechanisms. The difference in dosage compensation mechanisms appears to reflect an independent origin of sex chromosomes in these species and, as a result, independent formation of the mechanisms directed toward the regulation of X-linked gene expression. Despite the difference in means of X-linked gene dosage compensation in *D. melanogaster*, *C. elegans*, and mammals, there are several common features. First, dosage compensation is achieved via mechanisms operating at the chromosomal level; these mechanisms do not appear *de novo*: existing proteins and protein complexes adapt to the regulation of X-linked gene expression. Second, up-regulation of the single X-chromosome in males is common to all three dosage compensation systems, though the mechanisms underlying this phenomenon may differ. Third, the required gene expression level is supported via a change in the X-chromosomal chromatin structure by chromatin-modifying complexes. In *D. melanogaster* and mammals, the effect of chromatin-modifying complexes in the course of X-linked gene dosage compensation is associated with the expression of noncoding RNAs. Tight association between the noncoding RNAs and regulation of gene expression implies that noncoding RNA is likely to be also found in the *C. elegans* dosage compensation system. Fourth, the X-chromosome contains a set of sequences responsible for binding and effective spreading of the dosage compensation complexes. Thus, the mechanisms of dosage compensation enable the leveling of the expression of autosomal and X-linked genes, as well as maintenance of an equal expression level of X-linked genes in both sexes. Transcriptional balance of X-linked genes is supported in different somatic cell types and the germinal cells of *D. melanogaster* and mammals in both sexes [7, 8], thus suggesting its importance to the organism.

DOSAGE COMPENSATION OF Z-LINKED GENES IN BIRDS AND BUTTERFLIES

By analogy with the XY system of sex chromosomes, one might expect that the ZW system of sex chromosomes should also be characterized by the up-regulation of genes on the single Z-chromosome in females (heterogametic sex). However, early studies on the expression of a small number of Z-linked genes in birds and butterflies showed an elevated expression of some Z-linked genes in males, as compared to females [65–68]. Thus, the existence of Z-linked gene dosage compensation was called into doubt for a long time.

The use of microarray techniques allowed to determine the level of Z-linked and autosomal gene expres-

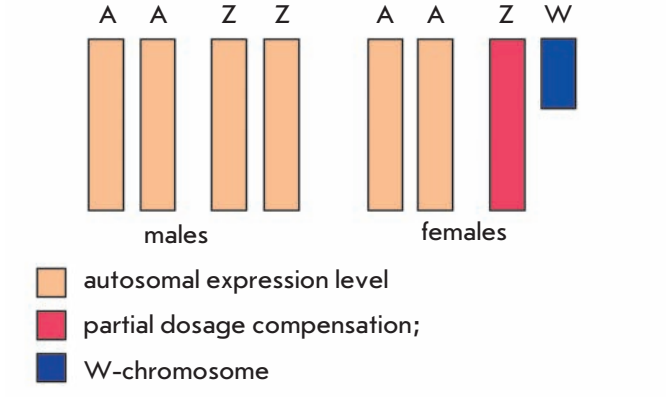


Fig. 3. Dosage compensation of Z-chromosome genes in birds and butterflies A – autosomal set, Z and W – sex chromosomes.

sion in two avian species (chicken and zebra-finch) and in silkworm. The ratio between male and female Z-linked gene expression levels in birds ranged between 1 and 2: i.e., one gets the impression that the Z-chromosome is between dosage compensation at the chromosomal level and the lack of dosage compensation [69, 70]. Similar data was obtained from the study of Z-linked gene expression in silkworm [71]. Moreover, in zebra-finch, Z-linked genes were distinctly divided into two groups: genes with an equal expression level in both sexes and those with a higher expression level in males [69]. Birds and butterflies do not appear to have mechanisms controlling the gene expression of the whole Z-chromosome; however, some Z-linked genes in females do undergo dosage compensation (Fig.3). The mechanisms involved in this process are not yet understood. Nonetheless, a specific locus, MHM (male hypermethylated), has been found on the avian Z-chromosome. This locus is hypermethylated in males, while in females a noncoding RNA is transcribed from this locus and accumulates in the region surrounding MHM. In females, this region is acetylated at Lys16 of histone H4 (H4K16). Besides, despite the genes undergoing dosage compensation being distributed along the Z-chromosome, the majority of them are concentrated near the MHM locus. It is likely that Z-linked gene dosage compensation in birds occurs the same way it does in *D. melanogaster*: noncoding RNA and H4K16 acetylation provide an elevated expression level of Z-linked genes in females [72].

The difference in the degrees of dosage compensation between the X- and Z-chromosomes might be ascribed to their age. When sex chromosomes are young enough, the mechanisms controlling gene expression might not have developed yet. While avian and mammalian sex chromosomes are close in age (no less than 150 and 166 MYA, respectively), the sex chromosomes of *D. melanogaster* are relatively young (~65 MYA), but the age is enough in order for the chromosomal dosage compensation mechanism to have developed. Hence, the age of sex chromosomes does not influence the extent of dosage compensation [22, 73]. It is known that hemizygosity at several genes or small genomic regions may remain, with no consequences for the organism. The avian and butterfly Z-chromosomes contain about 840 and 600 genes, respectively, which is considerably less than the number of genes on the *D. melanogaster*, *C. elegans*, and human X-chromosomes (2,300; 3,100; and 1,100 genes, respectively). It is likely that it is the lower number of genes on the sex chromosomes of birds and butterflies that allows them to do without chromosomal dosage compensation mechanisms. However, hemizygosity at several hundred genes must be lethal anyway; so the limited dosage compensation in birds and butterflies cannot be the result of lower gene density on sex chromosomes [22, 73]. As of now, the local dosage compensation is only found in species whose heterogametic sex is female (ZW). Since the Z-linked gene expression level was only examined in representatives of two taxa, it remains unclear whether this path of dosage compensation is characteristic of organisms with the ZW sex chromosome system or whether it is a coincidence. The study of other taxa, whose heterogametic sex is female, will likely answer this question [22, 73].

CONCLUSION

The data on sex-chromosomal gene expression in birds and butterflies force us to look anew at the problem of gene dosage compensation. It is becoming obvious that sex-chromosomal genes undergo dosage compensation to different extents, up to its complete escaping. It is possible that dosage compensation mechanisms evolved to control the expression of a distinct gene set, rather than the entire sex chromosome. This postulate seems to be correct not only for the Z-chromosome, but also for the X-chromosome, because genes escaping dosage compensation were found in mammals and *D. melanogaster*. Further studies will probably focus on the identification of the sex-chromosomal genes requiring dosage compensation, as well as on the mechanisms that determine the extent of dosage compensation for individual genes. Another important line of inquiry may be uncovering the mechanisms underlying the up-regulation of the X-linked genes in mammals and *C. elegans*. Studies on heteromorphic sex chromosomes in new taxa could shed light on the matter. ●

This study was supported by the RAS Presidium Program Molecular and Cell Biology.

REFERENCES

1. Bridges C.B. // Am. Nat. 1925. V. 59. P. 127–137.
2. Koopman P., Gubbay J., Vivian N., Goodfellow P., Lovell-Badge R. // Nature. 1991. V. 351. P. 117–121.
3. Charlesworth B. // Science. 1991. V. 251. P. 1030–1033.
4. Charlesworth B. // Curr. Biol. 1996. V. 6. P. 149–162.
5. Dementyeva E.V., Shevchenko A.I., Zakian S.M. // Bioessays. 2009. V. 31. P. 21–28.
6. Akhtar A. // Curr. Opin. Genet. Dev. 2003. V. 13. P. 161–169.
7. Gupta V., Parisi M., Sturgill D., Nuttall R., Doctolero M., Dudko O.K., Malley J.D., Eastman P.S., Oliver B. // J. Biol. 2006. V. 5. P. 3.
8. Nguyen D.K., Disteche C.M. // Nat. Genet. 2006. V. 38. P. 47–53.
9. Johnston C.M., Lovell F.L., Leongamornlert D.A., Stranger B.E., Dermitzakis E.T., Ross M.T. // PLoS Genet. 2008. V. 4. e9.
10. Straub T., Becker P.B. // Nat. Rev. Genet. 2007. V. 8. P. 47–57.
11. Akhtar A., Becker P.B. // Mol. Cell. 2000. V. 5. P. 367–375.
12. Lucchesi J.C., Kelly W.G., Panning B. // Annu. Rev. Genet. 2005. V. 39. P. 615–651.
13. Kind J., Vaquerizas J.M., Gebhardt P., Gentzel M., Luscombe N.M., Bertone P., Akhtar A. // Cell. 2008. V. 133. P. 813–828.
14. Prestel M., Feller C., Straub T., Mitlohner H., Becker P.B. // Mol. Cell. 2010. V. 38. P. 815–826.
15. Raja S.J., Charapitsa I., Conrad T., Vaquerizas J.M., Gebhardt P., Holz H., Kadlec J., Fraterman S., Luscombe N.M., Akhtar A. // Mol. Cell. 2010. V. 38. P. 827–841.
16. Taipale M., Rea S., Richter K., Vilar A., Lichter P., Imhof A., Akhtar A. // Mol. Cell Biol. 2005. V. 25. P. 6798–6810.
17. Cai Y., Jin J., Swanson S.K., Cole M.D., Choi S.H., Florens L., Washburn M.P., Conaway J.W., Conaway R.C. // J. Biol. Chem. 2010. V. 285. P. 4268–4272.
18. Lerach S., Zhang W., Deng H., Bao X., Girton J., Johansen J., Johansen K.M. // Genesis. 2005. V. 43. P. 213–215.
19. Ebert A., Schotta G., Lein S., Kubicek S., Krauss V., Jenuwein T., Reuter G. // Genes Dev. 2004. V. 18. P. 2973–2983.
20. Park Y., Kuroda M.I. // Science. 2001. V. 293. P. 1083–1085.
21. Li F., Schiemann A.H., Scott M.J. // Mol. Cell Biol. 2008. V. 28. P. 1252–1264.
22. Vicoso B., Bachtrog D. // Chromosome Res. 2009. V. 17. P. 585–602.
23. Alekseyenko A.A., Peng S., Larschan E., Gorchakov A.A., Lee O.K., Kharchenko P., McGrath S.D., Wang C.I., Mardis E.R., Park P.J., Kuroda M.I. // Cell. 2008. V. 134. P. 599–609.
24. Larschan E., Alekseyenko A.A., Gortchakov A.A., Peng S., Li B., Yang P., Workman J.L., Park P.J., Kuroda M.I. // Mol. Cell. 2007. V. 28. P. 121–133.
25. Hamada F.N., Park P.J., Gordadze P.R., Kuroda M.I. // Genes Dev. 2005. V. 19. P. 2289–2294.
26. Gilfillan G.D., Straub T., de Wit E., Greil F., Lamm R., van Steensel B., Becker P.B. // Genes Dev. 2006. V. 20. P. 858–870.
27. Legube G., McWeeney S.K., Lercher M.J., Akhtar A. // Genes Dev. 2006. V. 20. P. 871–883.
28. Heard E., Disteche C.M. // Genes Dev. 2006. V. 20. P. 1848–1867.
29. Hellman A., Chess A. // Science. 2007. V. 315. P. 1141–1143.
30. Meyer B.J., McDonel P., Csankovszki G., Ralston E. // Cold Spring Harb. Symp. Quant. Biol. 2004. V. 69. P. 71–79.
31. Chan R.C., Severson A.F., Meyer B.J. // J. Cell Biol. 2004. V. 167. P. 613–625.
32. Chuang P.T., Albertson D.G., Meyer B.J. // Cell. 1994. V. 79. P. 459–474.
33. Lieb J.D., Capowski E.E., Meneely P., Meyer B.J. // Science. 1996. V. 274. P. 1732–1736.
34. Lieb J.D., Albrecht M.R., Chuang P.T., Meyer B.J. // Cell. 1998. V. 92. P. 265–277.
35. Tsai C.J., Mets D.G., Albrecht M.R., Nix P., Chan A., Meyer B.J. // Genes Dev. 2008. V. 22. P. 194–211.
36. Nagy P.L., Griesenbeck J., Kornberg R.D., Cleary M.L. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. P. 90–94.
37. Meyer B.J. // WormBook. 2005. P. 1–14.
38. Dawes H.E., Berlin D.S., Lapidus D.M., Nusbaum C., Davis T.L., Meyer B.J. // Science. 1999. V. 284. P. 1800–1804.
39. Chuang P.T., Lieb J.D., Meyer B.J. // Science. 1996. V. 274. P. 1736–1739.
40. Davis T.L., Meyer B.J. // Development. 1997. V. 124. P. 1019–1031.
41. Yonker S.A., Meyer B.J. // Development. 2003. V. 130. P. 6519–6532.
42. Jans J., Gladden J.M., Ralston E.J., Pickle C.S., Michel A.H., Pferdehirt R.R., Eisen M.B., Meyer B.J. // Genes Dev. 2009. V. 23. P. 602–618.
43. Lyon M.F. // Nature. 1961. V. 190. P. 372–373.
44. Heard E. // Curr. Opin. Genet. Dev. 2005. V. 15. P. 482–489.
45. Shevchenko A.I., Pavlova S.V., Dement'eva E.V., Golubeva D.V., Zakian S.M. // Genetika. 2006. V. 42. P. 1225–1234.
46. Wutz A., Gribnau J. // Curr. Opin. Genet. Dev. 2007. V. 17. P. 387–393.
47. de Napoles M., Mermoud J.E., Wakao R., Tang Y.A., Endoh M., Appanah R., Nesterova T.B., Silva J., Otte A.P., Vidal M., Koseki H., Brockdorff N. // Dev. Cell. 2004. V. 7. P. 663–676.
48. Fang J., Chen T., Chadwick B., Li E., Zhang Y. // J. Biol. Chem. 2004. V. 279. P. 52812–52815.
49. Silva J., Mak W., Zvetkova I., Appanah R., Nesterova T.B., Webster Z., Peters A.H., Jenuwein T., Otte A.P., Brockdorff N. // Dev. Cell. 2003. V. 4. P. 481–495.
50. Cao R., Zhang Y. // Curr. Opin. Genet. Dev. 2004. V. 14. P. 155–164.
51. Ohhata T., Tachibana M., Tada M., Tada T., Sasaki H., Shinkai Y., Sado T. // Genesis. 2004. V. 40. P. 151–156.
52. Nishioka K., Rice J.C., Sarma K., Erdjument-Bromage H., Werner J., Wang Y., Chuikov S., Valenzuela P., Tempst P., Steward R., Lis J.T., Allis C.D., Reinberg D. // Mol. Cell. 2002. V. 9. P. 1201–1213.
53. Duret L., Chureau C., Samain S., Weissenbach J., Avner P. // Science. 2006. V. 312. P. 1653–1655.
54. Elisaphenko E.A., Kolesnikov N.N., Shevchenko A.I., Rogozin I.B., Nesterova T.B., Brockdorff N., Zakian S.M. // PLoS ONE. 2008. V. 3. e2521.
55. Reik W., Lewis A. // Nat. Rev. Genet. 2005. V. 6. P. 403–410.
56. Zakharova I.S., Shevchenko A.I., Zakian S.M. // Chromosoma. 2009. V. 118. P. 279–290.
57. Carrel L., Willard H.F. // Nature. 2005. V. 434. P. 400–404.
58. Disteche C.M., Filippova G.N., Tsuchiya K.D. // Cytogenet. Genome Res. 2002. V. 99. P. 36–43.
59. Yen Z.C., Meyer I.M., Karalic S., Brown C.J. // Genomics. 2007. V. 90. P. 453–463.
60. Disteche C.M. // Trends Genet. 1995. V. 11. P. 17–22.
61. Brown C.J., Greally J.M. // Trends Genet. 2003. V. 19. P. 432–438.
62. Lyon M.F. // Cytogenet. Cell Genet. 1998. V. 80. P. 133–137.
63. Bailey J.A., Carrel L., Chakravarti A., Eichler E.E. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. P. 6634–6639.

64. Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D., Platzer M., Howell G.R., Burrows C., Bird C.P., Frankish A., Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G., Jones M.C., Hurles M.E., Andrews T.D., Scott C.E., Searle S., Ramser J., Whittaker A., Deadman R., Carter N.P., Hunt S.E., Chen R., Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L., Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A., Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhya S., Ashwell R.I., Babbage A.K., Bagguley C.L., Ballabio A., Banerjee R., Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M., Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N., Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D., Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W., Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C., Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C., Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S., Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S., Corby N., Connor R.E., David R., Davies J., Davis C., Davis J., Delgado O., Deshazo D., Dhami P., Ding Y., Dinh H., Dodsworth S., Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I., Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L., Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P., Gilbert J., Gill R., Glockner G., Gregory S.G., Gribble S., Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A., Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J., Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J., Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S., de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z., Kioschis P., Klages S., Knights A.J., Kosiura A., Kovar-Smith C., Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W., Lloyd C., Lloyd D.M., Loulseged H., Loveland J.E., Lovell J.D., Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H., McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T., Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Muller I., Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'Dell C.N., Okwuonu G., Palmer S., Pandian R., Parker D., Parrish J., Pasternak S., Patel D., Pearce A.V., Pearson D.M., Pelan S.E., Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S., Ridler K.A., Schlessinger D., Schueler M.G., Sehra H.K., Shaw-Smith C., Shen H., Sheridan E.M., Shownkeen R., Skuce C.D., Smith M.L., Sotheran E.C., Steingruber H.E., Steward C.A., Storey R., Swann R.M., Swarbreck D., Tabor P.E., Taudien S., Taylor T., Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S., Tromans A.C., d'Urso M., Verduzco D., Villasana D., Waldron L., Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A., Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L., Williams G., Williams L., Williamson A., Williamson H., Wilming L., Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H., Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A., Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F., Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A., Nelson D.L., Weinstock G., Sulston J.E., Durbin R., Hubbard T., Gibbs R.A., Beck S., Rogers J., Bentley D.R. // Nature. 2005. V. 434. P. 325–337.
65. Baverstock P.R., Adams M., Polkinghorne R.W., Gelder M. // Nature. 1982. V. 296. P. 763–766.
66. McQueen H.A., McBride D., Miele G., Bird A.P., Clinton M. // Curr. Biol. 2001. V. 11. P. 253–257.
67. Suzuki M.G., Shimada T., Kobayashi M. // Heredity. 1998. V. 81. P. 275–283.
68. Suzuki M.G., Shimada T., Kobayashi M. // Heredity. 1999. V. 82. P. 170–179.
69. Itoh Y., Melamed E., Yang X., Kampf K., Wang S., Yehya N., van Nas A., Replogle K., Band M.R., Clayton D.F., Schadt E.E., Lusis A.J., Arnold A.P. // J. Biol. 2007. V. 6. P. 2.
70. Ellegren H., Hultin-Rosenberg L., Brunstrom B., Dencker L., Kultima K., Scholz B. // BMC Biol. 2007. V. 5. P. 40.
71. Zha X., Xia Q., Duan J., Wang C., He N., Xiang Z. // Insect Biochem. Mol. Biol. 2009. V. 39. P. 315–321.
72. Melamed E., Arnold A.P. // Genome Biol. 2007. V. 8. P. R202.
73. Mank J.E. // Trends Genet. 2009. V. 25. P. 226–233.

The Role of p66shc in Oxidative Stress and Apoptosis

E. R. Galimov
Belozersky Institute of Physico-Chemical Biology, Moscow State University
E-mail: e.r.galimov@gmail.com
Received 03.11.2010

ABSTRACT ***p66shc*** **is a gene that regulates the level of reactive oxygen species (ROS), apoptosis induction, and lifespan in mammals. Mice knocked out for** ***p66shc*** **have a lifespan ~30% longer and demonstrate an enhanced resistance to oxidative stress and age-related pathologies such as hypercholesterolemia, ischemia, and hyperglycemia. In this respect, p66shc is a promising pharmacological target for the treatment of age-related diseases. In this review, an attempt has been made to survey and put to a critical analysis data concerning the involvement of p66shc in the different signaling pathways that regulate oxidative stress and apoptosis.**
KEYWORDS **apoptosis, reactive oxygen species, p66shc, mitochondria.**
ABBREVIATIONS **Akt – protein kinase B, Cdc42 (cell division control protein 42 homolog) – a small GTPase of the Rho-subfamily, Cyt C – cytochrome** *c*, **DMTU – dimethylthiourea, Grb2 – growth factor receptor-bound protein 2, ERK – extracellular signal-regulated kinases, HIV-1 – human immunodeficiency virus 1, IMS – intermembrane space of mitochondria, JNK – c-Jun N-terminal kinases, MAPK – mitogen-activated protein kinases, mHSP70 – mitochondrial 70 kilodalton heat shock proteins, MnSOD – mitochondrial superoxide dismutase, PKCβ – protein kinase Cβ, Pin-1 – peptidyl-prolyl** *cis/trans* **isomerase, PP2A – Protein phosphatase 2, Prx1 – peroxiredoxin 1, PTP – permeability transition pore, PTP-PEST – protein tyrosine phosphatase that contains a C-terminal PEST motif, Rac-1 – ras-related C3 botulinum toxin substrate 1, RAS – rat sarcoma viral oncogene, REF-1 – redox factor 1, ROS – reactive oxygen species, SOS1 (Son of Sevenless) – guanine nucleotide-binding protein, TIM – translocase of the inner membrane of mitochondria, TMPD – N,N,N',N'- tetramethyl-p-phenyldiamine, TOM – translocase of the outer membrane of mitochondria.**

INTRODUCTION

The identification of the mutations that lead to the prolongation of the lifespan of various model organisms shows that aging can be considered as a genetic program [1]. One of these genes is *p66shc*, the deletion of which results in a 30% increase in the lifespan. It is important to note that mice knocked out for *p66shc*, in comparison with other mouse models with a prolonged lifespan (e.g., mice with the deleted gene of a growth hormone receptor), are fertile and exhibit a normal phenotype [2]. These mice are resistant to oxidative stress and age-related pathologies such as atherosclerosis [3], endothelial disorders [4], AGE (advanced glycation end products)-dependent glomerulopathy related to diabetes mellitus [5, 6], and ethanol-induced liver affection [7].

P66shc is an adaptor protein which is coded for by a single locus in *Drosophila (dShc)*, and by four loci in mammals – *Shc* (*ShcA*), *Sli* (*ShcB*), *Rai* (*ShcC*) [8], and *RalP* [9]. The four mammalian loci code for at least 7 proteins due to the usage of alternative start codons and alternative splicing. Three isoforms encoded by the *ShcA* locus are designated according to their molecular weights as p46shc, p52shc, and p66shc, respectively. These proteins participate in the regulation of proliferation (p46shc and p52shc) and apoptosis (p66shc) [8].

P66shc is considered to be a relatively "young" protein since it is not found in yeast (*Saccharomyces*), nematodes (*Caenorhabditis*), and insects (*Drosophila*) but appears in amphibians (*Xenopus*), fishes (*Fugu rubripes*), and mammals [8]. Since it is the longest isoform, p66shc contains all the most ancient domains that are found in short isoforms p46shc and p52shc (Fig.1), including: the N-terminal phosphotyrosine-binding domain (PTB), the middle collagen homology domain (CH1), and the C-terminal Src-homology domain (SH2). The isoforms that are longer than p46shc contain some additional N-terminal domains: the cytochrome c binding domain (CB), which is common to both p52shc and p66shc, and the collagen-homology domain (CH2), which is unique to p66.

P46shc and p52shc fulfill the function of adaptor proteins transmitting signals from various tyrosine-kinase receptors, which phosphorylate tyrosine residues in these proteins. Phosphorylation of p46shc/p52shc induces the formation of a complex between GRb2

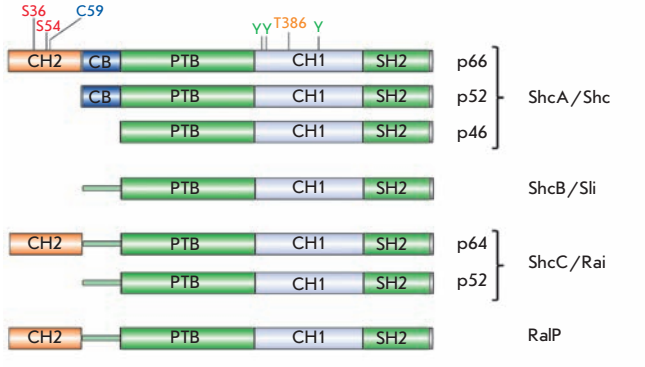


Fig. 1. Schematic presentation of domain structure of SHC-like proteins. Modified amino acid residues denoted as: serine 36 and 54 (S36 and S54), cysteine 59 (C59), threonine 386 (T386) and tyrosines (Y) phosphorylated during signal transduction from tyrosine kinase receptors.

(adaptor protein) and SOS (nucleotide exchange factor), which leads to RAS activation and the induction of the mitogene-activated protein kinases (MAPK) pathway. Although p66shc as p46shc/p52shc is phosphorylated by tyrosine-kinase receptors and interact with the GRb2/SOS complex, it, apparently, does not activate MAPK [10]. The competition between p66shc and p52shc for binding with GRb2 [11] and p66shc-induced displacement of SOS from its complex with GRb2 [12, 13] can mediate the regulation of signal transduction from receptors. However, experiments on animal models knocked out for *p66shc* showed that the role of this protein in lifespan regulation, oxidative stress, and apoptosis is accounted for by the CH2 and CB domains, which are not essential for the primary adaptor function of the protein.

P66, OXIDATIVE STRESS, AND APOPTOSIS.

Studies on mice with a knocked out *p66shc* gene revealed decreased levels of intracellular ROS, as determined by means of ROS-sensitive probes, as well as reduced levels of oxidative damages to DNA and proteins, estimated by measuring 8-oxo-deoxiguanosine and nitrotyrosines [3,4,14-16]. The mutant mice exhibited higher resistance to paraquat-induced oxidative stress [2]. Studies on various *p66shc*-deficient cell lines (ones with a deleted *p66shc*-gene, or cells with a dominant-negative phenotype caused by a Ser36Ala substitution in the target gene) derived from mice, rats, and a human showed that p66shc plays an important role in apoptosis induced by various agents (Table 1). P66shc-mediated oxidative stress is assumed to be the key factor in these experimental models of apoptosis. In particular, certain published data indicate the importance of p66shc-induced oxidative stress in the p53-dependent apoptotic pathway [14]. Data derived from physiological experiments have led to similar conclusions (Table 2).

Posttranslational modifications have been shown to play an important role in the pro-apoptotic activity of p66shc. For instance, phosphorylation at Ser36 located in the CH2-domain is indispensable in hydrogen peroxide- or UV-induced apoptosis [2]. Phosphorylation itself is carried out by such kinases as JNK (in response to UV or amyloid β-peptide) [23, 33], ERK [34], and PKCβ in response to treatment by H_2O_2 [35]. Phosphorylation at Ser36 promotes interaction with 14-3-3 proteins [36], tyrosine phosphatase PTP-PEST [37], and prolylisomerase Pin-1 [35]. While the significance of the first two interactions remains unclear, Pin-1-mediated p66shc isomerization plays a crucial role in the regulation of p66shc transport into mitochondria and activation of mitochondrion-dependent apoptosis (see below).

However, despite the great amount of research confirming the pro-apoptotic functions of p66shc, some studies suggest that it can also exhibit antiapoptotic effects. In a human breast cancer model, as well as in a human stem cell model, it was shown that the suppression of *p66shc* expression protects from hypoxia-induced cytotoxicity. It was also found that low oxygen concentrations lead to p66shc activation, which, in turn, induces expression of the *Notch-3* gene. The latter accounts for the self-renewal of stem cells and their survival under hypoxia. Notch-3 induces the expression of carboanhydrase IX, which also accounts for a hypoxia resistant phenotype [38]. Therefore, a connection exists between p66shc and the pathway that underlies the protection of stem cells from hypoxia. These data clarify the role of p66shc in homeostasis and the self-renewal of stem cells, which are known to occupy specialized tissue compartments or niches containing small amounts of blood vessels and, hence, low oxygen concentrations. It is interesting to note that since the small GTPase Rac-1 is a well-known activator of p66-induced oxidative stress [25], it also plays an essential part in the maintenance and self-renewal of epidermal stem cells. Taken together, these data show that p66shc has a more complex role, acting as a "double-edged sword" in the regulation of apoptosis, based on environmental conditions and genetic context.

MECHANISMS OF P66-DEPENDENT INCREASE IN CELLULAR ROS LEVELS

By now, we know the mechanisms p66shc exploits to increase intracellular ROS levels: activation of membrane-bound NADPH-oxidases, down-regulation of antioxidant enzymes synthesis, and generation of ROS in mitochondria.

Table 1. Cellular models of p66shc inactivation (via either *p66shc* gene deletion or expression of a dominant-negative mutant of p66shc at S36) with phenotype of apoptosis resistance

Cells	Cell line name	Organism	Apoptosis inducer	References
Embryonic fibroblasts	MEF	mouse	H_2O_2, UV, staurosporine, isothiocyanate, chloroform	[2, 15, 17, 18]
Primary cardiomyocyte		mouse	angiotensin II	[19]
Transformed renal epithelial cells	TKPTS	mouse	H_2O_2, cisplatin	[13]
Hepatocytes transgenic for human TGFa	AML12	mouse	hypoxia-reoxygenation	[20]
Endothelial progenitor cells	BM c-kit+	mouse	high glucose in media	[21]
Osteoblastic cells	OB-6, UAMS-32	mouse	H_2O_2	[22]
Pheochromocytoma	PC12	rat	beta amyloid, constitutively active Rac1 mutant	[23]
Cardiomyocyte	ARVM	rat	high glucose in media	[24]
Transformed fibroblast-like cells	COS7	green monkey	constitutively active Rac1 mutant	[25]
Neuroblastoma	SH-SY5Y	human	beta amyloid	[23]
Human podocytes immortalized with SV40-T-antigen	CIDHPs	human	HIV-1 transfection	[26]
Prostatic carcinoma	PC3, LNCaP	human	isothiocyanate	[18]
Cervical carcinoma	HeLa	human	H_2O_2	[27]
Osteosarcoma	SaOs-2	human	H_2O_2	[27]
Retinal pigmented epithelial Cells	RPE	human	H_2O_2	[28]
Lymphoma	Jurcat	human	hypoxia, calcium ionophores	[29]
Transformed renal epithelial cells	φNx-293	human	cell detachment from a solid matrix	[30]
Endothelial cells	HuVec	human	cell detachment from a solid matrix	[30]

p66-dependent Activation of Membrane-bound NADPH-oxidases

First of all, it should be noted that p66shc can cause oxidative stress in the cell carrying out its primary function of adaptor protein. As was mentioned above, this protein can negatively regulate RAS-activation by means of displacing the nucleotide exchange factor SOS from its complex with GRb2. It has turned out that this chain of events can be accompanied by the SOS-dependent activation of small GTPase Rac-1, which consequently promotes the assembly of membrane-bound NADPH-oxidases and the production of ROS [12] (Fig. 2). It has been shown that p66shc is necessary for apoptosis induced by a constitutively active Rac-1 mutant. However, this apoptosis scenario does not involve the phosphorylation of Ser36 but phosphorylation of Ser54 and Thr386, instead, which promotes Rac-1-dependent

Table 2. p66 involvement in the development of pathologies associated with oxidative stress and demonstrated on physiological models (experiments with *p66shc* knock-out animals or comparison between young and elderly individuals)

Pathology associated with apoptosis	Organism and genetic line	Reference
Experimental diabetic glomerulopathy	mouse, SV/129	[5]
Vascular cell apoptosis and atherogenesis induced by high-fat diet	mouse, SV/129	[3]
Cardiomyocyte apoptosis in experimental model of diabetes induced by streptozotocin	mouse, SV/129	[31]
Cerebral cortex hypoxia	rat, Sprague–Dawley	[32]

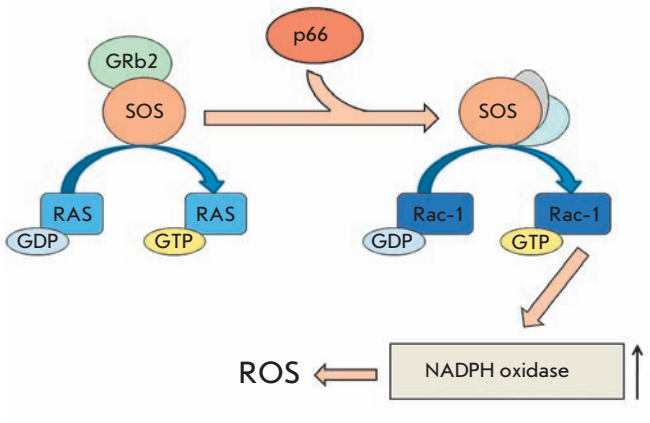


Fig. 2. P66-dependent activation of plasma membrane NADPH oxidases. P66shc displaces nucleotide exchange factor SOS from Grb2 complex and promotes activation of small GTPase Rac1. Activated Rac1 stimulates NADPH oxidase complex formation and ROS generation.

stabilization of p66shc and protects against ubiquitin-mediated degradation [25]. In macrophages, where NADPH-oxidase serves as the major source of ROS, knockout of the *p66shc* gene undermines the formation of the active NADPH-oxidase complex, ultimately leading to a 40% decrease in ROS-formation.

p66shc and Regulation of Expression of Antioxidant Enzymes

Previous research has shown that p66shc reduces the expression of such antioxidant enzymes and regulatory factors as glutathione peroxidase-1 [28], MnSOD [7, 20, 28], and REF-1 [20] by means of down-regulation of Forkhead-type transcription factors (e.g., Foxo3a) [23, 40, 41]. In the course of oxidative stress, serine/threonine protein kinase Akt undergoes phosphorylation and in turn phosphorylates and inactivates Foxo3a. This reaction requires the presence of p66shc in the cell [40] and also its phosphorylation at Ser36 [42]. Outside of that, some data suggest that p66shc, in complex with βPix (a nucleotide exchange factor for Rac-1 and Cdc42), can cause Akt-independent phosphorylation and the inactivation of Foxo3a [43] (Fig. 3).

It should be mentioned that these data are in disagreement with a number of publications which argue that p66shc does not affect the levels of antioxidant enzymes [4, 15-17].

p66shc and Mitochondria-mediated Apoptosis

The fact that cells carrying deletion of the *p66shc* gene are resistant to various inductors of mitochondria-mediated apoptosis implies a direct interaction between p66 and mitochondria.

Mitochondrial localization of p66shc and its transport to mitochondria. Studies on the cellular localization of p66shc showed that 32% of this protein is localized in the cytoplasm; 24% is in the endoplasmic reticulum; and 44%, in mitochondria [17]. Inside the mitochondria, p66shc is distributed in the following manner: 35% is in the intramembrane space, 56% is associated with the inner membrane, and 9% is located in the mitochondrial matrix [15]. According to other data, mitochondria contain only 10% of cellular p66shc [44]. These differences told arise from the changes in p66shc intracellular localization caused by external influences.

The stimuli that promote p66shc translocation into mitochondria are usually pro-apoptotic factors, such as UV radiation and treatment with H_2O_2 [17, 35]. However, the mechanism of p66shc transport into mitochondria remains unknown. It is known that a short Shc isoform p46shc is also localized in mitochondria and contains the signal of mitochondrial import [45]. However, mutations in a similar mitochondrial import sequence in p66 did not affect its localization - which probably means that this signal is somehow masked by the N-terminal CH2-domain [44]. It was also shown that p66shc is associated with protein complexes that contain mHSP70, TIM, and TOM subunits and mediate protein transport into mitochondria. P66 is thought to be inactive in these complexes; however, it is believed that under oxidative stress p66 dissociates and thus acquires active conformation [46].

According to data presented in [35], the signal pathway that initiates p66shc translocation into mitochondria upon H_2O_2 treatment includes the activation of PKCβ, which phosphorylates p66 at Ser36. Phosphorylated p66 becomes a target of prolylpolymerase Pin-1 that recognizes a proline residue, which follows phosphorylated serine. After isomerization, p66shc is dephosphorylated by PP2A phosphatase and transported into mitochondria. The latter event is confirmed by the

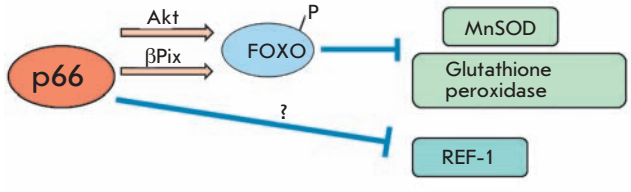


Fig. 3. P66 role in regulation of the cellular antioxidant defense system. P66 can downregulate antioxidative ferments and regulatory factors. This downregulation can be induced through both Akt-dependent and Akt-independent inactivation of Forkhead transcription factors.

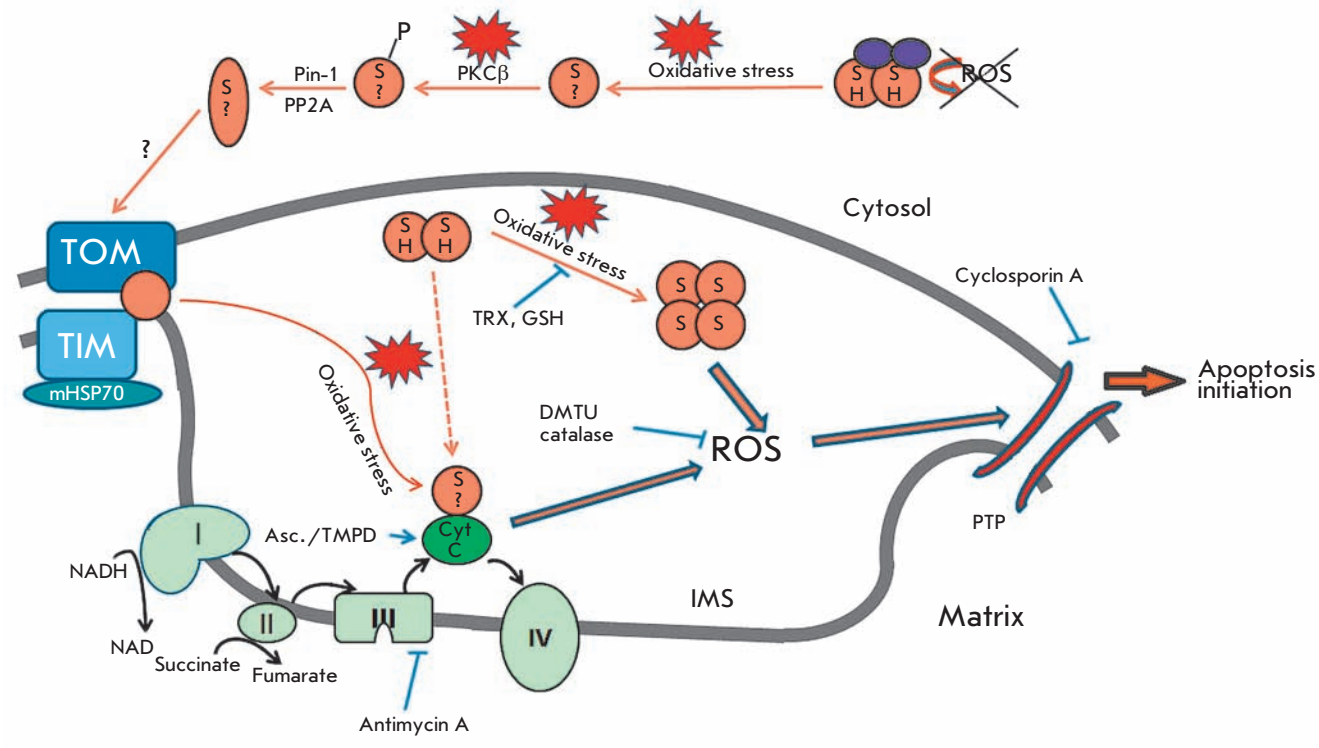


Fig. 4. Model for induction proapoptotic signaling in mitochondria. p66shc (denoted as orange circle) can exist as dimer (circles where "SH" indicates reduced cysteine residue 59) and tetramer (circles with "S" indicating an oxidized cysteine residue 59). Circles with "S?" indicating unknown redox state of a cysteine residue 59 that is responsible for tetramerisation p66. Interaction of p66shc with peroxiredoxin 1 (Prx1, denoted as violet oval) in cytosol results in a complex presumably consisting of dimeric p66 and dimeric Prx1 in which Prx1 peroxidase activity degrades p66-generated ROS. Oxidative stress leads to dissociation of the complex; PKCβ phosphorylates released p66 on S36. Phosphorylated p66 becomes a target for prolyl isomerase Pin1, which recognizes proline following phosphorylated serine residue. After izomerization PP2A dephosphorylates p66shc, which is then transported into mitochondria. During oxidative stress p66shc is released from the high-molecular-mass complex that contains TOM, TIM, and mHsp70. Released p66shc acts as oxidoreductase and transfers electrons from reduced Cyt C to oxygen. As a result, generated ROS lead to permeability-transition pore (PTP) opening and induction of apoptosis. These consequences may appear after tetrameric p66 formation during oxidative stress and following depletion of the thoredoxin and glutathione pool (for full explanation, see text). Asc. – ascorbate; DMTU – dimethylthiourea; GSH – glutathione; IMS – intermembrane space of mitochondria; PTP – permeability transition pore; TMPD – N,N,N′,N′- tetramethyl-p-phenyldiamine; TRX – thioredoxin.

fact that the pool of mitochondrial p66shc is not phosphorylated [15] (Fig.4). It was also shown in [35] that the absence of p66shc also results in altered Ca-signaling and prevents the fragmentation of mitochondria, which is related to resistance to apoptosis.

Redox proapoptotic activity of p66shc and the production of ROS. Current views hold that, in the course of mitochondrial-dependent apoptosis, various signals (ROS, elevated Ca, uncoupled oxidative phosphorylation) promote the formation of a multi-subunit protein complex, which eventually forms a pore in the inner mitochondrial membrane. The permeabilization of both mitochondrial membranes results in the release of Cyt C and other proteins in the cytoplasm, apoptosome formation, and caspase activation [47].

Pelicci *et al.* have suggested a mechanism explaining the role of p66shc in mitochondrium-dependent apoptosis [15]. It turns out that p66shc is necessary for a drop in the membrane's potential and Cyt C release in cytosol. Moreover, the addition of cyclosporine A, an inhibitor of permeability transition pore formation, blocked the pro-apoptotic function of p66shc [17]. *In vitro* showed that the addition of recombinant p66shc to isolated mitochondria with a permeabilized outer membrane results in mitochondria swelling as a result of permeability transition pore formation. This effect was also inhibited by cyclosporine A, catalase, antioxi-

dant dimethylthiourea, and the inhibitor of complex III of the respiratory chain antimycin A, as well as being dependent on respiratory substrates. Therefore, the obtained data is evidence that permeability-transition pore formation is a key step in p66-mediated apoptosis and that ROS and respiration are also necessary in this process.

The application of electrochemical analysis and fluorescent redox-sensitive probes revealed that p66 acts as a redox enzyme and transfers electrons from reduced Cyt C to oxygen. The interaction of p66 with Cyt C is mediated by the CB-domain, which was confirmed by ELISA and site-directed mutagenesis [17]. Incomplete oxygen reduction leads to ROS production, which in turn promotes the formation of a permeability-transition pore. In confirmation of this hypothesis, ROS formation was shown upon the addition of p66 to mitochondria in the absence of substrates of respiration in a medium supplemented with ascorbate/TMPD., a redox couple that selectively reduces Cyt C. ROS formation was also observed even in the absence of mitochondria upon mixing of Cyt C and p66shc. However, in this case, the presence of copper ions was indispensable.

Therefore, the Pelicci group suggested the following mechanism of pro-apoptotic action of p66shc (Fig. 4): under normal conditions, p66shc is inactivated and represents a part of the multi-subunit complex comprising TIM, TOM, and mHSP70. Oxidative stress causes p66shc to dissociate from the complex. As a result, p66shc acts as a redox enzyme and transfers electrons from the reduced Cyt C to oxygen. The incomplete reduction of oxygen results in ROS production, which promotes the formation of a permeability-transition pore, mitochondrial swelling, Cyt C release to cytosol, apoptosome assembly, and caspase activation.

It is important to realize that the suggested mechanism of p66shc redox activity, which includes Cyt C as an electron donor, is only a reality *in vitro*. In order to validate this hypothesis, one needs to carry out additional experiments based on fluorescent probes for measuring ROS levels in the mitochondria of cells with an inactive respiratory chain (Rho-0 cells), as well as in cells devoid of *Cyt C*.

It should be noted that the described model has many weak points. For example, p66shc does not contain any well-known redox domains or metal-binding domains. Therefore, its ability to generate ROS only in the presence of copper ions can be an artifact, since copper ions are known to be able to generate ROS during Fenton's reaction, which could affect the results obtained by means of redox-sensitive fluorescent probes. It is also known that Cyt C, upon oligomerization, can produce ROS as a result of auto-oxidation [48-52]. Taking this into account, one can suggest that p66shc, upon binding with Cyt C, can act as a factor promoting oligomerization and the auto-oxidation of Cyt C, similarly to ptothymosine α, but not as a redox enzyme [53].

Studies on an isolated CH2-CB-domain. The mechanism of action of p66shc was also studied in *in vitro* experiments with an isolated CH2-CB-domain. It turns out that the recombinant CH2-CB-domain can exist in two forms: reduced, as a dimer, and oxidized, as a tetramer (or a dimer of dimers) which contains disulfide bridges between residues Cys59. The addition of both the CH2-CB-domain and the full-length protein to mitochondria resulted in the formation of a permeability transition pore and swelling; however, only the tetramer exhibited such a pro-apoptotic activity. In contradiction to the data presented in [15], which suggested that ROS production is directly connected to permeability-transition pore formation, the tetramer form of the recombinant CH2-CB-domain generated less ROS than the dimer when mixed with isolated mitochondria.

Apart from that, the CH2-CB-domain was shown to trigger ROS production in the presence of dithionite (as an electron donor) and copper ions. However, this effect was not reproduced when dithionite was replaced with Cyt C as an electron donor. Apparently, the other domains of p66shc are necessary either for interaction with Cyt C or for maintenance of the CH2-CB-domain in proper conformation.

As a result of the above-described findings, a refined model of p66-mediated apoptosis was suggested. The model assumes that under normal conditions the tetrameric form of p66 is reduced by mitochondrial antioxidant systems. Under stress conditions, however, antioxidant systems are not able to retain p66shc in their reduced dimeric state, so the oxidized tetrameric forms generate ROS locally, which triggers permeability-transition pore formation and apoptosis.

It is important to note that a direct connection between the ROS-generating and pro-apoptotic functions of p66shc does not necessarily exist. For example, a confirmation of this fact can be the lower level of ROS production in the case of pro-apoptotic tetrameric forms of p66shc in comparison to the dimer, which does not induce apoptosis.

p66 as a redox sensor. Extensive experiments on the isolated CH2-CB-domain revealed a new interaction partner of p66shc – peroxiredoxine 1 (Prx1). Prx1 is a member of the peroxidase family. These enzymes regulate the redox balance in the cell. Prx1 can exist in the form of a dimer, a decamer (5 dimers) or a multimer. Prx1 is basically localized in cytosol, though recent proteomic studies have revealed its presence in

the perimitochondrial compartment. Under normal conditions, Prx1 exists predominantly in dimeric form, which functions as peroxidase. When the cell is under stress, Prx1 undergoes oxidation and forms decamers that possess lower peroxidase activity. However, upon these transformations Prx1 acquires a chaperone function. Severe oxidative stress leads to the formation of the multimer, which is also devoid of peroxidase activity but can function as a chaperone [56-58].

Interactions between the CH2-CB-domain of p66shc and Prx1 result in the destabilization of the decameric form of Prx1 and favor dimeric form transition. In turn, Prx1, in complex with p66shc, retains p66shc in dimeric form by means of a disulfide exchange between the two proteins. The resulting hybrid complex consists of dimeric p66shc, which generates ROS but has low pro-apoptotic activity, and dimeric Prx1, which functions as a peroxidase. Thus, under normal conditions dimeric Prx1 stabilizes the inactive state of p66shc and degrades ROS which are produced by p66shc. Under stress conditions, cystein residues of proteins are oxidized, resulting in disassembly of the complex: so, the uncomplexed p66shc can now participate in apoptosis induction [55]. Therefore, the complex of dimeric p66shc and dimeric Prx1 can be considered as a sensor that detects the level of cellular ROS and induced apoptosis when excessive amounts of ROS are accumulated (Fig. 4).

CONCLUSION

According to current views, p66shc is a pro-apoptotic protein which regulates oxidative stress and induces the mitochondrial apoptosis pathway by means of redox activity. Although the physiological role of p66shc has been extensively studied, little is known about the precise mechanism of action underlying its redox functions and participation in apoptosis induction. Further studies will endeavor to elucidate the precise mechanism of p66shc translocation into mitochondria and the localization of p66shc-dependent ROS production in the cell. Although p66shc phosphorylation at Ser36 is one of the indispensible steps of apoptosis, it was shown that the mitochondrial pool of p66 is not phosphorylated. This indicates that p66shc phosphorylated at Ser36 probably participates in pro-apoptotic events outside mitochondria.

P66shc is undoubtedly interesting in terms of the investigation of oxidative stress, apoptosis, and the aging related to it. It is important to note that, to estimate the involvement of p66shc in age-related disorders, it is necessary to better understand its role in cancer. A deeper understanding of the regulatory pathways and structural and mechanistic bases of p66shc redox activity can be important for developing pharmacological approaches to the treatment of age-related disorders. ●

The author is grateful for the valuable comments and help provided in the preparation of the manuscript by B.V. Chernyak and V.P. Skulachev.

REFERENCES

1. Longo V.D., Mitteldorf J., Skulachev V.P. // Nat. Rev. Genet. 2005. V. 6. № 11. P. 866–872.
2. Migliaccio E., Giorgio M., Mele S., Pelicci G., Reboldi P., Pandolfi P.P., Lanfrancone L., Pelicci P.G. // Nature. 1999. V. 402. № 6759. P. 309–313.
3. Napoli C., Martin-Padura I., de Nigris F., Giorgio M., Mansueto G., Somma P., Condorelli M., Sica G., De Rosa G., Pelicci P. // Proc. Natl. Acad. Sci. USA. 2003. V. 100. № 4. P. 2112–2116.
4. Camici G.G., Schiavoni M., Francia P., Bachschmid M., Martin-Padura I., Hersberger M., Tanner F.C., Pelicci P., Volpe M., Anversa P., Luscher T.F., Cosentino F. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. № 12. P. 5217–5222.
5. Menini S., Amadio L., Oddi G., Ricci C., Pesce C., Pugliese F., Giorgio M., Migliaccio E., Pelicci P., Iacobini C., Pugliese G. // Diabetes. 2006. V. 55. № 6. P. 1642–1650.
6. Menini S., Iacobini C., Ricci C., Oddi G., Pesce C., Pugliese F., Block K., Abboud H.E., Giorgio M., Migliaccio E., Pelicci P.G., Pugliese G. // Diabetologia. 2007. V. 50. № 9. P. 1997–2007.
7. Koch O.R., Fusco S., Ranieri S.C., Maulucci G., Palozza P., Larocca L.M., Cravero A.A., Farre S.M., De Spirito M., Galeotti T., Pani G. // Lab. Invest. 2008. V. 88. № 7. P. 750–760.
8. Luzi L., Confalonieri S., Di Fiore P.P., Pelicci P.G. // Curr. Opin. Genet. Dev. 2000. V. 10. № 6. P. 668–674.
9. Fagiani E., Giardina G., Luzi L., Cesaroni M., Quarto M., Capra M., Germano G., Bono M., Capillo M., Pelicci P., Lanfrancone L. // Cancer Res. 2007. V. 67. № 7. P. 3064–3073.
10. Migliaccio E., Mele S., Salcini A.E., Pelicci G., Lai K.M., Superti-Furga G., Pawson T., Di Fiore P.P., Lanfrancone L., Pelicci P.G. // EMBO J. 1997. V. 16. № 4. P. 706–716.
11. Okada S., Kao A.W., Ceresa B.P., Blaikie P., Margolis B., Pessin J.E. // J. Biol. Chem. 1997. V. 272. № 44. P. 28042–28049.
12. Khanday F.A., Santhanam L., Kasuno K., Yamamori T., Naqvi A., Dericco J., Bugayenko A., Mattagajasingh I., Disanza A., Scita G., Irani K. // J. Cell Biol. 2006. V. 172. № 6. P. 817–822.
13. Arany I., Faisal A., Nagamine Y., Safirstein R.L. // J. Biol. Chem. 2008. V. 283. № 10. P. 6110–6117.
14. Trinei M., Giorgio M., Cicalese A., Barozzi S., Ventura A., Migliaccio E., Milia E., Padura I.M., Raker V.A., Maccarana M., Petronilli V., Minucci S., Bernardi P., Lanfrancone L., Pelicci P.G. // Oncogene. 2002. V. 21. № 24. P. 3872–3878.
15. Giorgio M., Migliaccio E., Orsini F., Paolucci D., Moroni M., Contursi C., Pelliccia G., Luzi L., Minucci S., Marcaccio M., Pinton P., Rizzuto R., Bernardi P., Paolucci F., Pelicci P.G. // Cell. 2005. V. 122. № 2. P. 221–233.

16. Francia P., delli Gatti C., Bachschmid M., Martin-Padura I., Savoia C., Migliaccio E., Pelicci P.G., Schiavoni M., Luscher T.F., Volpe M., Cosentino F. // Circulation. 2004. V. 110. № 18. P. 2889–2895.
17. Orsini F., Migliaccio E., Moroni M., Contursi C., Raker V.A., Piccini D., Martin-Padura I., Pelliccia G., Trinei M., Bono M., Puri C., Tacchetti C., Ferrini M., Mannucci R., Nicoletti I., Lanfrancone L., Giorgio M., Pelicci P.G. // J. Biol. Chem. 2004. V. 279. № 24. P. 25689–25695.
18. Xiao D., Singh S.V. // Cancer Res. 2010. V. 70. № 8. P. 3150–3158.
19. Graiani G., Lagrasta C., Migliaccio E., Spillmann F., Meloni M., Madeddu P., Quaini F., Padura I.M., Lanfrancone L., Pelicci P., Emanueli C. // Hypertension. 2005. V. 46. № 2. P. 433–440.
20. Haga S., Terui K., Fukai M., Oikawa Y., Irani K., Furukawa H., Todo S., Ozaki M. // J. Hepatol. 2008. V. 48. № 3. P. 422–432.
21. Di Stefano V., Cencioni C., Zaccagnini G., Magenta A., Capogrossi M.C., Martelli F. // Cardiovasc. Res. 2009. V. 82. № 3. P. 421–429.
22. Almeida M., Han L., Ambrogini E., Bartell S.M., Manolagas S.C. // Mol. Endocrinol. 2010. V. 24. № 10. P. 2030–2037.
23. Smith W.W., Norton D.D., Gorospe M., Jiang H., Nemoto S., Holbrook N.J., Finkel T., Kusiak J.W. // J. Cell Biol. 2005. V. 169. № 2. P. 331–339.
24. Malhotra A., Vashistha H., Yadav V.S., Dube M.G., Kalra S.P., Abdellatif M., Meggs L.G. // Am. J. Physiol. Heart Circ. Physiol. 2009. V. 296. № 2. P. H380–388.
25. Khanday F.A., Yamamori T., Mattagajasingh I., Zhang Z., Bugayenko A., Naqvi A., Santhanam L., Nabi N., Kasuno K., Day B.W., Irani K. // Mol. Biol. Cell. 2006. V. 17. № 1. P. 122–129.
26. Husain M., Meggs L.G., Vashistha H., Simoes S., Griffiths K.O., Kumar D., Mikulak J., Mathieson P.W., Saleem M.A., Del Valle L., Pina-Oviedo S., Wang J.Y., Seshan S.V., Malhotra A., Reiss K., Singhal P.C. // J. Biol. Chem. 2009. V. 284. № 24. P. 16648–16658.
27. Tiberi L., Faisal A., Rossi M., Di Tella L., Franceschi C., Salvioli S. // Biochem. Biophys. Res. Commun. 2006. V. 342. № 2. P. 503–508.
28. Wu Z., Rogers B., Kachi S., Hackett S.F., Sick A., Campochiaro P.A. // J. Cell. Physiol. 2006. V. 209. № 3. P. 996–1005.
29. Carraro F., Pucci A., Pellegrini M., Pelicci P.G., Baldari C.T., Naldini A. // J. Cell Physiol. 2007. V. 211. № 2. P. 439–447.
30. Ma Z., Myers D.P., Wu R.F., Nwariaku F.E., Terada L.S. // J. Cell Biol. 2007. V. 179. № 1. P. 23–31.
31. Rota M., LeCapitaine N., Hosoda T., Boni A., De Angelis A., Padin-Iruegas M.E., Esposito G., Vitale S., Urbanek K., Casarsa C., Giorgio M., Luscher T.F., Pelicci P.G., Anversa P., Leri A., Kajstura J. // Circ. Res. 2006. V. 99. № 1. P. 42–52.
32. Rapino C., Bianchi G., Di Giulio C., Centurione L., Cacchio M., Antonucci A., Cataldi A. // Aging Cell. 2005. V. 4. № 4. P. 177–185.
33. Le S., Connors T.J., Maroney A.C. // J. Biol. Chem. 2001. V. 276. № 51. P. 48332–48336.
34. Hu Y., Wang X., Zeng L., Cai D.Y., Sabapathy K., Goff S.P., Firpo E.J., Li B. // Mol. Biol. Cell. 2005. V. 16. № 8. P. 3705–3718.
35. Pinton P., Rimessi A., Marchi S., Orsini F., Migliaccio E., Giorgio M., Contursi C., Minucci S., Mantovani F., Wieckowski M.R., Del Sal G., Pelicci P.G., Rizzuto R. // Science. 2007. V. 315. № 5812. P. 659–663.
36. Foschi M., Franchi F., Han J., La Villa G., Sorokin A. // J. Biol. Chem. 2001. V. 276. № 28. P. 26640–26647.
37. Faisal A., el-Shemerly M., Hess D., Nagamine Y. // J. Biol. Chem. 2002. V. 277. № 33. P. 30144–30152.
38. Sansone P., Storci G., Giovannini C., Pandolfi S., Pianetti S., Taffurelli M., Santini D., Ceccarelli C., Chieco P., Bonafe M. // Stem Cells. 2007. V. 25. № 3. P. 807–815.
39. Tomilov A.A., Bicocca V., Schoenfeld R.A., Giorgio M., Migliaccio E., Ramsey J.J., Hagopian K., Pelicci P.G., Cortopassi G.A. // J. Biol. Chem. 2010. V. 285. № 2. P. 1153–1165.
40. Nemoto S., Finkel T. // Science. 2002. V. 295. № 5564. P. 2450–2452.
41. Berniakovich I., Trinei M., Stendardo M., Migliaccio E., Minucci S., Bernardi P., Pelicci P.G., Giorgio M. // J. Biol. Chem. 2008. V. 283. № 49. P. 34283–34293.
42. Lebiedzinska M., Karkucinska-Wieckowska A., Giorgi C., Karczmarewicz E., Pronicka E., Pinton P., Duszynski J., Pronicki M., Wieckowski M.R. // Biochim. Biophys. Acta. 2010. V. 1797. № 6–7. P. 952–960.
43. Chahdi A., Sorokin A. // Mol. Biol. Cell. 2008. V. 19. № 6. P. 2609–2619.
44. Nemoto S., Combs C.A., French S., Ahn B.H., Fergusson M.M., Balaban R.S., Finkel T. // J. Biol. Chem. 2006. V. 281. № 15. P. 10555–10560.
45. Ventura A., Maccarana M., Raker V.A., Pelicci P.G. // J. Biol. Chem. 2004. V. 279. № 3. P. 2299–2306.
46. Orsini F., Moroni M., Contursi C., Yano M., Pelicci P., Giorgio M., Migliaccio E. // Biol. Chem. 2006. V. 387. № 10–11. P. 1405–1410.
47. Green D.R., Kroemer G. // Science. 2004. V. 305. № 5684. P. 626–629.
48. Margoliash E., Lustgarten J. // J. Biol. Chem. 1962. V. 237. P. 3397–3405.
49. Schejter A., Glauser S.C., George P., Margoliash E. // Biochim. Biophys. Acta. 1963. V. 73. P. 641–643.
50. Person P., Fine A.S., Mora P.T., Zipper H. // J. Biol. Chem. 1965. V. 240. P. 3159–3164.
51. Margoliash E., Schejter A. // Adv. Protein Chem. 1966. V. 21. P. 113–286.
52. Petersen L.C., Cox R.P. // Biochem. J. 1980. V. 192. № 2. P. 687–693.
53. Markova O.V., Evstafieva A.G., Mansurova S.E., Moussine S.S., Palamarchuk L.A., Pereverzev M.O., Vartapetian A.B., Skulachev V.P. // Biochim. Biophys. Acta. 2003. V. 1557. № 1–3. P. 109–117.
54. Gertz M., Fischer F., Wolters D., Steegborn C. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 15. P. 5705–5709.
55. Gertz M., Fischer F., Leipelt M., Wolters D., Steegborn C. // Aging (Albany NY). 2009. V. 1. № 2. P. 254–265.
56. Jang H.H., Lee K.O., Chi Y.H., Jung B.G., Park S.K., Park J.H., Lee J.R., Lee S.S., Moon J.C., Yun J.W., Choi Y.O., Kim W.Y., Kang J.S., Cheong G.W., Yun D.J., Rhee S.G., Cho M.J., Lee S.Y. // Cell. 2004. V. 117. № 5. P. 625–635.
57. Fourquet S., Huang M.E., D'Autreaux B., Toledano M.B. // Antioxid. Redox Signal. 2008. V. 10. № 9. P. 1565–1576.
58. Matsumura T., Okamoto K., Iwahara S., Hori H., Takahashi Y., Nishino T., Abe Y. // J. Biol. Chem. 2008. V. 283. № 1. P. 284–293.

Chromatin Diminution Process Regulates rRNA Gene Copy Number in Freshwater Copepods

M. V. Zagoskin[1*], T. L. Marshak[2], D. V. Mukha[1], A. K. Grishanin[1]
[1] Vavilov Institute of General Genetics, Russian Academy of Sciences
[2] Koltsov Institute of Developmental Biology, Russian Academy of Sciences
*E-mail: zagoskin_mv@mail.ru
Received 20.09.2010

ABSTRACT **The results of quantitative PCR (qPCR) presented in the paper clearly demonstrate that the sixteen-fold genome reduction in *Cyclops kolensis* during chromatin diminution (from 15.3 pg to 0.98 pg) results in a dramatic decrease in ribosomal RNA gene copy numbers in the genome of a somatic cell line by more than two orders of magnitude. The results presented allow for the consideration of the chromatin diminution as a mechanism of rDNA copy number regulation.**
KEYWORDS **Chromatin diminution, ribosomal RNA genes, Copepoda, real-time PCR, gene copy number**

INTRODUCTION

Chromatin diminution (CD) is the programmed process of elimination of a considerable fraction of chromatin from the genome of somatic precursor cells and occurs during early developmental stages of some multicellular eukaryotes, or from the somatic nucleus (the macronucleus) during its formation in protozoa. Knowledge of the phenomenon of chromatin diminution has existed for over a hundred years. Yet, it has been established to exist only in less than 100 belonging to only a few taxa: including protozoa, nematodes, and copepods [1, 2]. CD has been described in approximately 20 cyclop species [1].

The rRNA genes in the genome of most eukaryote species are represented by a large copy number and organized as cistrons. Copies of the ribosomal cistron are repeated in tandem form in one or several clusters, which may be located on one or several chromosomes [3]. A large amount of data, mainly concerning the structural-functional organization of rRNA genes, has been accumulated over the past decades [4–6]. Recent studies have been directed primarily towards the investigation of the mechanisms controlling the regulation of the transcriptional activity of rRNA genes. It should be noted that rRNA molecules represent more than half of all RNA synthesized in a cell [6], and rRNA genes are responsible for approximately 35–60% of the total transcriptional activity in a cell [7].

It is known that the copy number of rRNA genes in the eucaryotic genome varies over an appreciably wide range: from 39 to 19,300 in animals and from 150 to 26,048 in plants [8]. A number of cases have been described with a variation of the rDNA copy number, including an increase in the amount of rDNA due to the amplification of extrachromosomal rRNA gene copies in *Xenopus laevis* oocytes [9–11] or in protozoa [12]. The rRNA gene number may also decrease, as occurs in *Drosophila melanogaster* with the so-called *bobbed (bb)* mutation; however, the number of rDNA repetitions completely recovers by the third–fourth generations [13–18]. Although this variation in the rDNA copy number in the genome is likely an exception, it suggests the existence and maintenance of mechanisms for regulation of the rRNA gene copy number at a certain level.

The programmed gene elimination during onthogenesis results from CD as well, although this has been detected so far only in two nematode species: *Ascaris lumbricoides* and *A. suum*. In *A. lumbricoides*, the gene encoding the ALEP-1 ribosomal protein is cleaved [19], whereas the three unique genes (*rpS19G*, *fert-1*, *aleg-3)* and a retransposon (*Tas)* [20–22]) that are eliminated by CD have been detected thus far only in *A. suum*. The remaining portion of the eliminated sequences being eliminated are noncoding. It has been repeatedly noted that CD could be an informative model for studying the excessive DNA problem and genome reorganization during ontogenesis. However, the significance of CD has been underappreciated because of the lack of data [23, 24].

Despite the fact that 94% of DNA is eliminated in *Cyclops kolensis* during CD [25], until recently, only either noncoding nucleotide sequences enriched in repeat regions or satellite DNA have been detected in the fraction eliminated from the genome [26–28]. Earlier, it was assumed that there is a possibility of elimination of a fraction of ribosmal cistrons from *C. kolensis* chromosomes as a result of CD [29]. Prokopovich *et al.* have noted the positive correlation between the genome's size and the rDNA copy number [8].

This study was aimed at ascertaining the ratio of rRNA gene copy number in prediminuted to post-

diminuted *C. kolensis* genomes and estimating the rRNA gene copy number in the Russian population of *C. insignis*, the species which lacks CD [30]. For this purpose, we used quantitative PCR (qPCR), along with other techniques for determining gene copy number. For the present purposes, qPCR is the most efficient and informative method [31–33].

EXPERIMENTAL

Cyclops *C. kolensis* Lill. and *C. insignis* Claus were collected from a pond in the Vorob'evy Gory, Moscow, Russia (55°42'35.40"N; 37°34'6.61"W) in April 2009–2010.

Cytophotometry

The procedure [25, 34] for the preparation of specimens of embryonic and somatic cells of adult *C. kolensis,* in order to carry out Feulgen staining, was subjected to modifications in which destruction of the shells of embryo sac and egg was performed. Sperm cells and erythrocytes of loach were used as an internal control for determining the absolute amount of DNA. The DNA content in loach (*Misgurnus fossilis*) cells is 1C = 2.4 pg [35]. Loach sperm cells contain 1C = 2.4 ± 0.2 (SD) pg of DNA; for erythrocytes, 2C = 5.1 ± 0.4 (SD) pg. The DNA content was measured in 100 erythrocytes and 127 sperm cells of loach. The procedure of sperm smear preparation included the trituration of the male gonads of the loach in the standard physiological solution – 0,9% NaCl; the resulting cell suspension was used to prepare the smears. After drying, the specimens were immobilized for 10 minutes in 96% ethanol, at room temperature. Cyclops were fixed in an ethanol–acetic acid mixture (3 : 1), squashed in 45% acetic acid to completely separate the shells of the embryo sac and embryo itself. Next, they were subjected to the Feulgen staining procedure in order to measure the amount of DNA. Feulgen staining of the cell nuclei was carried out under the following conditions: hydrolysis in 5 N HCl for 11 min at 37°C and staining with Schiff's reagent (1 hr at room temperature). The measurements were carried out on a Vickers M86 microdensitometer (England) (wavelength of 540 nm). All specimens used for the measurements were processed in the same staining batch. Fifty-nine cells of the second polar body were analyzed at the metaphase state in the pre-diminution genome and 140 somatic line cells of an adult cyclops after CD. The results were processed using Microsoft Excel 2007 software (the descriptive statistics).

Data collection and DNA isolation

The *C. kolensis* embryo sacs with embryos at stages of four to eight cells were taken as the pre-diminution material. The selection of the embryos was carried out according to the aforementioned procedure [25], which allowed *in vivo* determination of the cell division stage of cyclops embryos using a light microscope. The *C. kolensis* antennae, comprising somatic cells only, served as the post-diminution material. Two antennae were severed with a scalpel from each individual and placed on a glass slide located on a liquid-nitrogen cooled table. The embryo sacs were immobilized in liquid nitrogen. The *C. insignis* individuals were selected because no CD was observed in their ontogenesis.

Since there was a very small amount of DNA material available, genomic DNA was isolated using the Diatom™ DNA Prep 100 reagent kit (Izogen, Moscow). This kit makes it possible to minimize DNA loss during isolating. Its operating principle is based on the lysis of a specimen in guanidine thiocyanate (a strong chaotropic agent) followed by DNA sorption on silica gel. The material was prehomogenized in a buffer solution (0.2 M Tris, 50 mM EDTA, 0.5% SDS, 200 μg/ml of proteinase K) and lyzed for 1 h at 50°C. The lysate was treated with RNase (0.1 mg/ml) for 5 min; DNA was then isolated according to the procedure recommended by the manufacturer.

Real-time PCR using EVA Green dye

Concentrations of the total *C. kolensis* DNA before and after CD and the total *C. insignis* DNA were determined on a Nanodrop 1000 spectrophotometer. The coefficient of variation of DNA concentration was calculated on the basis of two replicates and was equal to 1.33% (before CD) and 5.91% (after CD) in *C. kolensis* and 9.18% in *C. insignis*.

For carrying out real-time PCR, specific primers were constructed (28real_for – 5'-GGTAGCCAAATGCCTCGTC-3' and 28real_rev– 5'-CGCCAAAGATGCTCCGCCAC-3'), which allowed the amplification of the 183 bp fragment of the 28S rRNA gene. The fragment length of the 28S rRNA gene was equal in C. *kolensis* and *C. insignis*.

Real-time PCR was carried out on an iCycler iQ4 amplificator (Bio-Rad, United States). The data were calculated using iQ5 Optical System Software. The threshold value of accumulation of the amplification products was determined by visual analysis of the PCR product accumulation curves. This value lay within the region of exponential growth of the curves and was equal to 100 in all calculations.

The real-time PCR was carried out using the "Reaction mixture 2.5x for carrying out real-time PCR in the presence of EVA green dye and ROX reference dye" kit (Sintol, Russia). According to the manufacturer's recommendations, the reaction was carried out in a volume of 25 μl: 11 μl of PCR standard water, 10 μl of the finished reaction mixture (including deoxynucle-

oside triphosphates, PCR buffer, $MgCl_2$, and Taq DNA polymerase with antibodies inhibiting enzymatic activity), 1 μl of each primer (the final concentration was 0.4 pmol/μl), and 2 μl of the DNA matrix. The conditions during real-time PCR were as follows: primary denaturation for 4.5 min at 95°C, followed by 50 cycles: 95°C – 15 s, 64°C – 15 s, 72°C – 20 s. The fluorescent signal was recorded at the annealing stage at 64°C. After the amplification, a fusion curve with the 0.5°C temperature gradient (from 55 to 94.5°C) was built, which attested to the presence of only one specific amplification product in each specimen.

Plasmid DNA preparation

The PCR product of the 28S gene in *C. kolensis* rRNA, with a length of 2,199 bp, was cloned into the pGEM-T Easy vector (Promega, United States), yielding pGEM-20b1 plasmid. Plasmid DNA (pDNA) was purified on columns using the Wizard Plus SV Minipreps DNA Purification System kit (Promega, United States), according to the manufacturer's recommendations and including treatment with RNase A. In order to approach the conditions of amplification of the linear DNA, the circular pDNA was cleaved by *Pst*I restriction enzyme, which recognizes the plasmid polylinker and is absent in the inserted fragment. pGEM-20b1 plasmid (2.6 μg) was treated with 2 μl of *Pst*I restrictase (Fermentas) in a 50 μl volume: 23 μl of purified water, 5 μl of 10×Buffer, and 20 μl of pDNA. After incubation for 3 h at 37°C, the plasmid was purified using the phenol-chloroform method and dissolved in the same buffer as the specimens of the genome DNA.

Construction of the calibration curve to determine the rDNA copy number

A calibration curve was used in the method of absolute determination of the rRNA gene copy number. A series of five-fold dilutions of pGEM-20b1/*Pst*I (from 1 ng to 0.32 pg) was used to construct the curve. Each dilution of pDNA was carried out using two replicates. The initial plasmid concentration was measured on a Nanodrop 1000 spectrophotometer with four replicates; the coefficient of variation was equal to 2.36%. The calibration curve had the following characteristics: the coefficient of correlation (R^2) – 0.996; slope of the curve = -3.760; efficiency (E) – 84.5% (Figure).

The size of the plasmid with an inserted fragment is 5216 bp. Based on the nucleotide's composition, the molar mass of the plasmid in double-stranded form was determined using OligoII Mass Calculator v.1.0 software (M = 3.23 × 10^6 g/mol). The number of plasmid DNA molecules was calculated using formula (1):

$$N_p(copies) = \frac{N_A \cdot m}{M} = \frac{6.02 \cdot 10^{23}(copies \cdot mol^{-1}) \cdot 1 \cdot 10^{-15}(g)}{3.23 \cdot 10^6(g \cdot mol^{-1})} = 186.378 \approx 186, \quad (1)$$

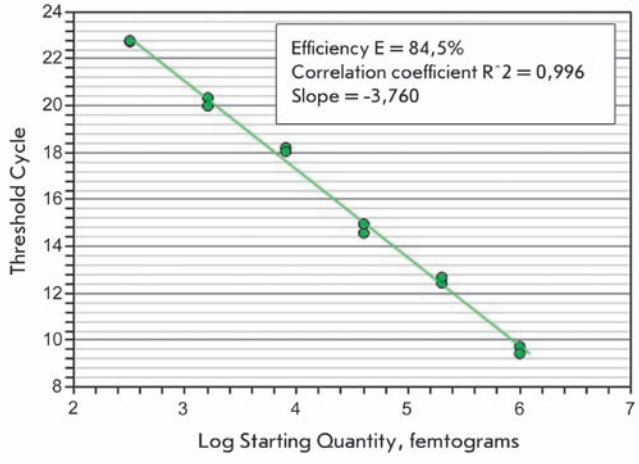
where N_p is the number of plasmid DNA molecules per 1 fg of pDNA; N_A is the Avogadro constant; m is the DNA amount for which the copy number is calculated; and M is the molar mass of a plasmid. It is shown that approximately 186 pDNA molecules correspond to 1 fg of pDNA.

RESULTS

Cytophotometry

It was previously shown that the haploid genome of embryomatic and somatic cells of adult *C. insignis* contains 2.1–2.15 pg of DNA [30].

During the initial stages of the study of CD, we experienced certain methodological difficulties in the preparation of *C. kolensis* specimens for cytophotometry [25, 36]. Therefore, in order to calculate the rRNA gene copy number in *C. kolensis* cells before and after CD, we repeatedly measured the DNA amount in *C. kolensis* cells by quantitative cytophotometry. The measurement results attest to the change in the absolute DNA amount in *C. kolensis* cells before and after CD. The pre-diminuted genome contained 1C = 15.3 ± 3.1 (SD) pg of nuclear DNA, while the somatic line cells after CD at the anaphase stage contained 0.98 ± 0.13 (SD) pg in equivalence to the haploid genome. The relative amount of DNA being eliminated remained constant and was equal to 94%, which is consisted with published reports [25, 34]. The results of cytophotometric measurements make it possible to es-



The calibration curve, which was used in the method of absolute determination of the rRNA gene copy number

timate the number of cells corresponding to different stages of development in *C. kolensis* and *C. insignis*.

Determination of the relative rRNA gene copy number using the $2^{-\Delta C'T}$ method [37]

No nucleotide sequences of any genes in the species subjected to study were known at the time this research was carried out; thus it was not possible to perform the experiment with this type of internal control. Accordingly, the external standardization was carried out relative to the amount of DNA at the beginning of the reaction; i.e., the equal amounts of *C. kolensis* DNA before and after CD and *C. insignis* DNA were compared. The reactions were performed with 500, 400, 300, and 200 pg dilutions for each DNA specimen. Each reaction with a particular starting amount of DNA was represented in three replicates. When determining the relative amounts using one gene, the $2^{-\Delta C'T}$ method is used for processing the PCR results. The specimens with post-diminution (antennae) *C. kolensis* DNA were used as reference in calculations of $\Delta C'_T$ (Table 1). Since the genome's size decreases by a factor of 15.6 as a result of CD, the same cell number before and after CD contains different amounts of DNA. Taking this fact into consideration, the factual ratio between the rDNA copy number will differ from the value obtained by $2^{-\Delta C'T}$ by a factor of 15.6. The same is valid for *C. insignis*; with its diploid genome being 2.2 times larger than the post-diminution genome of *C. kolensis* (Table 1).

Determination of the absolute rRNA gene copy number using the calibration curve

We estimated rRNA gene copy number using a calibration curve. The calibration curve was constructed based on a series of five-fold dilutions of pDNA with pGEM-20b1/*Pst*I; each copy contained a 183 bp fragment of the 28S rRNA gene. The initial amounts of rDNA templates were determined in relation to 500, 400, 300, and 200 pg of pDNA in the same manner as was done for the calculation using the previous method.

Taking into account the differences in the genome size of the two species, the copy number of rDNA in the diploid genome was determined for each sample of DNA (200 – 500 pg) using data generated from real-time PCR reactions (Table 2). Formula (2) was used for the calculation:

$$N_r(copies) = \frac{N_m(fg) \cdot N_p(copies \cdot fg^{-1}) \cdot 2C(pg)}{k(pg)}, \quad (2)$$

where N_r is the 28S rDNA copy number, $2C$ is the size of the diploid genome, N_m is the experimentally determined initial amount of 28S rRNA gene tamplates, N_p is the number of pDNA molecules per 1 fg of pDNA (see formula 1), and k is the amount of the analyzed DNA templates in qPCR.

Thus, the average value of the ratio between the numbers of rRNA genes of prediminuted to post-diminuted genome of *C. kolensis* is 329.94 ± 19.09 (Table 2), while that for the *C. insignis* genome is 11.73 ± 1.16.

DISCUSSION

The methods for calculating the gene copy number used in this study require that a series of assumptions be made. The main assumption for the $2^{-\Delta C'T}$ method is that the reaction is 100% efficient, which is practically

Table 1. Relative quantification of rRNA genes copy number amounts in prediminuted genomes of *C. kolensis* and soma of *C. insignis* using the $2^{-\Delta C'T}$ method

Parameters	DNA samples	DNA quantity taken in reaction (pg)				Mean	Standard deviation	Coefficient of variation
		500	400	300	200			
$\Delta C'_T$	*C. kolensis* before CD	-5.08	-4.99	-4.93	-4.88	-4.97	±0.09	1.7%
	C. insignis	-2.96	-2.72	-2.67	-2.63	-2.75	±0.15	5.4%
$2^{-\Delta C'T}$	*C. kolensis* before CD	33.82	31.78	30.48	29.45	31.38	±1.88	6.0%
	C. insignis	7.78	6.59	6.36	6.19	6.73	±0.72	10.7%
Relative number of copies (post diminuted *C. kolensis* serves as a reference sample)	*C. kolensis* before CD	528.01	496.16	475.86	459.78	489.95	±29.41	6.0%
	C. insignis	121.48	102.86	99.36	96.64	105.09	±11.22	10.7%

Table 2. Absolute quantification of rRNA genes copy number in *C. kolensis* and *C. insignis* using calibration curve

Parameters	DNA samples	DNA quantity taken in reaction (pg)			
		500	400	300	200
Mean of template DNA starting quantity (fg) and standard deviation	*C. kolensis* before CD	135±23.2	102±21.3	71.7±18.1	50.3±1.89
	C. kolensis after CD	5.99±0.953	4.73±0.553	3.45±0.579	2.56±0.447
	C. insignis	37±8.11	25.3±6.13	17.8±3.36	12.7±1.2
rDNA copy number per diploid genome, Nr	*C. kolensis* before CD	1539.86	1454.31	1363.06	1434.35
	C. kolensis after CD	4.38	4.32	4.20	4.68
	C. insignis	58.62	50.10	47.00	50.30
rDNA copy number ratio – N_r(*C. kolensis*-before)/N_r(*C. kolensis*-after)		351.86	336.67	324.46	306.76
rDNA copy number ratio – N_r(*C. insignis*)/N_r(*C. kolensis*-after)		13.39	11.60	11.19	10.76

unattainable. When using the calibration curve that was constructed using pDNA, it is assumed that during the PCR, pDNA is amplified with the same efficiency as the specimens of *C. kolensis* and *C. insignis* under study. Hence, the observed differences in the ratios between the rDNA copy numbers result directly from the features of the methods used in the calculation.

In this study, it was ascertained that a considerable fraction of rRNA genes were eliminated from the genome of presomatic cells of *C. kolensis* during CD. Along with the almost 16-fold decrease in the genome size, the rRNA gene copy number decreases, as well. Let us note that this is the first description of gene elimination by CD in cyclops. Moreover, no evidence of rDNA elimination in multicellular organisms resulting from CD had previously been described.

It is possible that the elimination of rRNA genes is attributable to the necessity of aligning the value of the rRNA gene copy number to genome size. A positive correlation between the genome's size and the rRNA gene copy number was also observed by Prokopovich *et al.* [8]. In a recent elegant experiment with yeast, it was demonstrated that a substantial number of rRNA gene repetitions is important for the general maintenance of the genome's stability [38]. In particular, it has been ascertained that excessive rDNA copies facilitate sister chromatid association, which is of importance for efficient recombinational repair. Let us note that the elimination of rDNA copies resulting from CD is not proportional to a genome size's decrease. This is not surprising, since rRNA genes are generally located in the nucleolar organizers of certain chromosomes and are likely to accumulate in one or several clusters [3], instead of distributing uniformly over the genome. Therefore, an accurate mechanism should exist which would make it possible to infallibly cleave the specific fragment of rRNA genes; their loss will not result in functional deficiency of these genes in somatic line cells. Moreover, not all the copies of rRNA genes have to be active; the number of active copies can vary during the ontogenesis. Three states of rDNAs are distinguished. In one state, the active transcription of rRNA genes takes place; in the two other states, genes are not transcribed. However, rRNA genes are prepared at the beginning of the transcription process and, just as in the previous case, have a euchromatin structure. The densely packed nontranscribable rDNA is also sliced out. It has a heterochromatic structure [6]. It is possible that one of these fractions (most probably the heterochromatin fraction) is eliminated from the genome during CD. The possibility of uniform elimination of rDNA copies from all three rDNA fractions should not be excluded, either.

It would be logical to assume that it is functional rDNA copies that predominately remain intact during CD. It has been known that a considerable portion of the rDNA copies of *Drosophila* and other organisms [39, 40] is affected by specific mobile elements (R1, R2); their incorporation results in the inactivation of rDNA copies. It is possible that it is precisely these copies that are eliminated in *C. kolensis* during CD.

Transcription of ribosomal genes is the key element in the regulation of the general level of protein synthesis in a cell [41, 42]. As has been shown in our study, the *C. insignis* genome which lacks CD is more similar to the post-diminuted *C. kolensis* genome, rather than to the pre-diminuted genome, in terms of the rDNA copy number.

The active rRNA gene expression and synthesis of a substantial number of ribosomes is occurs during early developmental stages. However, CD takes place at the stage of the fourth cleavage division, when the active

expression of any genes is yet to begin. At that stage the developmental process occurs at the expense of the storage compounds. Therefore, the excessive genes in pre-diminution blastomeres should not affect the ribosome number in embryonic cells. It is possible that the eliminated copies participate in gamete maturation, where a greater number of ribosomes may be required.

There that has been an assumption that CD in *C. kolensis* is a developmental stage, during which a transition is made from the cytoplasmic regulation of gene expression during the first cleavage divisions (which is conditioned by the determinants that are already present in the cytoplasm of an unfertilized egg) to the nuclear regulation [1, 29]. Therefore, an initially high abundance of rRNA genes at the pre-diminuted developmental stage of a *C. kolensis* embryo is not possible, whereas the number of rRNA genes in the developing oocyte is over several hundred times larger than the gene number in the post-diminuted genome of somatic cells, and is capable of perfectly determining the higher level of rRNA gene expression in oogenesis, when rRNAs (necessary for the first cleavage divisions) are accumulated.

In conclusion, we would like to note that studying the intra-genomic variation of the rDNA amount in *C. kolensis* as a result of CD is directly linked to understanding the mechanisms of regulation and maintenance of the rDNA copy number in the eukaryotic genome. ●

This work was supported by the Russian Foundation for Basic Research (grant No. 10-04-01376), the Fundamental Research Program of the Russian Academy of Sciences " Biodiversity and Dynamics of Gene Pools", and the Federal Target-Oriented Program "Scientific and Scientific-Pedagogical Personnel of the Innovative Russia" for 2009-2013 (Government Contract No. 16.740.11.0238). We are grateful to T.A. Ketova for her invaluable assistance with data collection.

REFERENCES

1. Grishanin, A.K., Shekhovtsov A.K., Boikova T.V., Akif'ev A.P., Zhimulev I.F. // Tsitologiia (Cytology). 2006. V. 48(5). P. 379-397.
2. Tobler H. // Germ line-soma differentiation; results and problems in cell differentiation / Ed. Hennig W. Berlin: Springer, 1986. V. 13. P. 1–69.
3. Gerbi S.A. // Mol. Evol. Genet. N.Y.: Plenum Press, 1985. P. 419–517.
4. Derenzini M., Pasquinelli G., O'Donohue M.F., Ploton D., Thiry M. // J. Histochem. Cytochem. 2006. V. 54. № 2. P. 131–145.
5. Hernandez-Verdun D. // Histochem. Cell Biol. 2006. V. 126. № 2. P. 135–148.
6. Huang S., Rothblum L.I., Chen D. // Biochem. Cell Biol. 2006. V. 84(4). P. 444–449.
7. Santoro R. // Cell. Mol. Life Sci. 2005. V. 62. P. 2067–2079.
8. Prokopowich C.D., Gregory T.R., Crease T.J. // Genome. 2003. V. 46. № 1. P. 48–50.
9. Bakken A.H. // J. Histochem. Cytochem. 1975. V. 23. №7. P. 463–474.
10. Gall J.G. // Proc. Natl. Acad. Sci. USA. 1968. V. 60. P. 553–559.
11. Gall J.G. // Genetics. 1969. V. 61. P. 121–131.
12. Engberg J. // Europ. J. Cell Biol. 1985. V. 36. P. 133–151.
13. Atwood K.C. // Genetics. 1969. V. 61. P. 319–324.
14. Boncinelli E., Graziani F., Polito L., Malva C. // Cell Diff. 1972. V. 1. P. 133–139.
15. Locker D. // Mol. Gen. Genet. 1976. V. 143. P. 261–272.
16. Ritossa F.M. // Proc. Natl. Acad. Sci. USA. 1968. V. 60. P. 509–519.
17. Ritossa F. // The genetics and biology of Drosophila. V. 1b /Eds Aschburner M., Novitski E. London–New York–San Francisco, 1976. P. 801–812.
18. Ritossa F. M., Scala G. // Genetics. 1969. V. 61. P. 305–314.
19. Tobler H., Etter A., Müller F. // Trends Genet. 1992. V. 8. P. 427–432.
20. Etter A., Aboutanos M., Tobler H., Müller F. // Proc. Natl. Acad. Sci. USA. 1991. V. 88. P. 1593–1596.
21. Etter A., Bernard V., Kenzelmann M., Tobler H., Müller F. // Science. 1994. V. 265. P. 954–956.
22. Spicher A., Etter A., Bernard V., Tobler H., Müller F. // Dev Biol. 1994. V. 164. P. 72–86.
23. Shestakov S.V. // Paleontological Journal. 2003. V. 37(6). P. 609-617.
24. Petrov D.A. // Trends Genetics. 2001. V. 17. P. 23–28.
25. Grishanin, A.K., Khudoli G.A., Shaikhaev G.O., Brodsky V.Ya., Makarov V.B., Akif'ev A.P. // Genetika (Genetics). 1996. V. 32. P. 492 – 499.
26. Zagoskin M.V., Grishanin A.K., Korolev A.L., Palenko M.V., Mukha D.V. // Doklady Biochemistry and Biophysics. 2008. V. 423. № 4. P. 551-555.
27. Degtyarev S., Boykova T., Grishanin A., Belyakin S., Rubtsov N., Karamysheva T., Makarevich G., Akifyev A., Zhimulev I. // Genome Res. 2004. V. 14. № 11. P. 2287–2294.
28. Drouin G. // Genome. 2006. V. 49. P. 657–665.
29. Grishanin A.K.// Ontogeny (Ontogenez). 1995. V. 26(3). P. 188–195.
30. Grishanin A.K., Dams H.-U., Akifiev A.P. // Zool. Stud. 2004. V. 43. P. 8–19.
31. Lee Ch., Lee S., Shin S.G., Hwang S. // Appl. Microbiol. Biotechnol. 2008. V. 78. № 2. P. 371–376.
32. Li Z., Hansen J.L., Liu Y., Zemetra R.S., Berger P.H. // Plant Mol. Biol. Rep. 2004. V. 22. P. 179–188
33. Yuan J.S., Burris J., Stewart N.R., Mentewab A., Stewart C.N., Jr. // BMC Bioinformatics 2007. 8(Suppl 7):S6
34. Grishanin, A.K., Boikova T.V., Marshak T.L., Mel'nik N.G., Naumova E.Yu., Zagoskin M.V., Akif'ev A.P., Zhimulev I.F. // Doklady Biochemistry and Biophysics. 2006. V. 408(1). P. 161-164.
35. Kafiyeny K.A., Timofeeva M.Ya. // Dokl. AN SSR. 1964. V. 154. P. 721–724.
36. Grishanin, A.K., Brodsky V.Ya., Akif'ev A.P. // Doklady Akademii Nauk. V. 338(5). P. 708-710.
37. Livak K.J., Schmittgen T.D. // Methods. 2001. V. 25. № 4. P. 402–408.
38. Ide S., Miyazaki T., Maki H., Kobayashi T. // Science. 2010. V. 327. № 5966. P. 693–696.
39. Kagramanova A.S., Kapelinskaya T.V., Korolev A.L., Mukha D.V. // Molecular Biology. 2007. V. 41(4). P. 546-553.
40. Jakubczak J., Zenni M.K., Woodruff R.C., Eickbush T.H. // Genetics. 1992. V. 131. P. 129–142.
41. Larson D.E., Zahradka P., Sells B.H. // Biochem. Cell. Biol. 1991. V. 69. P. 5–22.
42. Moss T., Stefanovsky V.Y. // Nucl. Acids Res. Mol. Biol. 1995. V. 50. P. 25–66.

Association Study of Xenobiotic Detoxication and Repair Genes with Malignant Brain Tumors in Children

L. E. Salnikova[1], N. I. Zelinskaya[2], O. B. Belopolskaya[1], M. M. Aslanyan[3], A. V. Rubanovich[1]
[1] Vavilov Institute of General Genetics, Russian Academy of Sciences
[2] Federal State Center "Russian Scientific Center of Roentgenoradiology"
[3] Lomonosov Moscow State University
*E-mail: rubanovich@vigg.ru
Received 31.08.2010

ABSTRACT **This study presents the results of research on DNA polymorphism in children with malignant brain tumors (172 patients, 183 in the control group). Genotyping was performed using an allele-specific tetraprimer reaction for the genes of the first (*CYP1A1* (2 sites)) and second phases of xenobiotic detoxication (*GSTM1, GSTT1, GSTP1, GSTM3*), DNA repair genes *XRCC1, XPD* (2 sites), *OGG1*, as well as *NOS1* and *MTHFR*. The increased risk of disease is associated with a minor variant of *CYP1A1* (606G) (p = 0.009; OR = 1.50) and a deletion variant of *GSTT1*, (p = 0.013, OR = 1.96). Maximum disease risk was observed in carriers of double deletions in *GSTT1-GSTM1* (p = 0.017, OR = 2.42). The obtained results are discussed in reference to literary data on the risk of malignant brain tumor formation in children and adults.**
KEYWORDS **gene polymorphism, malignant brain tumors in children, genes of xenobiotic detoxication, DNA repair genes.**

INTRODUCTION

The causes behind the formation of malignant tumors of the central nervous system (CNS) in children, of which 80% are cerebral tumors, are unknown. Risk factors for this type of pathology include inherited susceptibility and the effects of irradiation. Several genetic syndromes, such as the Li-Fraumeni syndrome, Turcot syndrome, neurofibromatosis, and tuberous sclerosis, are known to cause CNS tumors. Moreover, there are families with an increased risk of cerebral tumor formation. For instance, a population cohort from Utah (USA) and a tumor register, which was created based on data from this cohort, indicate the importance of the inheritance factor in most common malignant diseases of the brain in adults (astrocytomas and glioblastomas) [1]. Studies of the Swedish tumor register indicate that first-degree relatives are 2 to 3 times more likely to develop a brain tumor of the same histopathological type as their probands [2]. The offspring of people who had a brain tumor in their childhood are twice as likely to develop a similar tumor [3], the same being true for such a patient's siblings, especially before the age of 5.

Relatives of patients with malignant diseases of the brain are also at risk of developing other oncological conditions. First-degree relatives of glioma patients have an increased standardized incidence ratio (SIR - ratio between the number of observed cases and the expected number) for developing any type of oncopathology (SIR = 1.21), especially before the age of 45 (SIR = 5.08). Relatives of glioma patients most often develop brain tumors (SIR = 2.14), melanomas (SIR = 2.02), and sarcomas (SIR = 3.83) [4].

Brain tumor incidence is now rising in the majority of highly developed countries, especially among children younger than 5 [5]. The role of environmental factors in childhood carcinogenesis, in general and in the CNS tumor development risk, is under investigation. An association has been established between *in utero* ionizing radiation and the risk of developing leucosis and other tumors in childhood [6]. Another such association has been observed for women using diethylstilboestrol during pregnancy and the risk of their daughters developing clear-cell vaginal adenocarcinoma [7]. It has also been shown that brain tumor development in offspring is often associated with parental occupational hazards, such as pesticides [8] or herbicides [9]. The association between maternal diets and the chance of their offspring developing brain tumors has also been researched. The most detrimental factors were found to be high amounts of nitrosamines, widely used for preserving meat and sausage products, as well as large amounts of fat [10, 11]. Transplacental carcinogens of alkyl-nitroseureas are highly carcinogenic in relation to rat brain tumors [12]. Children with excessive [14] or insufficient [13] birth

weight, as well as children with an excessive head circumference (OR = 1.27 for every centimeter of excess after stratification of the cohort for sex, weight and height of the newborn) [15], and children whose mothers have had miscarriages in their anamneses are also at higher risk of developing brain tumors [16]. Intensive smoking (> 10 cigarettes a day) during pregnancy is also among the risk factors contributing to CNS tumors in the offspring [13].

If hereditary syndromes associated with the risk of malignant tumor formation in the nervous system are absent, then genes with low penetrance take on the role of genetic risk factors [17]. Even though the structure of neuro-oncological disease incidence in adults and children differs considerably [18, 19], it is the study of children with sporadic tumors that allows for the effective identification of genetic susceptibilities, as compared to studies of adults. The higher the hereditary risk of cancer development, the easier it is for any environmental factor of even the slightest risk to trigger tumor formation.

Despite the fact that 20% of all the solid tumors in children are brain tumors, there have only been several associative studies of brain tumors on children from various ethnic populations. In a cohort of 73 children in Thailand with various types of CNS tumors it was demonstrated an increased number of homozygous carriers of the minor variant of the *MTHFR* gene (polymorphism A1298C), which is involved in folate metabolism [20]. A study in the United States analyzed the distribution of xenobiotic detoxification gene alleles of *GSTM1* (insertion/deletion), *GSTT1* (insertion/deletion), and *GSTP1* (Ala114Val) genes among 173 child patients and registered the association of a functional allele of *GSTM1* and a rare genotype of *GSTP1* (Val114/Val114) with pediatric astrocytoma [21]. The same researchers showed that a combined cohort of adults (92) and children (43) with brain tumors displayed a distribution of Arg72Pro genotype frequencies for the *P53* gene that was considerably different from the control group. It has also been reported that highly malignant astrocytoma patient cohorts exhibit an increased number of

Table 1. Studied genes and polymorphisms

Gene	Latin name	Polymorphism	dbSNP assigned reference SNP ID	Locus	Gene functions
CytochromeP450 1A1	*CYP1A1*	T606G	rs2606345	15q24.1	The 1-st phase of detoxification - metabolic activation of the aromatic hydrocarbons
		A4889G Ile462Val	rs1048943		
Glutathione S-transferase mu 1	*GSTM1*	Insertion-deletion	-	1p13.3	The 2-nd phase of detoxification – detoxification proper by conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles
Glutathione S-transferase theta 1	*GSTT1*	Insertion-deletion	-	22q11.2	
Glutathione S-transferase mu 3 (brain)	*GSTM3*	G670A V224I	rs7483	1p13.3	
Glutathione S-transferase pi 1	*GSTP1*	A313G Ile105Val	rs1695	11q13	
X-ray repair, complementing defective, in chinese hamster, 1	*XRCC1*	C589T Arg194Trp	rs1799782	19q13.2	Base excision repair
Excision-repair, complementing defective, in chinese hamster, 2	*ERCC2 (XPD)*	A2251C Lys751Gln	rs13181	19q13.3	Nucleotide excision repair
		G862A Asp312Asn	rs1799793		
8-oxoguanine-DNA-glycosylase	*OGG1*	C977G Ser326Cys	rs1052133	3p26.2	Base excision repair - removal 8-oxodeoxyguanosine
Nitric oxide synthase, neuronal	*nNOS (NOS1)*	C276T	rs2682826	12q24.2	NO production in neuronal tissues
5,10-methylenetetrahydrofolate reductase	*MTHFR*	C677T Ala222Val	rs1801133	1p36.3	Conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a cosubstrate for homocysteine remethylation to methionine

Table 2. Age-specific mortality in Russia in the 0-24-year age range

Age, years	Deaths per 1000 population		
	2006	2007	2008
0	10.2	9.4	8.5
1-4	0.7	0.6	0.6
5-9	0.4	0.3	0.3
10-14	0.4	0.4	0.3
15-19	1.1	1.1	1.1
20-24	2.2	2.1	1.9

heterozygous individuals for this *P53* gene polymorphism [22].

Interaction of the environment and the genotype in relation to brain tumor incidence in childhood has been analyzed in two studies [23, 24]. In the case of exposure to phosphoorganic insecticides *in utero* or after birth, the increased risk of developing brain tumors is significantly associated with a polymorphism of the *PON1* (C108T) detoxification gene, for which the above-said compounds are a substrate [23]. This study moved on to confirm the effects of *PON1* on a larger cohort (201 people) and also showed associations with the risk of brain tumor development for two other detoxification genes involved in insecticide metabolism, *FMO1* (C9536A) and *BCHE* (A539T) [24].

This study presents the results of an associative study of genetic risk factors related to the formation of brain tumors in children. The choice of genotyping loci was based on literary data and on personal results obtained in a study of susceptibility genes that increase somatic mutability [25]. This study also includes genes which are primarily expressed in the brain (*GSTM3* aka brain *GSTM*) and in neural tissues (*NOS1,* or *nNOS* – neuronal) and which exhibit association with some oncological diseases [26, 27]. The involved loci are described in Table 1.

EXPERIMENTAL PROCEDURES

A cohort of 172 children with malignant CNS tumors (92 boys and 80 girls) aged 2-16 were included in this study. These children were under treatment in the laboratory of the Children's X-ray Radiology of the Russian Scientific Center of Roentgenoradiology from 2007 to 2010. The average age of the child patients was 8.96±0.38. The most common tumors in the studied cohort were medulloblastomas (N = 58) and brain stem tumors (N = 26). Apart from these, there were also cases of apoplastic ependymoma (N = 19), glioblastoma (N = 10), germinogenic tumors (N = 6), low malignancy astrocytoma (N = 5), high malignancy astrocytoma (N = 5), primitive neuroectodermal tumors (N = 5), and others (N = 38). The control group consisted of 183 people (102 males and 81 females) aged 17 to 21, an average age of 19.90 ± 0.08 years. All the sick children and youths from the control group were of Caucasian race. The patient database contains information on their places of birth and residence. The children's parents gave informed consent for the genotyping procedure. The ten-year difference in the average age of the patient and control groups could not have any significant effect on the allelic variant frequencies in the groups, since mortality in this age group does not exceed 0.1% (Table 2) [28]. Moreover, the first four main causes of death in the 15–24-age group are violence-related: unintentional bodily harm, suicide, undefined bodily harm and murder [29]. The criteria for involvement into the control group were age, nationality, birthplace inside the central regions of the European territory of the Russian Federation, and informed consent to the procedures.

DNA was extracted from peripheral blood lymphocytes using a Diatom DNA Prep 200 kit, which uses guanidine isocyantate and Nucleus–sorbent (Isogen Laboratory, Russia). Genotyping was performed using allele-specific tetraprimer PCR [30]. This method allows the amplification of DNA fragments of alternative alleles in a single test tube. The amplification products were separated using agarose gel electrophoresis and then stained with ethidium bromide.

The statistical analysis was performed using standard methods available in the WinSTAT 2003.1 software integrated into Microsoft Excel.

Estimation of the odds ratios (OR) and the significance of the odds ratio according to the precise Fischer test was accomplished using the free-use software WinPepi: http://www.brixtonhealth.com/pepi4windows.html.

RESULTS

We identified the genotypes of the studied individuals at 12 polymorphic sites of 10 genes. The genotype frequencies in the control group and the patient group were in accordance with the Hardy-Weinberg distribution.

Table 3 compares the frequencies of allele and genotype occurrence for 12 polymorphic sites in children with various tumors of the CNS, as well as youths in the control group. We also distinguished two major groups in the child patients cohort - a group with medulloblastoma and a group with brain stem tumors.

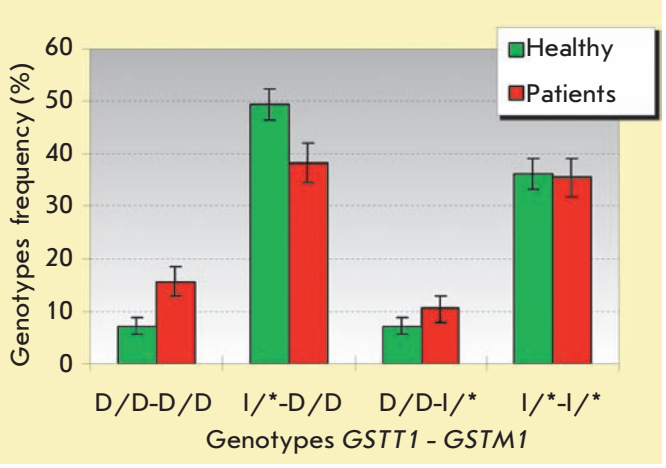
In cases where the polymorphism was of an insertion/deletion nature (genes *GSTM1, GSTT1*), we compared two genotypes: "zero" – homozygous deletion (D/D) and "functional" – a genotype with a functional

Table 3. Genotypes frequencies among the brain tumors patients and in the control group

Loci, alleles, genotypes		Frequencies (%)			
		All brain tumors (N*=172)	Medulloblastoma (N*=63)	Brainstem tumor (N*=26)	Healthy (N*=183)
CYP1A1 T606G rs2606345	T	187 (54.68)	67 (53.18)	30 (57.69)	236 (64.48)
	G	155 (45.32)	59 (46.83)	22 (42.31)	130 (35.52)
	T/T	57 (33.33)	22 (34.92)	10 (38.46)	78 (42.62)
	T/G	73 (42.69)	23 (36.51)	10 (38.46)	80 (43.72)
	G/G	41 (23.98)	18 (28.57)	6 (23.08)	25 (13.66)
CYP1A1 A4889G rs1048943	A	329 (95.64)	120 (95.24)	49 (94.23)	352 (96.18)
	G	15 (4.36)	6 (4.76)	3 (5.77)	14 (3.83)
	A/A	157 (91.28)	57 (90.48)	23 (88.46)	169 (92.35)
	A/G	15 (8.72)	6 (9.52)	3 (11.54)	14 (7.65)
	G/G	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
GSTM1	D/D	93 (54.07)	35 (55.56)	16 (61.54)	95 (51.91)
	I/*	79 (45.93)	28 (44.44)	10 (38.46)	88 (48.09)
GSTT1	D/D	45 (26.16)	20 (31.75)	9 (34.62)	28 (15.30)
	I/I*	127 (73.84)	43 (68.25)	17 (65.38)	155 (84.70)
GSTP1 A313G rs1695	A	242 (70.35)	87 (69.05)	31 (62.00)	247 (67.49)
	G	102 (29.65)	39 (30.95)	19 (38.00)	119 (32.51)
	A/A	80 (46.51)	29 (46.03)	8 (32.00)	79 (43.17)
	A/G	82 (47.67)	29 (46.03)	15 (60.00)	89 (48.63)
	G/G	10 (5.81)	5 (7.94)	2 (8.00)	15 (8.20)
GSTM3 G670A rs 7483	G	203 (59.01)	80 (63.49)	28 (53.85)	222 (60.66)
	A	141 (40.99)	46 (36.51)	24 (46.15)	144 (39.34)
	G/G	63 (36.63)	26 (41.27)	8 (30.77)	73 (39.89)
	G/A	77 (44.77)	28 (44.44)	12 (46.15)	76 (41.53)
	A/A	32 (18.60)	9 (14.29)	6 (23.08)	34 (18.58)
NOS1 C276T rs 2682826	C	243 (70.64)	90 (71.43)	33 (63.46)	271 (75.70)
	T	101 (29.36)	36 (28.57)	19 (36.54)	87 (24.30)
	C/C	84 (48.84)	33 (52.38)	9 (34.62)	103 (57.54)
	C/T	75 (43.60)	24 (38.10)	15 (57.69)	65 (36.31)
	T/T	13 (7.56)	6 (9.52)	2 (7.69)	11 (6.15)
MTHFR C677T rs1801133	C	228 (70.81)	76 (66.67)	36 (72.00)	221 (67.79)
	T	94 (29.19)	38 (33.33)	14 (28.00)	105 (32.21)
	C/C	80 (49.69)	25 (43.86)	13 (52.00)	70 (42.94)
	C/T	68 (42.24)	26 (45.61)	10 (40.00)	81 (49.69)
	T/T	13 (8.07)	6 (10.53)	2 (8.00)	12 (7.36)
XRCC1 C589T rs 1799782	C	322 (93.61)	119 (94.44)	46 (88.46)	337 (94.13)
	T	22 (6.40)	7 (5.56)	6 (11.54)	21 (5.87)
	C/C	150 (87.21)	56 (88.89)	20 (76.92)	160 (89.39)
	C/T	22 (12.79)	7 (11.11)	6 (23.08)	17 (9.50)
	T/T	0 (0.00)	0 (0.00)	0 (0.00)	2 (1.12)
XPD T2251G rs 13181	T	212 (61.63)	82 (65.08)	32 (61.54)	248 (68.13)
	G	132 (38.37)	44 (34.92)	20 (38.46)	116 (31.87)
	T/T	63 (36.63)	25 (39.68)	9 (34.62)	84 (46.15)
	T/G	86 (50.00)	32 (50.79)	14 (53.85)	80 (43.96)
	G/G	23 (13.37)	6 (9.52)	3 (11.54)	18 (9.89)
XPD G862A rs1799793	G	211 (61.70)	82 (66.13)	32 (61.54)	242 (66.48)
	A	131 (38.30)	42 (33.87)	20 (38.46)	122 (33.52)
	G/G	64 (37.43)	26 (41.94)	9 (34.62)	80 (43.96)
	G/A	83 (48.54)	30 (48.39)	13 (50.00)	82 (45.05)
	A/A	24 (14.04)	6 (9.68)	4 (15.38)	20 (10.99)
OGG1 C977G rs1052133	C	274 (80.59)	92 (77.97)	40 (76.92)	270 (78.04)
	G	66 (19.41)	26 (22.03)	12 (23.08)	76 (21.97)
	C/C	116 (68.24)	36 (61.02)	18 (69.23)	105 (60.69)
	C/G	42 (24.71)	20 (33.90)	4 (15.38)	60 (34.68)
	G/G	12 (7.06)	3 (5.08)	4 (15.38)	8 (4.62)

*The number of individuals genotyped at certain loci may differ.

Note. Genotypes associated with diseases are highlighted in grey.



Frequencies of *GSTT1-GSTM1* allele combinations among children with malignant brain tumors and in the healthy group.

allele in either homo- or heterozygous form (I/∗). Hence and further ∗ depicts an unspecified allele.

Increased susceptibility to brain tumor development was observed for carriers of the D/D genotype of the *GSTT1* gene. A two-side Fischer test for all the CNS tumor types yields $p = 0.013$, OR = 1.96, 95% confidence interval 1.16–3.32; for medulloblastoma patients - $p = 0.009$, OR = 2.57, 95% confidence interval 1.33–4.99; for children with brain stem tumors - $p = 0.026$, OR = 2.93, 95% confidence interval 1.21–7.12. Of all the analyzed two-loci combinations, the one associated with the highest risk of malignant brain tumors turned out to be a double deletion of *GSTM1-GSTT1* (27 people – 15.7% patients; $p = 0.017$, OR = 2.42, 95% confidence interval 1.18–4.95) (Figure).

The risk of developing any type of brain tumor, and specifically a medulloblastoma, turned out to be up in carriers of the minor 606G allele of the *CYP1A1* gene (for all types of tumor - $p = 0.009$, according to the two-sided Fischer test, OR= 1.50, 95% confidence interval 1.11–2.03, for medulloblastoma - $p = 0.026$, OR = 1.60, 95% confidence interval 1.06–2.41).

Among the brain stem tumor patients there was an elevated number of *NOS1**276T minor allele carriers in both homo- and heterozygous forms, where $p = 0.035$, OR = 2.56, 95% confidence interval 1.10–5.96. The whole patient group also exhibited an increased occurrence of the minor allele; however, these data were statistically insignificant - $p = 0.11$, OR = 1.42, 95% confidence interval 0.93–2.16.

There was also a tendency for association between the 2251G minor allele of the nucleotide excision DNA repair gene *XPD* in both homo- and heterozygous forms and an increased chance of developing a brain tumor ($p = 0.084$, OR = 1.48, 95% confidence interval 0.97–2.27).

DISCUSSION

The detoxification of xenobiotics consists of two main stages of detoxification and a third stage - secretion of the detoxified products out of the cell. The first stage involves activation of the xenobiotic compounds by P-450 cytochromes and a number of other enzymes. The second stage is the detoxification, *per se*, and it involves glutathione-S-transferases and other proteins. The intermediary electrophilic metabolites that were formed in the first stage are toxic, and effective detoxification requires a fine balance between the activity of the first- and second-stage enzymes. This balance is deregulated both by insufficient activity of the polymorphic variants of the second-stage enzymes and by the increased activity of the first-stage enzymes [31]. Increased activity of the first-stage detoxification enzymes and insufficient activity of the second-stage enzymes (GST) cause an increase in the level of activated electrophilic metabolites, thus increasing the deleterious effects of the xenobiotic compounds.

This study demonstrates that there is an association between certain xenobiotic detoxification gene alleles and the development of brain tumors in children. The risk of developing malignant tumors in the brain during childhood is increased in carriers of a minor variant of *CYP1A1* (606G).

The role of *CYP1A1* polymorphism has not been studied in relation to child neurooncology. Associative studies on adults have shown no association between *CYP1A1* A4889G (Ile462Val) polymorphism and the risk of developing glioma or several other types of malignant brain tumors [32–34]. The *CYP1A1* gene is located in the 15q22-24 region, and people with a hereditary predisposition towards glioma have exhibited associations between the disease and low-penetrance markers in the 15q23-q26.3 region which overlaps this locus [35].

Data on the role of *CYP1A1* gene polymorphism in carcinogenesis are contradictory, and it seems that their role considerably depends on the interaction between the genotype and the environment [36]. The T606G site is located in the first intron of the *CYP1A1* locus. The single nucleotide substitutions (SNP) located in the intron regions do not usually influence gene activity. However, the T606G polymorphism has been associated with lung cancer [37], hormone-dependent tumors [38], and with the level of sex hormones, which are substrates of CYP1A1 [39]. There are data on the T606G site which indicate that in the absence of specific substrates, the allelic 606T (SNP T606G) vari-

ant of the *CYP1A1* gene is expressed more actively, whereas the 606G variant is induced in the presence of specific substrates (polycyclic aromatic hydrocarbons of exogenous origins, such as foods, industrial waste, tobacco smoke, as well as endogenous compounds, such as estrogens). The differential effect of the allelic variants 606G and 606T on the observed effects under ecologically unfavorable conditions (industrial pollution of air, smoking), as well as in their absence, has been demonstrated in two independent studies [40, 41]. The two studied sites in the *CYP1A1* locus which were studied in this work are in strong linkage disequilibrium - D'=0.913, r=0.229, p=0, and the minor alleles (4889G, 606G) belong to a single linkage group. Linkage disequilibrium data in the patient and control groups were identical. In this work, we have confirmed our previous data obtained on different cohorts and indicating that the polymorphic sites A4889G and T606G are linked [25, 42]. According to data from HapMap, the frequency of the 606G allele in Caucasians is 0.36–0.45, while the frequency of 4889G is 0.04–0.07. Until recently, researchers had studied three major polymorphic sites in the *CYP1A1* gene in European populations: T3801C, A4889G, and C4887A [43]. Besides, polymorphism T606G has a functional character, the frequency of the 606G allele is higher than the frequency of the minor alleles at other polymorphic sites. Thus, this allele seems to be a new promising marker for associative studies of multifactor diseases.

Our study also shows associations between the formation of malignant brain tumors and the possession of deletion variants of *GSTT1* (D/D) (OR=1.96, p= 0.013). Association between polymorphism of glutathione-S-transferase genes, which encode enzymes for the second phase of xenobiotic detoxification, and the development of brain tumors in children, was analyzed in study [21]. Statistically significant results were obtained for the functional allele of *GSTM1* and a minor allele of *GSTP1* 313G. Association between the development of malignant brain tumors in adults and polymorphisms of glutathione-S-transferase-encoding genes was analyzed much more thoroughly; for instance, 10 studies were conducted in Europe [32–34, 44–50]. The results of seven of these studies and the results of the aforementioned work on a cohort of sick children [21] were combined in a meta-analysis [51], which was performed for two of the most common nosological forms: gliomas and meningiomas. According to this meta-analysis, in Caucasians the deletion variant of *GSTT1* occurred significantly more often in meningioma patients (OR = 1.95). No differences in the frequencies of the *GSTM1* (Ins/Del) and *GSTP1* A313G (Ile105Val) and the adjacent C341T (Ala114Val) allele were observed between the patient and control groups. Another large-scale study obtained data indicating the absence of an association between polymorphisms of *GSTM1* (Ins/Del), *GSTM3* (A63C), *GSTT1* (Ins/Del) and the development of gliomas, glioblastomas, and meningiomas. It was demonstrated that the 105G-114C (Val-Ala) haplotype of *GSTP1* has a weak protective effect on the chance of developing glioma [32].

No significant differences in DNA repair gene allele frequencies were found (Table 3); however, there was a tendency level association of the minor 2251G allele in the *XPD* locus with an elevated chance of developing the disease.

Associative studies of malignant tumors of any localization concerning DNA repair gene polymorphisms most often involve *XPD* nucleotide excision repair genes and *XRCC1* base excision repair genes [52], which are located on the same region of the chromosome (19q13.2–3). Most of the results of associative studies of brain tumors are summarized in review [53]. It was shown that in adults, the most common malignant tumors of neuroepithelial tissues are associated with the nucleotide excision repair genes *XPD, ERCC1* and a gene located in the same (19q13.2-3) region of the chromosome - *GLTSCR1* (glioma tumor suppressor candidate of an unknown function) [54]. Caggana *et al.* [55] showed that of 7 polymorphic sites in the *XPD* gene, maximum association with an increased risk of glioma was observed for the least studied synonymous Arg156Arg polymorphism, which may be a marker of another unknown gene that predisposes potential patients to this disease. Sites T2251G (Lys751Gln) and G862A (Asp312Asn) of the *XPD* gene are located 12340 b.p. apart and are linked. This work has obtained the following linkage disequilibrium data: D'=0.674, r=0.662, p=0 (no difference between the patient and control groups), which is in agreement with the published data on Caucasians [56]. Despite the absence of significant results concerning DNA repair genes in this work, studying polymorphic loci in the 19q13.2-3 chromosome region seems a promising line of research that could lead to the discovery of risk markers for malignant brain tumors in children.

This study also shows that a minor allele of the neuronal nitric oxide synthase occurs significantly more often in patients with brain stem tumors (Table 3); differences for the whole patient group are statistically insignificant.

Genes from the nitric oxide synthase family, which includes the neuronal nitric oxide synthase gene, are usually studied in connection with inflammatory processes. However, *nNOS* polymorphism is associated with melanoma predisposition [27]. Melanoma is included into the nerve-ending tumor group. Families that have a hereditary predisposition towards brain tumors are

often predisposed towards developing melanomas as well [5]. Taking into account the elevated expression of *nNOS* in nervous tissue, as well as the putative cross-sensitivity to melanoma and glioma, we resolved to analyze the C276T site of the *nNOS* gene. This polymorphism is considered to be functional, since the single nucleotide substitution in the untranslated region results in elevated mRNA expression of the minor variant [57]. Significant results on the association of the minor allele with increased risk of developing brain tumors were only observed for the small group of patients with brain stem tumors and will of course require further study.

Our previous associative studies of xenobiotic detoxification genes have shown that women with reproductive system diseases (mainly myomas and chronic cystic mastitis) carrying the 606G, 4889G alleles of the *CYP1A1* gene have an increased frequency of somatic mutations at the T-cell receptor (TCR) locus in peripheral blood lymphocytes (phentoype $CD^{3-}CD^{4+}$). It is known that the number of such TCR-mutant lymphocytes is elevated in cancer patients (cancer of the larynx and hypopharynx, thyroid gland tumor, cervical cancer and Hodgkin's lymphoma) and in people with hereditary predispositions towards oncological diseases (ataxia-teleangiectosia) [58, 59]. The single direction of the effects in two separate studies may indicate the pleiotropic effect of detoxification genes, which leads to insufficient resistance of the organism in the genotype-environment interaction process. Besides the possible increased risk of disease due to altered detoxification enzyme activity, allelic variants associated with somatic mutability and with predisposition to the formation of malignant tumors in childhood may act as markers of a linked group of unknown genes that can be responsible for some of the observed effects. The obtained results, if confirmed by independent studies, can be useful for identifying the genetic risk factors involved in the formation of malignant tumors in children. ●

This work was supported under the Biological Variety RAS Presidium program for basic research, under the Genepools and Genetic Variety subprogram.

REFERENCES

1. Blumenthal D.T., Cannon-Albright L.A. // Neurology. 2008. V. 71. P. 1015–1020
2. Malmer B., Henriksson R., Grönberg H. // Int. J. Cancer. 2003. V. 106. № 2. P. 260–263
3. Sankila R., Olsen J.H., Anderson H., Garwicz S., Glattre E., Hertz H., Langmark F., Lanning M., Møller T., Tulinius H. // N. Engl. J. Med. 1998. V. 338. № 13. P. 1339–1344
4. Scheurer M.E., Etzel C.J., Liu M., El-Zein R., Airewele G.E., Malmer B., Aldape K.D., Weinberg J.S., Yung W.K., Bondy M.L. // Cancer Epidemiol. Biomarkers Prev. 2007. V. 16. № 11. P. 2491–2495.
5. Hemminki K., Li X., Vaittinen P., Dong C. // Br. J. Cancer. 2000. V. 83. № 3. P. 407–411.
6. Busby C.C. // Int. J. Environ. Res. Public Health. 2009. V. 6. № 12. P. 3105–3114.
7. Herbst A.L., Ulfelder H., Poskanzer D.C. // N. EngI. J. Med. 1971. V. 284. P. 878–881.
8. Bassil K.L., Vakil C., Sanborn M., Cole D.C., Kaur J.S., Kerr K.J. // Can. Fam. Physician. 2007. V. 53. № 10. P. 1704–1711.
9. Shim Y.K., Mlynarek S.P., van Wijngaarden E. // Environ. Health Perspect. 2009. V. 117. № 6. P. 1002–1006.
10. Dietrich M., Block G., Pogoda J.M., Buffler P., Hecht S., Preston-Martin S. // Cancer Causes & Control. 2005. V. 16. № 6. P. 619–635.
11. Pogoda J.M., Preston-Martin S., Howe G., Lubin F., Mueller B.A., Holly E.A., Filippini G., Peris-Bonet R., Mccredie M.R.E., Cordier S., Choi W. // Ann. Epidemiol. 2009. V. 19. № 3. P. 148–160.
12. Idowu O.E., Idowu M.A. // Afr. Health Sci. 2008. V. 8. № 1. P. 1–4.
13. Schüz J., Kaletsch U., Kaatsch P., Meinert R., Michaelis J. // Med. Pediatr. Oncol. 2001. V. 36. № 2. P. 274–282.
14. Harder T., Plagemann A., Harder A. // Am. J. Epidemiol. 2008. V. 168. № 4. P. 366–373.
15. Samuelsen S.O., Bakketeig L.S., Tretli S., Johannesen T.B., Magnus P. // Lancet Oncol. 2006. V. 7. № 1. P. 39–42.
16. Cantwell M.M., Forman M.R., Middleton R.J., Murray L.J. // Br. J. Cancer. 2008. V. 99. № 5. P. 796–799.
17. de Andrade M., Barnholtz J.S., Amos C.I., Adatto P., Spencer C., Bondy M.L. // Genet. Epidemiol. 2001. V. 20. № 2. P. 258–270.
18. Merchant T.E., Pollack I.F., Loeffler J.S. // Semin. Radiat. Oncol. 2010. V. 20. № 1. P. 58–66.
19. Arora R.S., Alston R.D., Eden T.O.B., Estlin E.J., Moran A., Birch J.M. // Neuro-Oncology. 2009. V. 11. № 4. P. 403–413.
20. Sirachainan N., Wongruangsri S., Kajanachumpol S., Pakakasama S., Visudtibhan A., Nuchprayoon I., Lusawat A., Phudhicharoenrat S., Shuangshoti S., Hongeng S. // Cancer Detect. Prev. 2008.V. 32. № 1. P. 72–78.
21. Ezer R., Alonso M., Pereira E., Kim M., Allen J.C., Miller D.C., Newcomb E.W. // J. Neurooncology. 2002. V. 59. № 2. P. 123–134.
22. Parhar P., Ezer R., Shao Y., Allen J.C., Miller D.C., Newcomb E.W. // Brain Res. Mol. Brain Res. 2005. V. 137. № 1–2. P. 98–103.
23. Nielsen S.S., Mueller B.A., De Roos A.J., Viernes H.-M.A., Farin F.M., Checkoway H. // Environ. Health Perspect. 2005. V. 113. № 7. P. 909–913.
24. Nielsen S.S., McKean-Cowdin R., Farin F.M., Holly E.A., Preston-Martin S., Mueller B.A. // Environ. Health Perspect. 2010. V. 118. № 1. P. 144–149.
25. Salnikova L.E., Zamulaeva I.A., Belopolskaya O.B., Ivanova T.I., Kuznetsova G.I., Saenko A.S., Abilev S.K., Rubanovich A.V. // Ekologicheskaya genetika (Ecological Genetics). 2010. V. 8. № 2. P. 18–23.

26. De Roos A.J., Rothman N., Brown M., Bell D.A., Pittman G.S., Shapiro W.R., Selker R.G., Fine H.A., Black P.M., Inskip P.D. // Neuro-Oncology. 2006. V. 8. № 2. P. 145–155.
27. Li C., Hu Z., Liu Z., Wang L.-E., Gershenwald J.E., Lee J.E., Prieto V.G., Duvic M., Grimm E.A., Wei Q. // Cancer. 2007. V. 109. № 8. P. 1570–1578.
28. Demograficheskiy Ezhegodnik Rossii (Demographic Yearbook of Russia). 2009. Statistics of Russia. Moscow: Rosstat, 2009. 557 p.
29. Demoscope. Weekly. Electronic version of the newsletter Naselenie i Obshestvo. Smertnost ot vneshnih prichin i vozrast (Population and Society Newsletter. Age-specific average annual mortality from external causes). 2001. № 29–30. URL: http://www.demoscope.ru/weekly/029/tema04.php
30. Hamajima N. // Exp. Rev. Mol. Diagnosis. 2001. V. 1. № 1. P. 119–123.
31. Baranov V.S., Baranova E.V., Ivastchenko T.E., Aseev M.V. Genom cheloveka i geni predraspolozhennosti. Vvedidnie v prediktivnuyu meditsinu. (Human Genome and Predisposition Genes: Introduction to Predictive Medicine), St. Petersburg: Intermedika, 2000. 272 p.
32. Trizna Z., de Andrade M., Kyritsis A.P., Briggs K., Levin V.A., Bruner J.M., Wei Q., Bondy M.L. // Cancer Epidemiol. Biomarkers Prev. 1998. V. 7. P. 553–555.
33. Schwartzbaum J.A, Ahlbom A., Lönn S., Warholm M., Rannug A., Auvinen A., Christensen H.C., Henriksson R., Christoffer Johansen C., Lindholm C., Malmer B., Salminen T., Schoemaker M.J., Swerdlow A.J., Feychting M. // Cancer Epidemiol. Biomarkers Prev. 2007. V. 16. № 3. P. 559–565.
34. De Roos A.J., Rothman N., Inskip P.D., Linet M.S., Shapiro W.R., Selker R.G., Fine H.A., Black P.M., Pittman G.S., Bell D.A. // Cancer Epidemiol. Biomarkers Prev. 2003. V. 12. № 1. P. 14–22.
35. Paunu N., Lahermo P., Onkamo P., Ollikainen V., Rantala I., Helén P., Simola K.O. J., Kere J., Haapasalo H. // Cancer Res. 2002. V. 62. P. 3798–3802.
36. Androutsopoulos V.P., Tsatsakis A.M., Spandidos D.A. // BMC Cancer. 2009. 9:187. URL: http://www.biomedcentral.com/1471-2407/9/187
37. Rotunno M., Yu K., Lubin J.H., Consonni D., Pesatori A.C., Goldstein A.M., Goldin L.R., Wacholder S., Welch R., Burdette L., Chanock S.J., Bertazzi P.A., Tucker M.A., Caporaso N.E., Chatterjee N., Bergen A.W., Landi M.T. // PLOS. 2009. V. 4. № 5. e5652. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2682568
38. Figueroa J.D., Sakoda L.C., Graubard B.I., Chanock S., Rubertone M.V., Erickson R.L., McGlynn K.A. // Cancer Causes & Control. 2008. V. 19. № 9. P. 917–929.
39. Sowers M.R., Wilson A.L., Kardia S.R., Chu J., McConnell D.S. // Am. J. Med. 2006. V. 119. № 9. Suppl 1. S44-51. URL: http://www.amjmed.com/article/S0002-9343%2806%2900828-X/pdf
40. Wang S., Chanock S., Tang D., Zhigang L., Jedrychowski W., Perera F.P. // Cancer Epidemiol. Biomarkers Prev. 2008. V. 17. № 2. P. 405–413.
41. Lam T.K., Rotunno M., Lubin J.H., Wacholder S., Consonni D., Pesatori A.C., Bertazzi P.A., Chanock S.J., Burdette L., Goldstein A.M., Tucker M.A., Caporaso N.E., Subar A.F., Landi M.T. // Carcinogenesis. 2010. V. 31. № 4. P. 634–642.
42. Salnikova L.E., Smelaya T.V., Moroz V.V., Golubev A.M., Lapteva N.Sh., Rubanovich A.V.//Obschaya reanimatologiya (General Reanimatology). 2010. V. 6. № 1. P. 5–10.
43. Georgiadis P., Topinka J., Vlachodimitropoulos D., Stoikidou M., Gioka M., Stephanou G., Autrup H., Demopoulos N.A., Katsouyanni K., Sram R., Kyrtopoulos S.A. // Carcinogenesis. 2005. V. 26. № 1. P. 93–101.
44. Pinarbasi H., Silig Y., Gurelik M. // Cancer Genet. Cytogenet. 2005. V. 156. P. 144–149.
45. Wrensch M., Kelsey K.T., Liu M., Miike R., Moghadassi M., Aldape K., McMillan A., Wiencke J.K. // Cancer Epidemiol. Biomarkers Prev. 2004. V. 13. P. 461–467.
46. Butler M.A., Ruder A.M., Daly A.K., Waters M.A., Carreón T., Schulte P.A. // Proc. Am. Assoc. Cancer Res. 2003. V. 44. P. 128.
47. Kondratyeva T.V., Imyanitov E.N., Togo A.V., Zaitseva O.A., Iatsuk O.S., Bersnev V.P., Khanson K.P. // Vopr Onkol. (Oncology Questions). 1999. V. 45. P. 523–527.
48. Elexpuru-Camiruaga J., Buxton N., Kandula V., Dias P.S., Campbell D., Mclntosh J., Broome J., Jones P., Inskip A., Alldersea J., Fryer A.A., Strange R.C. // Cancer Res. 1995. V. 55. № 19. P. 4237–4239.
49. Kelsey K.T., Wrensch M., Zuo Z.F., Miike R., Wiencke J.K. // Pharmacogenetics. 1997. V. 7. P. 463–468.
50. Wiencke J.K., Wrensch M.R., Miike R., Zuo Z., Kelsey K.T. // Carcinogenesis. 1997. V. 18. P. 1431–1433.
51. Lai R., Crevier L., Thabane L. // Cancer Epidemiol. Biomarkers Prev. 2005. V. 14. № 7. P. 1784–1790.
52. Vineis P., Manuguerra M., Kavvoura F.K., Guarrera S., Allione A., Rosa F., Di Gregorio A., Polidoro S., Saletta F., John P.A., Ioannidis J.P.A., Giuseppe Matullo G. // J. Natl. Cancer Inst. 2009. V. 101. № 1. P. 24–36.
53. Bethke L., Webb E., Murray A., Schoemaker M., Johansen C., Christensen H.C., Muir K., McKinney P., Hepworth S., Dimitropoulou P., Lophatananon A., Feychting M., Lönn S., Ahlbom A., Malmer B., Henriksson R., Auvinen A., Kiuru A., Salminen T., Swerdlowand A., Houlston R. // Human Mol. Genet. 2008. V. 17. № 6. P. 800–805.
54. Yang P., Kollmeyer T.M., Buckner K., Bamlet W., Ballman K.V., Jenkins R.B. // Cancer. 2005. V. 103. № 11. P. 2363–2372.
55. Caggana M., Kilgallen J., Conroy J.M., Wiencke J.K., Kelsey K.T., Miike R., Chen P., Wrensch M.R. // Cancer Epidemiol. Biomarkers Prev. 2001. V. 10. P. 355–360.
56. Butkiewicz D., Rusin M., Enewold L., Shields P.G., Chorazy M., Harris C.C. // Carcinogenesis. 2001. V. 22. № 4. P. 593–597.
57. Venturelli E., Villa A., Scarpini E., Fenoglio C., Guidi I., Lovati C., Marcone A., Cortini F., Scalabrini D., Clerici F., Bresolin N., Mariani C., Cappa S., Galimberti D. // Eur. J. Neurol. 2008. V. 15. № 1. P. 77–81.
58. Zamulaeva I.A., Saenko A.S., Orlova N.V., Smirnova S.G., Tsyb A.F. Posobie dlya vrachey (Manual for Doctors). Obninsk, 2007. 19 p.
59. Kyoizumi S., Akiyama M., Hirai Y., Kusunoki Y., Tanabe K., Umeki S. // J. Exp. Med. 1990. V. 171. № 6. P. 1981–1999.

Modeling of the Full-Size 3D Structure of Human Chaperone Hsp70 and Study of Its Interdomain Interactions

K. A. Chernorizov[1], V.K.Švedas[1,2*]
[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University
[2]Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University
*E-mail: vytas@belozersky.msu.ru
Received 02.11.2010

ABSTRACT **Hsp70 is a chaperone protein that participates in the folding of de novo synthesized proteins, protection of the hydrophobic regions of denaturated proteins, the regulation of apoptosis, the immune response, and several other cellular processes. Despite the large number of publications devoted to the functioning and structure of Hsp70, a reliable full-size 3D structure of this protein remains currently unavailable. Several probable full-size models of human Hsp70 have been constructed based on the structures of individual domains and their components from different organisms and using molecular modeling methodology. The stability of the obtained structures was studied using molecular dynamics. As a result of such an analysis, the most adequate model was selected. The model was built on the basis of Hsp70 elements from *Bos Taurus* and *Caenorhabditis elegans*. Using the method of steered molecular dynamics, the key salt bridges responsible for the interdomain interactions were identified: Arg171: Glu516 and Arg416: Glu218. Based on the performed molecular modeling, the scheme of the mechanism triggering ATP hydrolysis and leading to the separation of ATPase and the substrate-binding domains was proposed.**
KEYWORDS **chaperone, Hsp70, model, tertiary structure, ATPase and substrate-binding domains, molecular dynamics.**

INTRODUCTION

Hsp70 (Heat shock 70 kDa protein 1A/1B) is a two-domain ATP-dependent chaperone. Its function is to protect from aggregation the hydrophobic surfaces of proteins denaturated under stress and to promote their correct folding or direct them toward degradation. Proteins of this class are involved in the transport of peptides through the mitochondrial and cellular membranes. Depending on whether their localization is extracellular or intracellular, Hsp70s are involved, respectively, in the autoimmune labeling of tumor cells or in the protection of cells from stress and apoptosis [1, 2]. Hsp70 consists of two functionally distinct elements: the N-terminal ATPase domain hydrolyzes ATP to ADP with the formation of an inorganic phosphate; and the C-terminal substrate-binding domain binds hydrophobic peptides. The ATPase domain consists of two large subdomains, I and II, each of which is divided into the sections A and B (see Fig. 1). ATP binds at the bottom of the cleft located between the subdomains I and II. The ATPase domain also contains a binding site of the nucleotide exchange factor, which promotes the "recharging" of ADP to ATP by pulling the subdomains I and II apart from each other [3, 4].

The substrate-binding domain (SBD) consists of a β-sheet subdomain (βSBD), which binds the substrate peptide, and an α-helical "cap" (αSBD), formed by five α-helices A, B, C, D and E (see Fig. 1). Due to conformational changes mediated by ATP hydrolysis and the action of cochaperone Hsp40, the substrate-binding domain changes its state from a closed to an open one, while the "cap" regulates its accessibility for the hydrophobic sites of the target peptides [5].

Mediated by Hsp40, ATP hydrolysis and substrate binding lead to separation of the Hsp70 substrate-binding domain from the ATPase domain. As a result, they remain connected only by the hydrophobic linker. It is known that Hsp70 is capable of forming oligomers in solution. Most likely, this is due to the fact that the substrate-binding domain of Hsp70 is able to bind the hydrophobic linker of another Hsp70 molecule when its domains are separated [8, 9].

ATP hydrolysis is induced by the J-domain of cochaperone Hsp40. A study of the binding of Hsp70 and Hsp40 showed that Asp35 of yeast Hsp40 J-domain interacts with Arg171 of Hsp70. This leads to ATP hydrolysis and subsequent separation of Hsp70 domains from each other. The signal for ATP hydrolysis is as-

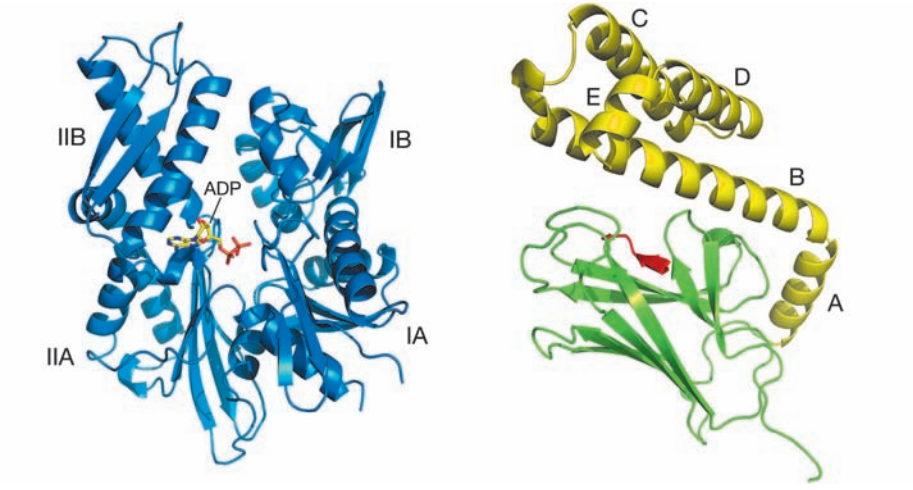


Fig. 1. ATPase domain of human Hsp70 in complex with ADP (left; pdb ID 1HJO [6]) and the substrate-binding domain of Hsp70 prokaryotic homolog from *E.coli* – chaperone DnaK (right), consisting of αSBD (yellow), βSBD (green) in complex with substrate peptide (red); pdb ID 1DKZ) [7].

sumed to pass through the chain J-domain of Hsp40 → Arg171 (interdomain linker of Hsp70) → Tyr371 (Hsp70 ATPase domain) → Ile181 (Hsp70 ATPase domain) [10-12]. Then, the substrate binds and domains separate from each other. The exact mechanism of this process remains unknown.

It is assumed that the structure of the substrate-binding domain of mammalian Hsp70 is similar to the structure of the analogous domain of its prokaryotic homolog DnaK (see Fig. 1) [7]. Although the structures of individual elements of the Hsp70 substrate-binding domain are known [11, 13], the full-size three-dimensional structure of mammalian Hsp70 is unknown. In this paper we attempt to construct a full-size three-dimensional structure of Hsp70 using molecular modeling methods. The structure is of interest not only for studying the interaction between ATPase and the substrate-binding domains of the chaperone, but also as a promising target for a search for anticancer agents. Because of their immunogenic properties, chaperones have been subjects of successful pharmacological studies [5, 14]. Establishing the plausible structure of human Hsp70 and the progress achieved in understanding the mechanism by which it functions could help to speed up research in the field of chaperone-associated cancer therapy strategies and reduce their cost.

EXPERIMENTAL

Homology modeling was performed using the Swiss Model [15, 16]. The systems for molecular dynamics simulation were prepared in tleap (program from the AmberTools 1.2 package). Molecular dynamics simulations were performed in the software package Amber10 [17]. The protein molecule was put in a cubic cell, with the distance between the protein and the cell boundary not less than 12 Å. TIP3P water was used as a solvent. The integration step in all stages was 2 fs. Energy minimization was restricted to 5,000 steps, and the steepest descent algorithm was replaced by the conjugated gradient after 2,500 steps. The cutoff distance was 10 Å. The structures under study were equilibrated in four stages: energy minimization with position restraints on all atoms, energy minimization without restraints, heating and molecular dynamics simulation. Heating from 0 to 300 K was performed at a constant volume for 50 ps. Molecular dynamics was implemented at a constant pressure. The Langevin thermostat was used for temperature control. The same parameters were used in running steered molecular dynamics. In simulations of salt bridges, the disruption force constant was taken to be 20 kcal/mol · Å^2. The visual analysis of three-dimensional structures was performed using VMD 1.8.6 [18].

RESULTS AND DISCUSSION

Modeling the full-size structure of Hsp70

For the modeling, the following elements from the Protein Data Bank were used: substrate-binding domain of DnaK from *Escherichia coli* (1DKZ; Fig. 1); α-subdomain of Hsp70 substrate-binding domain from *Caenorhabditis elegans* (2P32 [19]); ATPase domain, βSBD and α-helix 'A' αSBD from bovine Hsc70 (1YUW). The latter structure describes the contact of the ATPase domain with the substrate-binding domain when the domains are rigidly coupled to each other.

Model hHsp70_1dkz

Modeling was conducted in three steps:

1. Based on a primary amino acid sequence of human Hsp70 (hHsp70) (PID: P08107) and a tertiary structure of its close homologue from *B. taurus* – Hsc70 (1YUW), a model was built describing the ATPase domain,

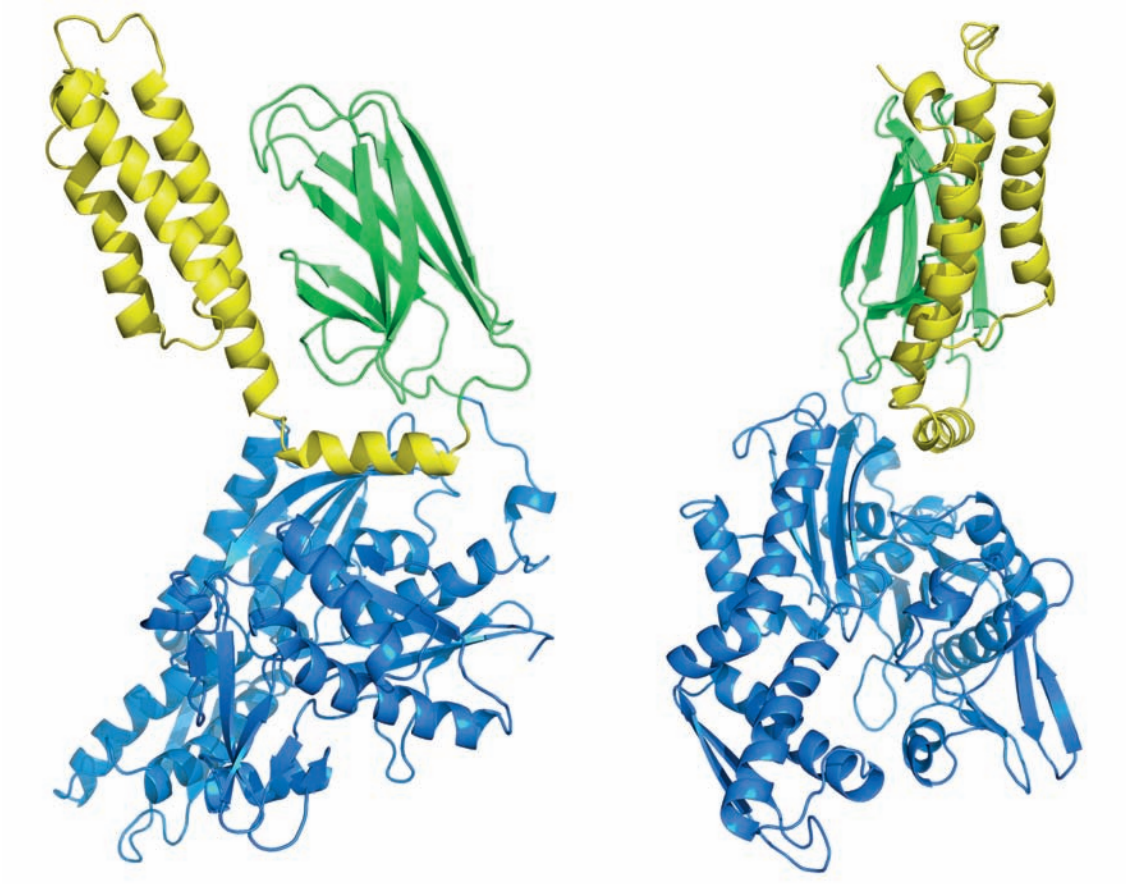


Fig.2. Model hHsp70_2p32 after molecular dynamics simulation. ATPase domain is colored blue, αSBD - yellow, βSBD – green.

βSBD and the helix 'A' of αSBD. The identity of the model with the template was 88.6%. To obtain a full-size structure, we had to construct the remaining part of the αSBD.

2. To construct the αSBD the structure of the prokaryotic homolog of Hsp70, DnaK (1DKZ) was used as a template. In 1DKZ, the α-subdomain, in a complex with the β-subdomain, forms a full-size substrate-binding domain. The identity of the constructed structure with the template was 44.7%.

3. The resulting structures were superimposed by βSBD. The model was equilibrated using molecular dynamics simulation for 6.5 ns.

In the resulting construct, the ATPase domain and βSBD were built by homology modeling with the close eukaryotic relative of Hsp70; and αSBD, by homology modeling with the prokaryotic chaperone DnaK from *E. coli*.

An analysis of molecular dynamic trajectories showed that the spatial structures of the ATPase domain and βSBD were stable, while the segment of αSBD, formed by the α-helices 'B'-'E', was prone to unfolding.

Model hHsp70_2p32

The structure of αSBD from nematode (*Caenorhabditis elegans*; PDB ID: 2P32) was taken as an alternative template for homologous modeling of the unstable segment. The identity of the resulting structure and the template was 63%. The created construction was superimposed on hHsp70_1dkz by the α-helix 'B', after which part of the α-helix 'B', as well as helices 'C', 'D' and 'E', was replaced by the corresponding elements constructed by homology modeling with the 2P32. The resulting model was equilibrated for 8.5 ns.

The model hHsp70_2p32 (after equilibration) is shown in Fig. 2. Helices 'B', 'C' and 'D' form a bundle similar in tertiary structure with 1DKZ; however, in contrast, the α-helix 'E' in hHsp70_2p32 is reduced. Based on an analysis of the molecular dynamic trajectory, the constructed structure showed high stability and was accepted as a working model (hereafter 'hHsp70') for further studies.

Study of interdomain interactions in hHsp70

In the process of chaperone functioning, its domains diverge with a disruption of all noncovalent interactions.

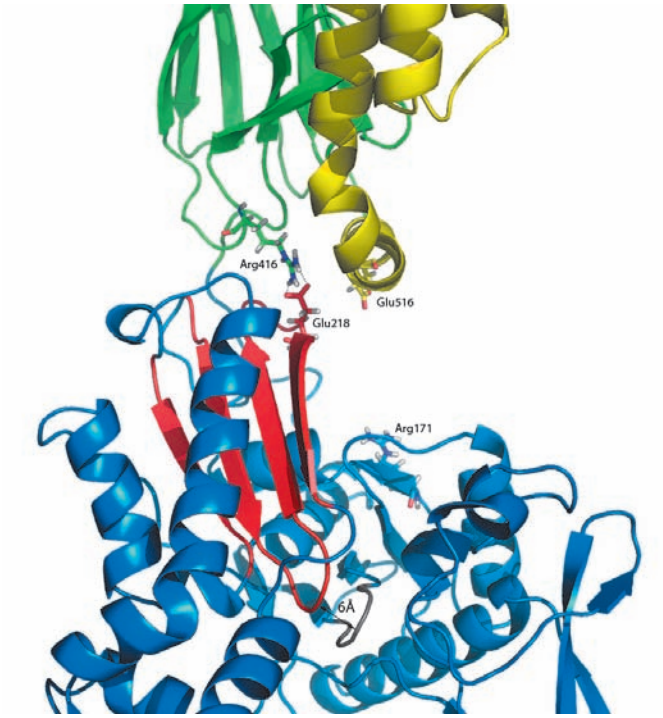


Fig.3. Rearrangement of the salt bridges during hHsp70 domains separation: on the right is the disrupted salt bridge Arg171: Glu516; on the left, the second bridge Arg416: Glu218 before simulation of its break. β-sheet including Glu218 and loop fixing the ATP is colored red. The second loop, which takes part in the fixation, is marked in gray.

As a result, they remain connected only by a cross-domain linker. The process of domain separation was studied using the method of steered molecular dynamics. The salt bridges that are crucial in this process were identified. It was shown that inverse convergence of the domains is not dependent on conformational changes in the ATPase domain.

Disruption of salt bridge Arg171 : Glu516

According to [12], separation of the domains starts with (mediated by cochaperone Hsp40) disruption of the salt bridge Arg171: Glu516, which connects the α-helix 'A' of the substrate-binding domain with the subdomain IA from the ATPase domain. Modeling of this process was conducted using the method of steered molecular dynamics. The Cα-atoms of the residues were pulled apart by 10 Å. The initial distance between the atoms was 12 Å. The amount of energy spent on pulling the residues apart was 34.3 kcal / mol. It is worth noting that after the actual disruption of the salt bridge Arg171: Glu516, the average distance between loops of the ATPase domain fixing the β-and γ-phosphates of ATP decreased by 2 Å (from 8 to 6 Å; Fig. 3). In the native enzyme a similar movement, theoretically, could cause the hydrolysis of ATP. In such a case, the salt bridge Arg171: Glu516 protects the ATPase domain of hHsp70 from spontaneous ATP hydrolysis. The cochaperone disrupting this bond thus initiates the process. The resulting structure was equilibrated for 1 ns. Signs of a restoration of the disrupted salt bridge Arg171: Glu516, as well as further domain separation, were not observed.

Disruption of salt bridge Arg416:Glu218

Disruption of the salt bridge Arg171: Glu516 led only to a partial divergence of the domains, but the final separation has not occurred. In this regard, it was interesting to identify the most stable noncovalent bonds whose disruption could lead to a complete separation of the hHsp70 domains.

Initially, the salt bridge Arg155: Glu523 connecting the α-helix 'A' with the subdomain IA was selected, but its disruption has not led to the domains' separation and has shown no signs of positive development of the process. The negative result led to the conclusion that the salt bridge Arg416: Glu218 is the bond most likely to be key in domain divergence. It links one of the βSBD loops with the subdomain IIA of the ATPase domain. The initial distance between the Cα-atoms of the residues forming the bond was 12.2 Å; after simulation of its disruption, it was - 22.2 Å. The total amount of energy spent on the process was 33.4 kcal/mol. On a curve representing the process, a leap was observed: an increase in distance from 13.4 to 14.7 Å, accompanied by an immediate rupture of the salt bridge, required about 7 kcal/mol (Fig. 4).

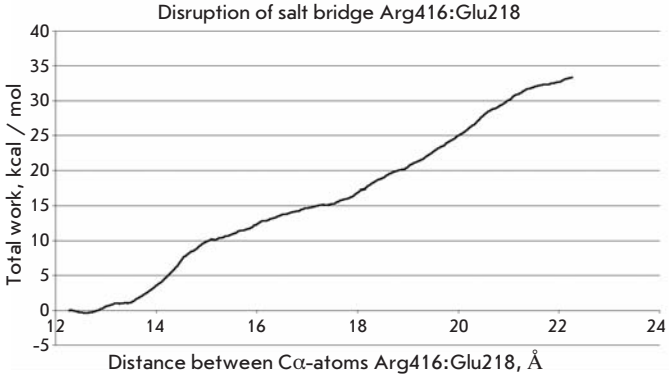


Fig. 4. The work expended on the second stage of the simulation of ATPase and substrate-binding domains separation.

The energy spent on breaking the salt bridge was equal to the energy released in ATP hydrolysis; it could, therefore, be assumed that breakage of the macroergic ATP bond in the ATPase domain (presumably caused in the disruption of the salt bridge Arg171: Glu516) is required to break the salt bridge Arg416: Glu218. Glu218 is part of the β-sheet (formed by the amino acid residues 192 – 226 and 332 – 338; Fig. 5), the second β-turn of which immediately fixes the β- and γ-phosphates of ATP. The first β-turn is involved in the contact of ATPase with substrate-binding domains. Considering the fact that the β-sheet is a rigid structure, its displacement, mediated by ATP hydrolysis, can cause a breakage of the second bond required for the divergence of domains.

As expected, simulation of the breakage of the bond Arg416: Glu218 has led to a complete separation of domains. The resulting structure was equilibrated for 1.5 ns, during which no significant changes in the structure were observed.

An interesting fact was observed in the simulation of domain separation: the residues Arg416 and Glu516 liberated from salt bridges later formed a new salt bridge with each other (see Fig. 5). Formation of a similar salt bridge is sterically possible also between Arg171 and Glu218.

Based on these data we can draw a conclusion on the important role of interdomain salt bridges in the functioning of hHsp70: the salt bridges Arg171: Glu516 and Arg416: Glu218 connect ATPase and substrate-binding domains, but as the domains diverge from each other the partners are rearranged and can form the salt bridges Arg171: Glu218 and Arg416: Glu516, which stabilize the new structural state of the chaperone.

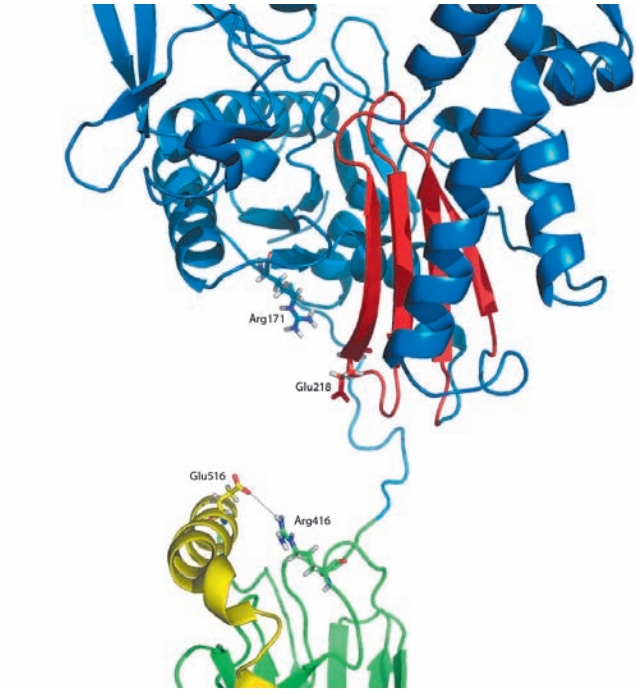


Fig. 5. hHsp70 after the simulation of domains separation. The β-sheet, whose motion is accompanied by the disruption of the second salt bridge Glu218:Arg416, is colored red. The dotted line denotes the salt bridge Glu516:Arg416 formed after domain separation.

Simulation of separation of subdomains I and II

In the process of nucleotide exchange, ATPase and substrate-binding domains converge back, but the force involved in this process has yet to be established. A stochastic nature of domain convergence seems unlikely. Because nucleotide exchange occurs while domains are separated and is accompanied by the interaction of the ATPase domain with the nucleotide exchange factor, it is reasonable to assume that pulling the subdomains I and II apart by the nucleotide exchange factor, theoretically, can affect the convergence of ATPase and the substrate-binding domains. We conducted a simulation of the process. In simulating pulling apart subdomain I from subdomain II, points of application of force were taken to be the nitrogen atoms belonging to residues Thr14 and Gly203, directly interacting with ATP / ADP. The initial distance between them was 7.3 Å; the final one, 14.3 Å. Despite the theoretical possibility that nucleotide exchange affects the convergence of domains, we obtained a predictable result: change in the distance between the subdomains I and II did not affect the interaction of the domains with each other. This result suggests that the convergence of domains is most likely mediated by some protein which may be another chaperone hHsp70. Theoretically, it can bind the hydrophobic linker released in domain divergence as a substrate and thus bring the diverged domains closer. However, this hypothesis requires further verification.

CONCLUSIONS

Basing on research, the following scheme of hHsp70 functioning can be suggested: at the interaction with the J-domain of cochaperone Hsp40, the salt bridge Arg171: Glu516 is broken. This leads to a convergence of the loops fixing ATP and promotes ATP hydrolysis. ATP hydrolysis, in turn, is accompanied by a shift of the β-sheet containing one of the loops fixing ATP and a shift of the residue Glu218 that is located in the site of interdomain contact, which leads to the disruption of the salt bridge Arg416: Glu218. Disruption of the salt bridge Arg416: Glu218 results in separation of the ATPase domain from the substrate-binding domain.

Arg416 forms a salt bridge with Glu516 that belongs to the substrate-binding domain. Arg171, in turn, forms a new salt bridge with Glu218. Thus, regrouping of the salt bridges that previously connected ATPase and the substrate-binding domains of hHsp70 occurs.

The role of the nucleotide exchange factor in the functioning of hHsp70, apparently, is not only in performing the exchange of ADP for a new molecule of ATP, but also in the fixation of the subdomains I and II in separated states, which is accompanied by the convergence of ATPase and substrate-binding domains with formation of the salt bridge Arg171: Glu516 that can be figuratively described as "cocking the hammer" for the next round of ATP hydrolysis. The "finger" that can pull the trigger is the J-domain of cochaperone Hsp40.

An equilibrated model of human Hsp70 can be found in the PMDB depositary [20, 21] (PMDB id: PM0076412). ●

This work was supported by the Russian Foundation for Basic Research (Joint project RFBR-DFG RECESS, grant 09-04-92744).

REFERENCES:

1. Multhoff G., Pfister K., Gehrmann M., Hantschel M., Gross C., Hafner M., Hiddemann W. // Cell Stress Chaperones. 2001. V. 6. № 4. P. 337–344.
2. Schmitt E., Gehrmann M., Brunet M., Multhoff G., Garrido C. // J. Leukoc. Biol. 2007. V. 81. № 1. P. 15–27.
3. Vogel M., Bukau B., Mayer M.P. // Mol. Cell. 2006. V. 21. № 3. P. 359–367.
4. Nollen E.A., Kabakov A.E., Brunsting J.F., Kanon B., Höhfeld J., Kampinga H.H. // J. Biol. Chem. 2001. V. 276. № 7. P. 4677–4682.
5. Guzhova I.V., Novoselov S.S., Margulis B.A. // Tsitologiia (Cytology). 2005. V. 47. №3. P. 187-199.
6. Osipiuk J., Walsh M.A., Freeman B.C., Morimoto R.I., Joachimiak A. // Acta Crystallogr. D Biol. Crystallogr. 1999. V. 55. № 5. P. 1105–1107.
7. Zhu X., Zhao X., Burkholder W.F., Gragerov A., Ogata C.M., Gottesman M.E., Hendrickson W.A. // Science. 1996. V. 272. № 5268. P. 1606–1614.
8. Chang Y., Sun Y., Wang C., Hsiao C. // J. Biol. Chem. 2008. V. 283. № 22. P. 15502–15511.
9. Freeman B.C., Myers M.P., Schumacher R., Morimoto R.I. // EMBO J. 1995. V. 14. № 10. P. 2281–2292.
10. Mitra A., Shevde L.A., Samant R.S. // Clin. Exp. Metastasis. 2009. V. 26. № 6. P. 559–567.
11. Chou C., Forouhar F., Yeh Y., Shr H., Wang C., Hsiao C. // J. Biol. Chem. 2003. V. 278. № 32. P. 30311–30316.
12. Jiang J., Maes E.G., Taylor A.B., Wang L., Hinck A.P., Lafer E.M., Sousa R. // Mol. Cell. 2007. V. 28. № 3. P. 422–433.
13. Jiang J., Prasad K., Lafer E.M., Sousa R. // Mol. Cell. 2005. V. 20. № 4. P. 513–524.
14. Mazzaferro V., Coppa J., Carrabba M.G., Rivoltini L., Schiavo M., Regalia E., Mariani L., Camerini T., Marchianò A., Andreola S., Camerini R., Corsi M., Lewis J.J., Srivastava P.K., Parmiani G. // Clin. Cancer Res. 2003. V. 9. № 9. P. 3235–3245.
15. Arnold K., Bordoli L., Kopp J., Schwede T. // Bioinformatics. 2006. V. 22. P. 195–201.
16. Kiefer F., Arnold K., Künzli M., Bordoli L., Schwede T. // Nucl. Acids Res. 2009. V. 37. P. 387–392.
17. Case D.A., Darden T.A., Cheatham T.E., Simmerling C.L., Wang J., Duke R.E., Luo R., Crowley M., Walker R.C., Zhang W., Merz K.M., Wang B., Hayik S., Roitberg A., Seabra G., Kolossvary I., Wong K.F., Paesani F., Vanicek J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Yang L., Tan C., Mongan J., Hornak V., Cui G., Mathews D.H., Seetin M.G., Sagui C., Babin V., Kollman P.A. // AMBER 10. San Francisco: University of California, 2008.
18. http://www.ks.uiuc.edu/Research/vmd/vmd-1.8.6/
19. Worrall L.J., Walkinshaw M.D. // Biochem. Biophys. Res. Commun. 2007. V. 357. P. 105–110.
20. http://mi.caspur.it/PMDB/main.php
21. Castrignanò T., De Meo P.D., Cozzetto D., Talamo I.G., Tramontano A. // Nucl. Acids Res. 2006. V. 34. P. 306–309.

Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets

R. V. Reshetnikov[1,4], A. M. Kopylov[2,3,4], A. V. Golovin[1,4]*
[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University
[2]Faculty of Chemistry, Lomonosov Moscow State University
[3]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University
[4]Apto-Pharm LLC
*E-mail: golovin@belozersky.msu.ru
Received 8.09.2010

ABSTRACT **The present work is devoted to the analysis of the G-quadruplex DNA structure using the bioinformatics method. The interest towards quadruplex DNAs is determined by their involvement in the functioning of telomeres and onco-promoters as well as by the possibility to create on their basis aptamers and nanostructures. Here, we present an algorithm for a general analysis of the polymorphism of the G-quadruplex structure from the data bank PDB using original parameters. 74 structures were grouped according to the following parameters: the number of DNA strands, the number of G-quartets, and the location and orientation of the connecting loops. Two quantitative parameters were used to describe the quadruplex structure: the twist angle between two adjacent quartets (analogous to that for the complementary pair in the duplex DNA) and the quartet planarity (an original parameter). The distribution patterns of these values are specific for each group of quadruplex structures and are dependent upon the type of connecting loops used (diagonal, lateral or propeller). The tetramolecular loopless parallel quadruplex was used as a comparison template. The lateral loops introduce the strongest distortion into the structure of quadruplexes: the values of the twist angles are the lowest and are not typical for the other quadruplex groups. The loops of the diagonal type introduce much weaker deformation into quadruplexes; the structures with propeller loops are characterized by the optimum geometry of G-quartets. Hence, the correlation between the twist angle and the tension in the structure of quadruplex DNA is revealed.**
KEYWORDS **G-quadruplex, G-quartet, twist angle, loops, structure.**

INTRODUCTION

G-quadruplexes

It is known that the strands of guanosine oligo- and polynucleotides are capable of aggregating with each other, on condition that the solution contains a monovalent cation, such as potassium or sodium. X-ray diffraction analysis revealed that these poly(G)-strands represent a novel type of DNA folding, a four-strand helix [1–3], where four guanine bases belonging to different strands form a planar structure retained by G-G pair interactions (Fig. 1). These structures are notable for their high stability and are known as guanine (G)-quartets, or G-tetrads. Each G-quartet is bound by a total of eight hydrogen bonds, which are formed as a result of the interaction between the Watson-Crick side of one guanine base and the Hoogsteen side of another one.

One of the fundamental reasons the nucleic acids containing the G-tetrad motif are of interest is that the guanine-rich sequences are widely represented in all the genomes that have been discovered to date. These motifs were found in the promoter regions and immunoglobulin switch regions, recombination hotspots, etc. [4]. G-quartets are also represented in DNA at the ends of eucariotic chromosomes known as telomeres [5]. Telomeric DNAs are tandem repeats of short poly-G-blocks, sometimes comprising adenine or thymine bases: G_nT_n, $G_nT_nG_n$, G_nA_n or $(TTAGGG)_n$; telomeric DNAs are associated with telomeric proteins. The repeat types are species-dependent; for instance, the $(TTAGGG)_n$ repeat is typical of mammals. The function of the telomeres is to protect the chromosomal ends from damage under recombination or action of nucleases. Human telomeric DNA in somatic cells contains, on average, 8–10 kbp. The terminal 100–200 nucleotides at the 3'-end represent a conformationally unrestricted single-strand "tail" [6]. In living cells, the "tail" is associated with POT1 protein [7]; while in the absence of this protein the single-stranded DNA is capable of folding and dimerizing to form four-stranded hairpins, which can be stabilized by the formation of guanine tetrads [8,

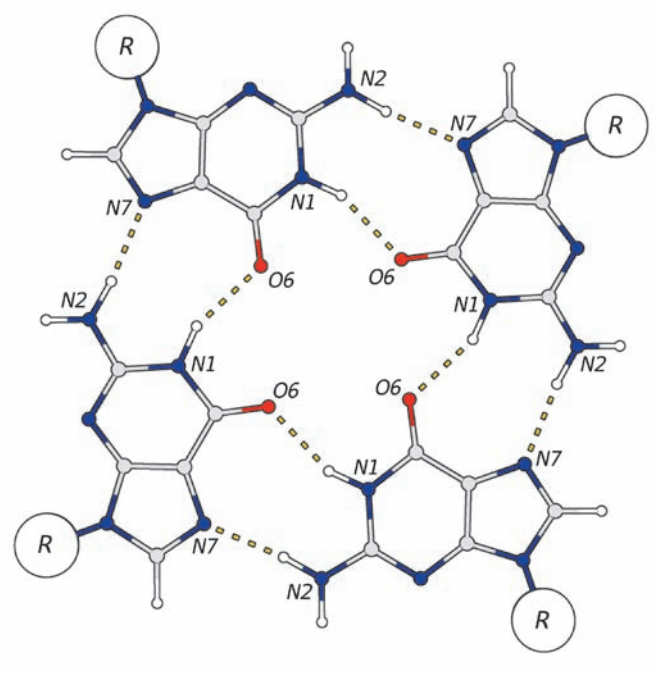


Fig. 1. Scheme of G-quartet. Four guanine residues form a square coplanar structure, where each nucleic acid base is a donor and acceptor of a hydrogen bond: N1 and N2 from one side of the heterocycle, O6 and N7 from the other side of the guanosine involved in formation of 8 hydrogen bonds in one quartet. Sugar-phosphate backbone of nucleic acids is denoted as R.

9]. An alternative method for the stabilization of this type of DNA consists in the formation of intramolecular G-quartets by repeated foldback. These G-quartet-containing structures are known as quadruplexes or tetraplexes [10]. Aside from telomeres, G-quadruplex sequences were localized in promoters of a number of oncogenes and cancer-associated genes, such as *k-ras* [11], *c-kit* [12], and *bcl2* [13]. Thus, the possibility of inhibition of the corresponding gene expression using such quadruplex-specific agents as porphyrin TMPyP4 [14, 15] appears to be encouraging.

G-quadruplexes can also be formed by short oligonucleotides with the corresponding sequence, which can be written as $G_mX_nG_mX_oG_mX_pG_m$, where m is the number of guanine per G-block. These guanines are generally directly involved in G-tetrad formation. X_n, X_o and X_p can be a combination of any residues, including G; these regions form loops between the G-tetrads.

Some of these sequences exhibit aptameric properties upon folding into quadruplex structures. Aptamers are short synthetic oligonucleotides or peptides, being functional analogs of monoclonal antibodies, and are capable of specifically recognizing a wide range of targets, from small molecules to whole cells [16, 17]. G-quadruplex aptamers targeting a wide range of proteins, such as thrombin [18] and STAT-3 [19], have been identified. There are G-quadruplex aptamers with anti-cancer properties, which are currently undergoing clinical trials. Their mechanism of action is connected with the nucleolin protein and its role in RNA processing [20].

Common structural features of quadruplex DNAs

Depending on the number of G-blocks, the formation of a quadruplex structure from a specified quadruplex sequence may occur via different paths. Four individual strands may associate with each other to form an intermolecular G-quadruplex (Fig. 2). Intramolecular tetraplexes are formed from a single-stranded molecule as a result of the complex spatial folding of the nucleotide strand (Fig. 3).

The G-quartet is a fundamental structural unit of all quadruplex structures. Within the quadruplex structure, the quartets are located one above the other; a minimum of two quartets are required for the structural stability of the tetraplex [21]. The number of guanines in each individual G-block is directly related to the number of G-tetrads in the final folded quadruplex. For example, in telomeric DNA found in mammals with tandem repeat d(TTAGGG), the quadruplexes formed by four of these repeats have three G-quartets located one above the other [22].

The factors that stabilize the quadruplex are the same as those that stabilize the duplex DNA; including the base stacking interaction, hydrogen bonds, electrostatic interactions, and the hydration shell. The hydration of the sugar-phosphate backbone plays a crucial role in the stability of the structure: within an ordered hydration shell, the water molecules with an extensive network of hydrogen bonds combine into a single unit with the bases, sugars, and charged phosphates which are localized on the outer surface of the quadruplex. [25–27].

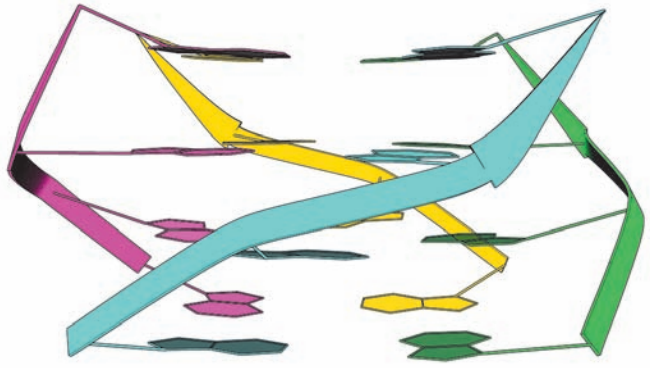


Fig. 2. Four-stranded intermolecular parallel G-quadruplex.

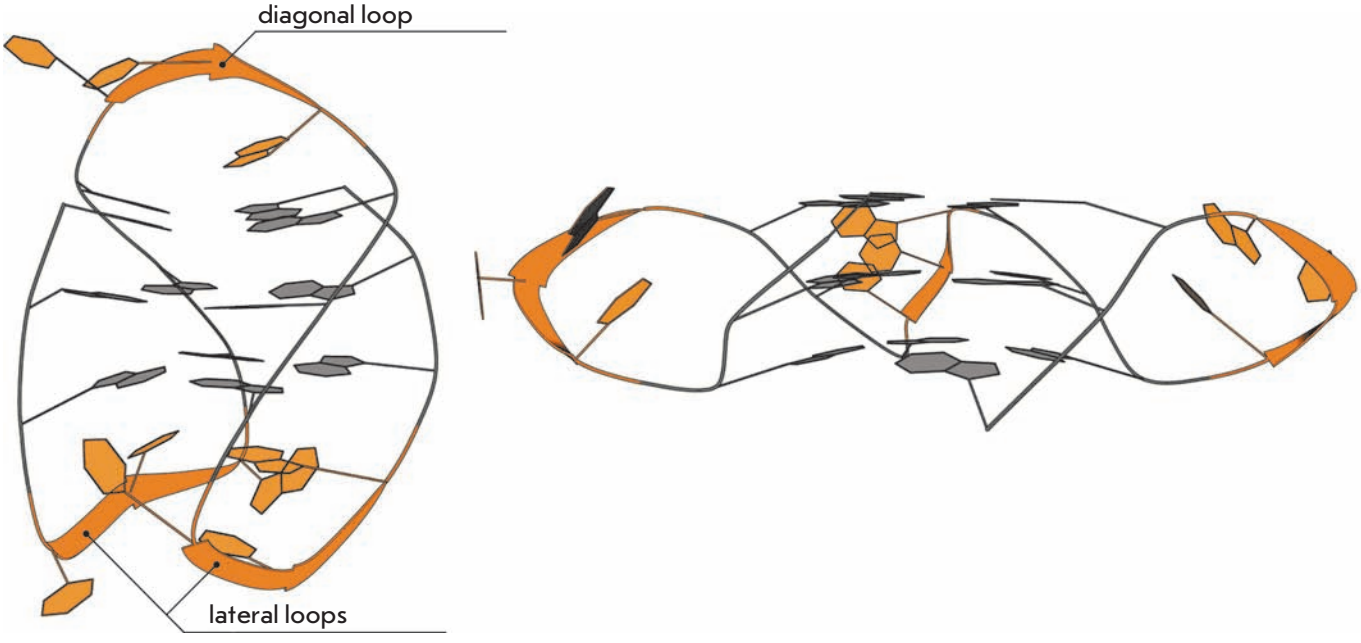


Fig. 3. Folding of a DNA strand into a monomolecular G-quadruplex with three G-quartets. Two variants of oligonucleotide d(AGGG(TTAGGG)$_3$) differ from each other in the direction of the polynucleotide chain in different strands of the quadruplex. Left: topology with loops of lateral and diagonal type [23]; right: topology with strand-reversal loops of propeller type [24].

In addition to these stabilizing factors being standard for the duplex DNA, the quadruplexes have a highly specific component: a significant contribution to stability is made by the coordination of O6 carbonyls by cations [28, 29]. Atoms O6 form a square in each quartet, forming a bipyramidal antiprism in the quadruplex with an interquartet distance of 3.3 Å [30]. This negatively charged space between the tetrads should be stabilized by coordination of a cation. The cation's nature (size/ionic radius and charge) has a considerable effect on the stability of the resultant quadruplex [29].

Four guanosine nucleosides within a tetrad can exist in either *anti*- or *syn*-conformation with respect to the glicoside bond; thus providing 16 possible combinations. The mutual orientation of the individual strands within a quadruplex has an effect on glicoside angles. For instance, with all four strands oriented in parallel (Fig. 2), all glicoside angles have an *anti*-conformation. For anti-parallel orientation, the quadruplexes contain both *syn*- and *anti*-guanines, regardless of whether the quadruplex is four- or single-stranded.

Various nucleotide sequences between G-blocks form extrahelical loops. These loops may be of three types. The parallel intramolecular quadruplex requires a loop that would connect the bottom G-tetrad with the top one, resulting in the formation of propeller-type loops (Fig. 3, right). Anti-parallel quadruplexes are ones in which at least one strand is located antiparallel to the other strands. Such topology of tetrads was revealed in most of the currently identified bimolecular and intramolecular quadruplex structures. In addition to the propeller-type loops, loops of two more types are observed in these structures. The adjacent G-strands are linked by the lateral (endgewise) loops. Two loops of this type can be located both at the same and opposite poles of the molecule, corresponding to the "head-to-head" or "head-to-tail" arrangement in bimolecular complexes. The second type of anti-parallel loop is represented by the diagonal loops connecting the opposite G-strands (Fig. 3, left).

All quadruplex structures have four grooves, unlike the double helix with only two grooves. The grooves are formed by the cavities restricted by sugar-phosphate backbones. The groove's dimensions widely vary depending on the total topology and nature of loops, and on glycoside angles, as well. In the quadruplexes with loops of only diagonal or lateral type, the grooves are structurally simple, while in the structures with propeller-type loops they possess more complex structural features.

Thus, there is a large number of structural variables (the number of G-tetrads; type, sequence, length, and orientation of loops), which leads to a wide topological and structural variety of quadruplexes. In this work, we attempted to reveal the interrelation between the

structure and properties of G-quadruplex DNAs and determine the factors that have an influence on a quadruplex's geometry.

METHODS

Sampling

The PDB data bank was used to compile the list of quadruplex-containing structures. All found structures were divided into 8 groups according to the geometry of spatial organization of a quadruplex. On the basis of the Perl programming language and modules Vector::Real and Statistics::Descriptive, we developed a software tool that would determine the presence of quartets in the structure, reveal their location, and measure their geometrical parameters. The quartet is determined as follows: a guanine should have a neighbor in contact with the O6 atom of the initial guanine via the N7 atom. The combination selected is recognized as a quartet on condition that the fourth guanine interacts with the first one, and all heterocyclic bases are located in one plane with the maximum permissible offset of the atoms from the surface being equal to 2 Å. The nearest quartet from the initial one with the distance between C1' atoms from the nearest nucleotides being no more than 10 Å was recognized as the next quartet of the quadruplex (thus, the variants with guanines from different quartets forming a tetrad were eliminated). The structures belonging to different NMR models were processed as independent quadruplexes.

The twist angle between two adjacent quartets in a quadruplex

In order to determine the quadruplex twist angle, the angle between the two vectors was measured. The first vector joined the C1' atoms of two adjacent nucleotides in a quartet, while the second vector joined the C1' atoms of the corresponding nucleotides in the adjacent quartet (Fig. 4).

The out-of-plane deviations of the quartet

The distance between the centers of mass of two tetragons (an original parameter) was proposed as a method for measuring the degree of disturbance of symmetry and planarity for an individual quartet. The first tetragon is formed by four N9 atoms of guanine in the quartet; the second tetragon is formed by four O6 atoms of the same guanines (Fig. 4). If the quartet is symmetrical and all guanines form hydrogen bonds with each other, this parameter fixes the planarity of the quartet. If the hydrogen bonds break, the symmetry of the quartet is disturbed; the parameter fixes the degree of quartet distortion, which includes deviations in both symmetry and planarity of a quartet.

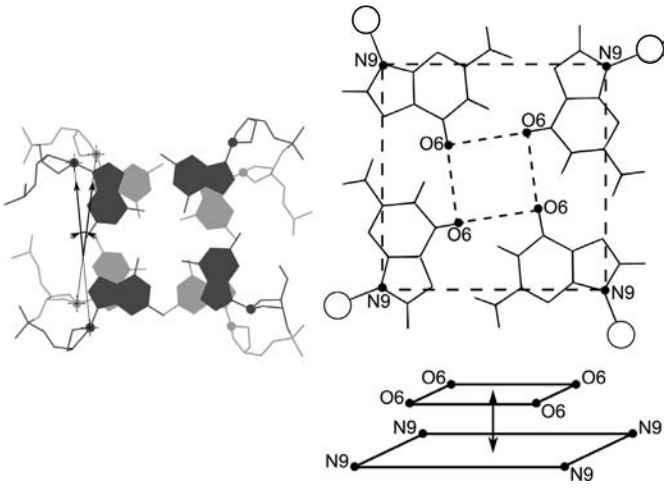


Fig. 4. Definition of the measured parameters. Left: Definition of the quadruplex twist angle. C1′ atoms are shown as spheres. Right: relative positions of two tetragons: the outer, formed by N9 atoms, and the inner, formed by O6 atoms. If the planarity of the quartet is disturbed, then the inner tetragon leaves the quartet plane. Distance between the centers of mass of the inner and outer tetragons is the numeric parameter of planarity deviation.

The histogram of distribution of the quadruplex twist angles and out-of-plane deviations of the quartet

For each of the eight groups obtained, the values of the parameter were combined to plot a distribution histogram. The twist angle range was selected from 0 to 60°, the out-of-plane deviations were analyzed within a range from 0 to 2 Å, the ranges of angles and distances were partitioned into 15 intervals. The data were analyzed using Gnuplot software (http://www.gnuplot.info).

RESULTS AND DISCUSSION

Parameters for description of geometry and conformational polymorphism of quadruplexes

Even a cursory analysis of the diversity of structure variants of the quartets detected in the PDB data bank leads to the conclusion that there is a necessity for systematization and development of universal parameters for the description of quartet structures and their polymorphism. Until now, no systematization attempts have been published.

In this work, two parameters were selected as structural characteristics of the quadruplexes. The first parameter was the twist angle of the quadruplex; i.e., the angle between two adjacent quartets, which is described by the angle between two vectors passing through the C1' atoms of two adjacent guanines (Fig. 4).

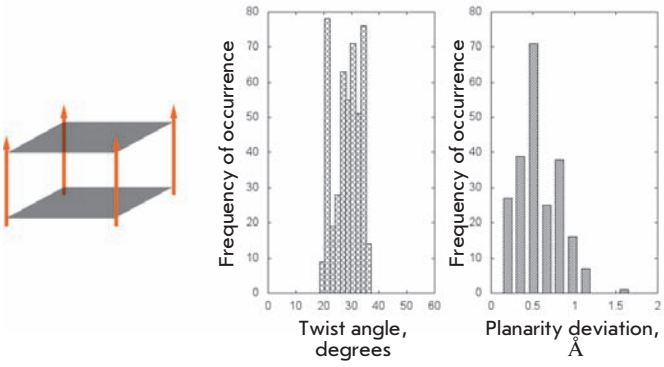


Fig. 5. Structural organization and geometrical features of the first group of structures.

This parameter has been widely used for the description of double-helix structures [31]. Previously it was demonstrated that the quadruplex twist angle represents the degree of tension of the oligonucleotide structure [32]. The second parameter was represented by an original parameter describing the planarity deviation of quartets; i.e., the distance between the centers of mass of two tetragons formed by the O6 and N9 atoms, respectively (Fig. 4). These two parameters make it possible to provide an adequate description of the conformational polymorphism and conformational mobility of the structure of quadruplex DNA.

Four-stranded parallel quadruplexes

Composition of the group (ID PDB):

- 1EVM [33], 1EVN [33], 1NP9 [34], 1NZM [35], 1O0K [36] – telomeric DNA (human);
- 139D [23, 37] – telomeric DNA (*Tetrahymena*);
- 1EMQ [38] – telomeric DNA (*Saccharomyces cerevisiae*);
- 1EVO [39] – fragment of SV40 viral genome.

The group under consideration is the simplest variant of the arrangement of quadruplex structures. The twist angle in such structures may be regarded as the ideal twist angle, since the association of four strands imposes no structural restrictions at all, something not observed for the bimolecular and monomolecular quadruplexes.

Four-stranded parallel quadruplexes are characterized by a wide spectrum of twist angles (Fig. 5), with two ranges of preferred values: the narrow range corresponds to 21° and the more diffuse range lies within 27°–34°. In addition, the planarity of the quartets with this structure varies. In most cases, the quartets have a small out-of-plane deviation amounting to 0.5 Å; however, the maximum deviations are above 1 Å. The determined polydispersity of parameters illustrates the wide range of possibilities of the conformational polymorphism of four-stranded parallel quadruplexes without any structural restrictions for their formation upon intermolecular association.

Chair-type structure

Composition of the group:

- 148D [40], 1C32 [41], 1C34 [41], 1C35 [41], 1C38 [41], 1QDF [42], 1QDH [42], 1RDE [43] – thrombin-binding DNA aptamer;
- 2KM3 [22] – telomeric repeat CTAGGG (human).

The chair-type structure is a monomolecular quadruplex which is connected by three lateral loops. It is a sufficiently unique structure represented by only two molecules: a 15-mer thrombin-binding DNA aptamer (148D, 1C32, 1C34, 1C35, 1C38, 1QDF, 1QDH, 1RDE) and a 22-mer oligonucleotide, which is formed by the telomeric repeat CTAGGG (2KM3). Such structures are characterized by significantly smaller twist angles in comparison with those of the preceding group of "loopless" intermolecular four-stranded quadruplexes (Fig. 6). The mean value of quadruplex twist angles for the chair-type structures is equal to 15° with a deviation of ± 5°. The quartets with such structure are characterized by a high planarity; the jump in values within the range of 0.8–0.9 Å corresponds to the structures that can exist in high ionic strength solutions. The planarity results from the fact that the heterocyclic bases of the

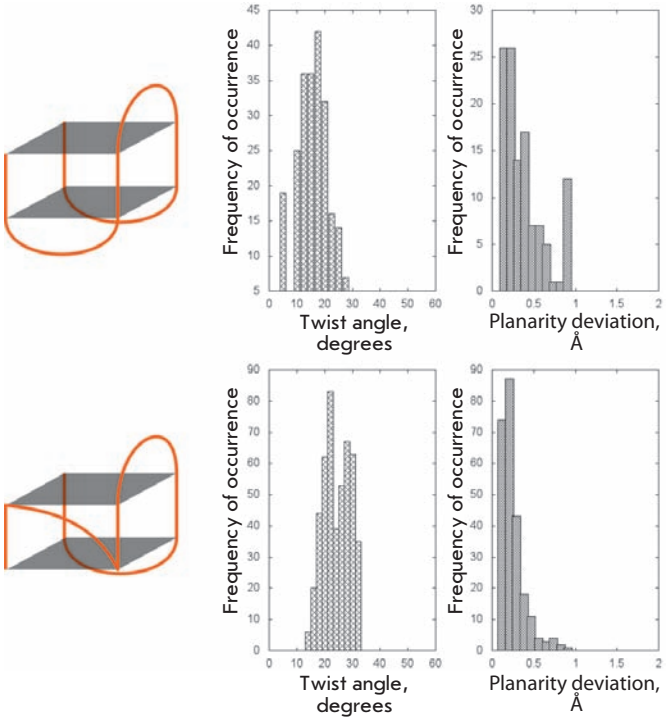


Fig. 6. Structural organization and geometrical features of the second and third groups of structures.

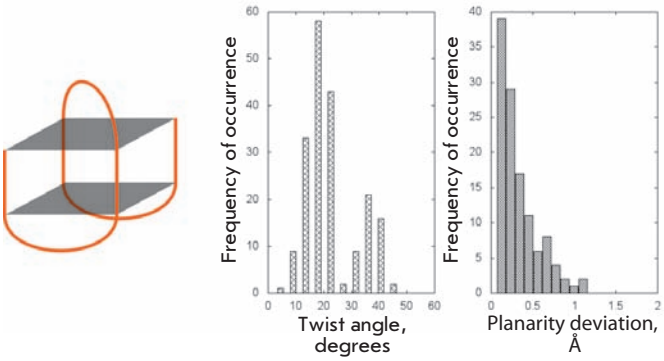


Fig. 7. Structural organization and geometrical features of the fourth group of structures.

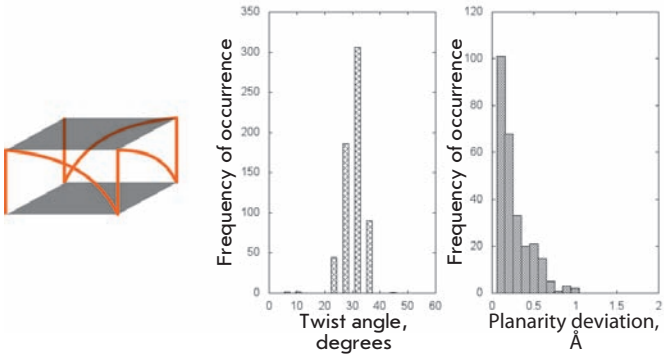


Fig. 8. Structural organization and geometrical features of the fifth group of structures.

lateral loops form stacking interactions with the bases of the quartets, thereby limiting their out-of-plane deviations.

Monomolecular quadruplexes with (3+1) strand topology

Composition of the group:

- 2JSK [44], 2JSQ [44], 186D [45], 2GKU [46], 2HY9 [47], 2JPZ [48], 2JSL [44], 2JSM [44] – telomeric DNA (human, *Tetrahymena*);
- 2F8U [49] – BCL2 promoter (human).

These quadruplexes comprise two lateral loops and one propeller-type loop, which reverses the direction of the polynucleotide strand. This group differs from the preceding quadruplex type only by the existence of a propeller strand, which significantly affects the values of the twist angle (Fig. 6). An additional distribution with a maximum at 28° ± 4° emerges in the pattern of the angle distribution. Furthermore, the values of the twist angles within the range of 10°–20°, characteristic of quadruplexes with lateral loops, are shifted toward higher values by 2°–4°. Quadruplex quartets of this type are highly planar structures. Similar to the quartets in the previous case, the planarity is fixed by stacking interactions with heterocyclic bases of lateral loops.

Basket-type structure

Composition of the group:

- 2KF8 [50], 2KKA [51], 2KOW [52], 143D [23], 230D [53], 201D [54] – telomeric DNA (human, *Oxytricha*).

The basket-type structures are represented by the monomolecular quadruplex connected by two lateral and one diagonal loops. As shown in the previous case, the substitution of a lateral loop by a diagonal one results in change in the character of the twist angle distribution and the emergence of two strongly pronounced shoulders with mean values of 18° ± 4° and 36° ± 4° (Fig. 7). Thus, a new set of structures, characterized by quadruplex twist angles higher than the maximum angles in loopless four-stranded parallel quadruplexes, emerging in the conformational landscape are conditioned by the emergence of a diagonal loop. The degree of planarity of basket-type quartet structures is lower than that of the earlier discussed structures with lateral loops, which is likely the result of a decrease in stacking interactions between the heterocyclic bases of the diagonal loop and the top quartet.

Monomolecular parallel quadruplexes with propeller loops

Composition of the group:

- 1KF1 [24], 3CDM [55] – telomeric DNA (human);
- 1XAV [56], 2A5P [57], 1A5R [57] – c-MYC promoter (human);
- 2KQG [58], 2KQH [58], 2KYP [59] – c-kit oncogene promoter (human);
- 1MYQ [60] – synthetic oligonucleotide $(GGA)_4$;
- 1Y8D [61] – aptamer-targeting HIV-1 integrase.

These unusual structures represent intramolecular quadruplexes where all loops are of the propeller type and the polynucleotide strand changes its direction thrice. The presence of propeller loops strictly determines the quadruplex structure; the distribution of twist angles has a pronounced maximum at 31° ± 3° (Fig. 8). This value is similar to that determined for loopless four-stranded parallel quadruplexes. Quadruplexes of this type have planar quartets. It is not unlikely that propeller loops ensure the optimal geometry for quadruplexes with a rigid sugar-phosphate backbone.

Bimolecular quadruplexes with lateral loops

Composition of the group:

- 1A8N [62], 1A8W [63] – tandem repeat GGGC;
- 1F3S [64] – synthetic oligonucleotide.

Unlike the monomolecular quadruplexes that are formed by intramolecular folding, bimolecular quadruplexes are formed upon dimerization of two self-folded

polynucleotide strands which contain guanine blocks. The members of this quadruplex group are notable for the relatively long length of lateral loops (5–6 nucleotides). They can be regarded as an intermediate group between monomolecular quadruplexes with lateral loops and quadruplexes with diagonal loops. The values of the twist angles in these quadruplexes, lying in the region between two extrema belonging to the adjacent groups, argue for an intermediate position for this group as well (Figs. 6, 7, 9). It should be noted that the structures under consideration have a tendency to form 20° ± 1° and 27° twists. A higher statistical significance is required in order to draw firmer conclusions. The quartets in quadruplexes of this type appear to be highly planar due to the intense stacking interactions with the heterocyclic bases of the loops.

Bimolecular quadruplexes with diagonal loops

Composition of the group:

- 156D [53, 65], 1JPQ [25], 1L1H [66], 1QDI [42], 1QDK [42], 3EM2 [67], 3EQW [67], 3ERU [67], 3ES0 [67], 3ET8 [67], 3EUM [67], 2AKG [68], 1K4X [53], 1JRN [25], 2HBN [69], 3EUI [67] – telomeric DNA (*Oxytricha*);
- 2KAZ [70], 1U64 [71], 1LVS [72], 1FQP [73] – synthetic oligonucleotides.

The molecules of this group form bimolecular quadruplexes with diagonal loops between the opposite angles of the quadruplex. Thus, these quadruplexes resemble the earlier described basket-type structures with two lateral loops substituted by a diagonal loop. In terms of the values of the twist angles, this substitution leads to the elimination of the shoulder with a maximum at 35°, retaining only one maximum at 19° ± 4° (Fig. 9). The distribution of the values of quartet planarity also resembles that of the basket-type structures.

Bimolecular parallel quadruplexes with propeller loops

Composition of the group:

- 1K8P [24], 2HRI [74], 3CE5 [75] – telomeric DNA (human); 2KYO [59] – c-kit oncogene promoter (human);
- 1NYD [76], 1EEG [77], 1XCE [78] – synthetic oligonucleotides.

For this type of quadruplexes, as well as for their monomolecular analog, propeller loops restrict the conformational polymorphism of the quadruplex: the distribution of the twist angle values has a maximum at 31° ± 3°; the quartets are planar (Fig. 10).

Classification of quadruplex structures

G-quadruplex structures are localized at the ends of telomeric regions of DNA and promoters of a number of oncogenes and cancer-associated genes. This fact makes the quadruplexes an attractive target for anti-cancer chemotherapy. The most preferable model of interaction between the anti-cancer agents and quadruplex DNA is based on stacking interactions with the quartets (terminal, or via intercalation), on one side, and interactions with the grooves, on the other side. Thus, thorough knowledge of the geometry of these elements and the factors that can influence them is a crucial part in a rational search for new efficient anti-cancer agents.

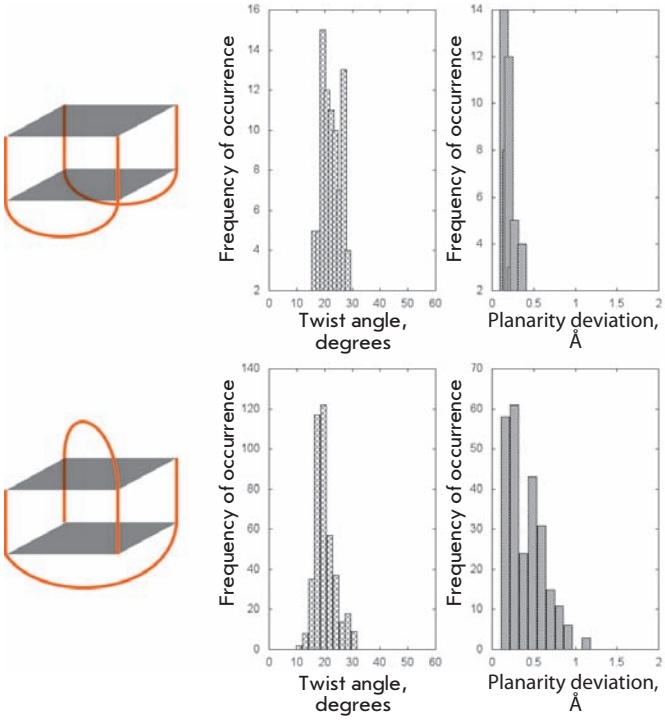


Fig. 9. Structural organization and geometrical features of the sixth and seventh groups of structures.

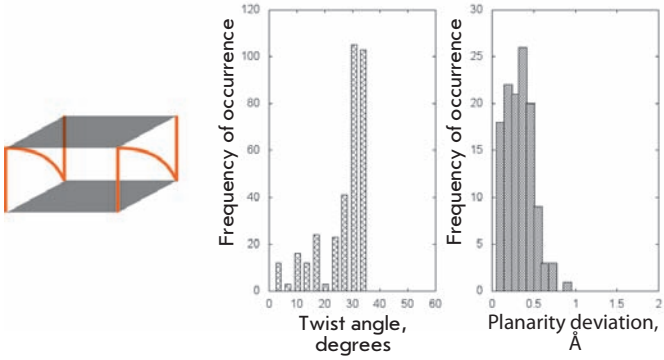


Fig. 10. Structural organization and geometrical features of the eight group of structures.

Nucleic acid quadruplexes are of interest not only as targets for anti-oncogenic drugs, but also as the structural core of aptamer-based therapeutic agents. In particular, this refers to aptamers targeting thrombin, HIV-1 integrase, the tumor marker – nucleolin protein which is involved in RNA processing. In order to optimize the stability and efficient self-assembling of these oligonucleotides, it is necessary to have knowledge pertaining to the nature of the forces able to have stabilizing or destabilizing effects on the aptamer molecule.

A wide topological and structural diversity conditioned by such variables as the number of G-quartets, the orientation, type, sequence and length of the loops is typical for quadruplexes;. For the description of this conformational ensemble, two parameters were selected: the twist angle between two adjacent quartets and the planarity of the quartets.

In our study it was shown that these parameters successfully characterize the most complex quadruplex structures. It should be noted that by the term "planarity" we refer to the generalization of two phenomena. When the quartet is symmetrical and all guanines form hydrogen bonds with each other, it is the planarity of the quartet that is fixed by this parameter. In the case of hydrogen bond breaking, the symmetry of the quartet is violated, and the parameter shows the degree of quartet distortion.

A four-stranded intermolecular parallel quadruplex is the simplest case, where only one of the specified parameters (the number of G-quartets) is realized. This variable determines the range in which these structures can exist without losing their integrity. This range is appreciably wide; in terms of the twist angle it stretches 19°–36°, with two regions of preferred values: 21° and 27°–34°. The motion of the terminal quartets in such structures is limited only by stacking interactions with the adjacent quartets and by coordination bindings with stabilizing cations if they exist in this case. Thus, G-quartets of these structures, in particular terminal ones, do not exhibit pronounced planarity.

Chair-type structures dramatically differ from the loopless four-stranded quadruplexes. The lateral loops of these structures restrict the geometry of quadruplexes; furthermore, they desrease the twist angles to values that are not characteristic of loopless quadruplexes. Whereas the minimum angles for loopless quadruplexes are equal to 19°, in contrast the structures containing only lateral loops have a range of preferred values that lies within 15° ± 5°. This points to the fact that considerable tension is introduced into the structure of the quadruplex by the lateral loops, primarily because of their small length. Meanwhile, on account of the location features evident from their name, loops of this type interact eagerly with the terminal quartets of quadruplexes at the terminal poles of G-quadruplexes via the formation of stacking interactions, which has a positive effect on the planarity of chair-type quartets.

It is of interest that the addition of diagonal or propeller loops to the chair-type structure has a considerable effect on the distribution pattern of twist angles. The preferred range of angle values shifts toward the higher values, typical of loopless parallel G-quadruplexes. In addition, a second shoulder emerges on the pattern, characterizing the region being adjacent to a loop other than the lateral one.

In basket-type structures the strength of the tension induced by the lateral loops located at one pole of the molecule is clearly visible. The moment of these forces turns the molecule region, which is adjacent to the diagonal loop, into a range of twist angles of 36° ± 4°. That is more than the required norm for loopless parallel quadruplexes; however, the diagonal loop compensates for this abnormality through its length and elasticity. It should be noted that these values of twist angles are not typical of structures with diagonal loops only. The quadruplex geometry is rather rigidly restricted by angle values of 19° ± 4° in these structures, which is still closer to values that are reasonable for loopless quadruplexes. Thus, it could be concluded that the tensions induced by the lateral loops are stronger than those induced by the diagonal loops. It is noteworthy that the *syn*/*anti*-conformation pattern, which has been frequently used for the description of quadruplexes, is represented by the largest number of combinations in the group of bimolecular quadruplexes with diagonal loops. Conversely, this group exhibits the smallest dispersion of quadruplex twist angles. This observation allows one to assume that the loop type is the dominant influence factor on the quadruplex structure.

The propeller loops determine the geometry of a quadruplex in the strictest manner. The values of twist angles fluctuate within 31° ± 3° both for monomolecular and bimolecular tetraplexes only with propeller loops. Even the introduction of lateral loops into such structures has a weak effect on the twisting of the regions adjacent to the propeller loops; it is characterized by a range of 28° ± 4°. These parameters agree with the second range of preferred values for the loopless quadruplexes (27°–34°). It seems likely that these twist angles correspond to the optimum geometry of a G-quartet, since the tetrads of structures with propeller loops are notable for high planarity, unlike quadruplexes with the diagonal loops.

CONCLUSIONS

In this work, we have investigated all known G-quadruplex structures. These structures can be divided into groups that are not only limited to the basis of

their topology; indeed, many groups are represented by sequences of common or related origin. It should be noted that the final discussion did not encompass all of the structures, since some of them are characterized by a topology too specific for the creation of representative samples. We have proposed two parameters to be used for the description of the geometry of quadruplexes: the twist angle of the tetraplex and the G-tetrad planarity. We demonstrated that the loops connecting the G-strand blocks are the major source of tension in the quadruplex structure. Lateral loops result in the strongest alteration of the geometry of G-quadruplexes; however, their influence is balanced by the introduction of diagonal and propeller loops into the structure. The diagonal loops provide the strict determination of the quadruplex structure, as well: however, the tensions induced by them are not as high as in the case with lateral loops. Propeller loops are characterized by an optimum quadruplex geometry. ●

REFERENCES:

1. Gellert M., Lipsett M., Davies D. // Proc. Natl. Acad. Sci. USA. 1962. V. 48. P. 2013–2018.
2. Arnott S., Chandrasekaran R., Marttila C.M. // Biochem. J. 1974. V. 141. P. 537–543.
3. Zimmerman S. // J. Mol. Biol. 1976. V. 106. P. 663–672.
4. Simonsson T. // Biol. Chem. 2001. V. 382. P. 621–628.
5. De Lange T. // Cold Spring Harb. Symp. Quant. Biol. 2005. V. 70. P. 197–204.
6. Wright W.E., Tesmer V.M., Huffman K.E., Levene S.D., Shay J.W. // Genes Dev. 1997. V. 11. P. 2801–2809.
7. Lei M., Podell E.R., Cech T.R. // Nat. Struct. Mol. Biol. 2004. V. 11. P. 1223–1229.
8. Sen D., Gilbert W. // Nature. 1988. V. 334. P. 364–366.
9. Sundquist W.I., Klug A. // Nature. 1989. V. 342. P. 825–829.
10. Gilbert D.E., Feigon J. // Curr. Opin. Struct. Biol. 1999. V. 9. P. 305–314.
11. Cogoi S., Xodo L.E. // Nucleic Acids Res. 2006. V. 34. P. 2536–2549.
12. Rankin S., Reszka A.P., Huppert J., Zloh M., Parkinson G.N., Todd A.K., Ladame S., Balasubramanian S., Neidle S. // J. Am. Chem. Soc. 2005. V. 127. P. 10584–10589.
13. Dexheimer T.S., Sun D., Hurley L.H. // J. Am. Chem. Soc. 2006. V. 128. P. 5404–5415.
14. Siddiqui-Jain A., Grand C.L., Bearss D.J., Hurley L.H. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. P. 11593–11598.
15. Hurley L.H., von Hoff D.D., Siddiqui-Jain A., Yang D. // Semin. Oncol. 2006. V. 33. P. 498–512.
16. Nimjee S.M., Rusconi C.P., Sullenger B.A. // Annu. Rev. Med. 2005. V. 56. P. 555–583.
17. Shamah S.M., Healy J.M., Cload S.T. // Acc. Chem. Res. 2008. V. 41. P. 130–138.
18. Bock L.C., Griffin L.C., Latham J.A., Vermaas E.H., Toole J.J. // Nature. 1992. V. 355. P. 564–566.
19. Jing N., Zhu Q., Yuan P., Li Y., Mao L., Tweardy D.J. // Mol. Cancer Ther. 2006. V. 5. P. 279–286.
20. McMicken H.W., Bates P.J., Chen Y. // Cancer Gene Ther. 2003. V. 10. P. 867–869.
21. Jayapal P., Mayer G., Heckel A., Wennmohs F. // J. Struct. Biol. 2009. V. 166. P. 241–250.
22. Lim K.W., Alberti P., Guédin A., Lacroix L., Riou J., Royle N.J., Mergny J., Phan A.T. // Nucl. Acids Res. 2009. V. 37. P. 6239–6248.
23. Wang Y., Patel D. // Structure. 1993. V. 1. P. 263–282.
24. Parkinson G.N., Lee M.P.H., Neidle S. // Nature. 2002. V. 417. P. 876–880.
25. Haider S., Parkinson G.N., Neidle S. // J. Mol. Biol. 2002. V. 320. P. 189–200.
26. Horvath M.P., Schultz S.C. // J. Mol. Biol. 2001. V. 310. P. 367–377.
27. Phillips K., Dauter Z., Murchie A., Lilley D., Luisi B. // J. Mol. Biol. 1997. V. 273. P. 171–182.
28. Hud N.V., Smith F.W., Anet F.A., Feigon J. // Biochemistry. 1996. V. 35. P. 15383–15390.
29. Kankia B.I., Marky L.A. // J. Am. Chem. Soc. 2001. V. 123. P. 10799–10804.
30. Deng J., Xiong Y., Sundaralingam M. // Proc. Natl. Acad. Sci. USA. 2001. V. 8. P. 13665–13670.
31. el Hassan M.A., Calladine C.R. // J. Mol. Biol. 1995. V. 251. P. 648–664.
32. Reshetnikov R., Golovin A., Spiridonova V., Kopylov A., Šponer J. // J. Chem. Theory Comput. 2010. V. 6. P. 3003–3014.
33. Patel P.K., Koti A.S., Hosur R.V. // Nucl. Acids Res. 1999. V. 27. P. 3836–3843.
34. Gavathiotis E., Searle M.S. // Org. Biomol. Chem. 2003. V. 1. P. 1650–1656.
35. Gavathiotis E., Heald R.A., Stevens M.F.G., Searle M.S. // J. Mol. Biol. 2003. V. 334. P. 25–36.
36. Clark G.R., Pytel P.D., Squire C.J., Neidle S. // J. Am. Chem. Soc. 2003. V. 125. P. 4066–4067.
37. Wang Y., Patel D. // Biochemistry. 1992. V. 31. P. 8112–8119.
38. Patel P.K., Hosur R.V. // Nucleic Acids Res. 1999. V. 27. P. 2457–2464.
39. Patel P.K., Bhavesh N.S., Hosur R.V. // Biochem. Biophys. Res. Commun. 2000. V. 270. P. 967–971.
40. Schultze P., Macaya R.F., Feigon J. // J. Mol. Biol. 1994. V. 235. P. 1532–1547.
41. Marathias V.M., Bolton P.H. // Nucleic Acids Res. 2000. V. 28. P. 1969–1977.
42. Marathias V.M., Wang K.Y., Kumar S., Pham T.Q., Swaminathan S., Bolton P.H. // J. Mol. Biol. 1996. V. 260. P. 378–394.
43. Mao X., Marky L.A., Gmeiner W.H. // J. Biomol. Struct. Dyn. 2004. V. 22. P. 25–33.
44. Phan A.T., Kuryavyi V., Luu K.N., Patel D.J. // Nucleic Acids Res. 2007. V. 35. P. 6517–6525.
45. Wang Y., Patel D.J. // Structure. 1994. V. 2. P. 1141–1156.
46. Luu K.N., Phan A.T., Kuryavyi V., Lacroix L., Patel D.J. // J. Am. Chem. Soc. 2006. V. 128. P. 9963–9970.
47. Dai J., Punchihewa C., Ambrus A., Chen D., Jones R.A., Yang D. // Nucl. Acids Res. 2007. V. 35. P. 2440–2450.
48. Dai J., Carver M., Punchihewa C., Jones R.A., Yang D. // Nucl. Acids Res. 2007. V. 35. P. 4927–4940.
49. Dai J., Chen D., Jones R.A., Hurley L.H., Yang D. // Nucl. Acids Res. 2006. V. 34. P. 5133–5144.
50. Lim K.W., Amrane S., Bouaziz S., Xu W., Mu Y., Patel D.J., Luu K.N., Phan A.T. // J. Am. Chem. Soc. 2009. V. 131. P. 4301–4309.
51. Zhang Z., Dai J., Veliath E., Jones R.A., Yang D. // Nucl. Acids Res. 2010. V. 38. P. 1009–1021.

52. Hu L., Lim K.W., Bouaziz S., Phan A.T. // J. Am. Chem. Soc. 2009. V. 131. P. 16824–16831.
53. Smith F., Feigon J. // Nature. 1992. V. 356. P. 164–168.
54. Wang Y., Patel D.J. // J. Mol. Biol. 1995. V. 251. P. 76–94.
55. Parkinson G.N., Cuenca F., Neidle S. // J. Mol. Biol. 2008. V. 381. P. 1145–1156.
56. Ambrus A., Chen D., Dai J., Jones R.A., Yang D. // Biochemistry. 2005. V. 44. P. 2048–2058.
57. Phan A.T., Kuryavyi V., Gaw H.Y., Patel D.J. // Nat. Chem. Biol. 2005. V. 1. P. 167–173.
58. Hsu S.D., Varnai P., Bugaut A., Reszka A.P., Neidle S., Balasubramanian S. // J. Am. Chem. Soc. 2009. V. 131. P. 13399–13409.
59. Kuryavyi V., Phan A.T., Patel D.J. // Nucl. Acids Res. 2010. doi: 10.1093/nar/gkq558.
60. Matsugami A., Ouhashi K., Kanagawa M., Liu H., Kanagawa S., Uesugi S., Katahira M. // J. Mol. Biol. 2001. V. 313. P. 255–269.
61. Phan A.T., Kuryavyi V., Ma J., Faure A., Andréola M., Patel D.J. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. P. 634–639.
62. Kettani A., Bouaziz S., Gorin A., Zhao H., Jones R.A., Patel D.J. // J. Mol. Biol. 1998. V. 282. P. 619–636.
63. Bouaziz S., Kettani A., Patel D.J. // J. Mol. Biol. 1998. V. 282. P. 637–652.
64. Kettani A., Basu G., Gorin A., Majumdar A., Skripkin E., Patel D.J. // J. Mol. Biol. 2000. V. 301. P. 129–146.
65. Smith F., Feigon J. // Biochemistry. 1993. V. 32. P. 8682–8692.
66. Haider S.M., Parkinson G.N., Neidle S. // J. Mol. Biol. 2003. V. 326. P. 117–125.
67. Campbell N.H., Patel M., Tofa A.B., Ghosh R., Parkinson G.N., Neidle S. // Biochemistry. 2009. V. 48. P. 1675–1680.
68. Gill M.L., Strobel S.A., Loria J.P. // J. Am. Chem. Soc. 2005. V. 127. P. 16723–16732.
69. Gill M.L., Strobel S.A., Loria J.P. // Nucleic Acids Res. 2006. V. 34. P. 4506–4514.
70. Balkwill G.D., Garner T.P., Williams H.E.L., Searle M.S. // J. Mol. Biol. 2009. V. 385. P. 1600–1615.
71. Sket P., Crnugelj M., Plavec J. // Bioorg. Med. Chem. 2004. V. 12. P. 5735–5744.
72. Crnugelj M., Hud N.V., Plavec J. // J. Mol. Biol. 2002. V. 320. P. 911–924.
73. Keniry M.A., Strahan G.D., Owen E.A., Shafer R.H. // Eur. J. Biochem. 1995. V. 233. P. 631–643.
74. Parkinson G.N., Ghosh R., Neidle S. // Biochemistry. 2007. V. 46. P. 2390–2397.
75. Campbell N.H., Parkinson G.N., Reszka A.P., Neidle S. // J. Am. Chem. Soc. 2008. V. 130. P. 6722–6724.
76. Webba da Silva M. // Biochemistry. 2003. V. 42. P. 14356–14365.
77. Kettani A., Gorin A., Majumdar A., Hermann T., Skripkin E., Zhao H., Jones R., Patel D.J. // J. Mol. Biol. 2000. V. 297. P. 627–644.
78. Webba da Silva M. // Biochemistry. 2005. V. 44. P. 3754–3764.

Biopharmacology of Enzyme Conjugates: Vasoprotective Activity of Supramolecular Superoxide Dismutase-Chondroitin Sulfate-Catalase Derivative

A.V. Maksimenko*, A.V. Vavaev, L.I. Bouryachkovskaya, V.P. Mokh, I.A. Uchitel, V.L. Lakomkin, V.I. Kapelko, E.G. Tischenko
Institute of Experimental Cardiology, Russian Cardiology Research-and-Production Complex
* E-mail: alexmak@cardio.ru
Received 07.10.2010

ABSTRACT **Bienzyme conjugate was obtained by the covalent connection of superoxide dismutase with catalase through endothelial glycocalyx glycosaminoglycan – chondroitin sulfate (SOD-CHS-CAT). This SOD-CHS-CAT conjugate has vasoprotective activity in respect to platelet interactions, tonus of the ring arterial fragment of a rat blood vessel, as well as normalization of hemodynamic parameters in rats and rabbits in conditions of oxidative stress caused by the administration of hydrogen peroxide. The SOD-CHS-CAT conjugate had antiplatelet potential due to its antiaggregation action manifested through the combination of enzyme activities and an acquired supramolecular structure. The influence on arterial fragment tonus was equivalent for SOD and CAT in native and conjugated form. Blood pressure and heart rate were significant and effectively normalized with SOD-CHS-CAT conjugate in rats and rabbits (after hydrogen peroxide administration as a perturbance stimulus). We have discovered the possibility of using the antioxidant bienzyme conjugate in chronic prophylaxis. It is important for a real development of the oral form of the SOD-CHS-CAT conjugate. These results indicate that the development of enzyme conjugates can be medically significant, as a promising approach for the creation of new drugs.**

KEYWORDS **antioxidant therapy, superoxide dismutase, catalase, chondroitin sulfate, vascular wall, oxidative stress, hydrogen peroxide, bienzyme conjugate, platelets, ring arterial fragment, hemodynamics, vasoprotective activity.**

ABBREVIATIONS **SOD – superoxide dismutase, CAT – catalase, CHS – chondroitin sulfate, SOD-CHS-CAT – bienzyme conjugate superoxide dismutase-chondroitin sulfate-catalase, L-NNA – N_{ω}-nitro-*L*-arginine, SNP – sodium nitroprusside, ADP – adenosine diphosphate, TRAP – thrombin receptor-activation peptide, NA – noradrenaline, BP – mean blood pressure, HR – heart rate, ECG – electrocardiogram, PRP – platelet-rich plasma, ROS – reactive oxygen species.**

INTRODUCTION

Enzymes are widely used as medication in thrombolytic therapy [1, 2]. The "golden molecule" of fibrinolysis and "golden time" of thrombolysis are known to increase the arsenal of biocatalysts used in treatment [3, 4]. However, the therapeutic applications of enzymes are far from having been exhausted, and there is a search for new forms that can be used in the development of original treatments for various pathologies [5, 6].

It is widely known today that the development of many pathologies is accompanied by oxidative stress [7, 8]. Normally, reactive oxygen species (ROS) participate in cellular signaling. However, the distortion of the balance between oxidative and antioxidant activity, as well as the mass production of ROS, results in oxidative stress. Excess ROS in the organism leads to the modification of macromolecules, disbalance of metabolic pathways and progression of the pathologic processes [9] that can be prevented or delayed by means of antioxidant administration [10]. Enzymes are considered as very efficient antioxidants, since they are specific and their mode of action is known in most cases [11]. Oxidative stress is known to play an important role in the pathogenesis of most cardiovascular disorders, which is why many cardiology studies are focused on antioxidants.

Current views hold that superoxide dismutase, catalase, and glutathione peroxidase are the main antioxidant enzymes of the body. The autonomous functioning of the first two makes them attractive in terms of the development of antioxidant medication to protect the cardiovascular system against oxidative stress. Based

on the described advantages, we chose Cu,Zn-superoxide dismutase (SOD), and catalase for the development of an enzymatic antioxidant derivate.

Our approach was based on biochemical coupling of SOD and CAT activity, when the product of SOD-reaction (H_2O_2) is used as a substrate by CAT, which finally yields water and oxygen as products [9, 11]. Accumulation of a glycosaminoglycan of endothelial glycocalyx – chondroitin sulfate (CHS) – in zones of initial atherosclerotic changes of vessels (i. e. in sites of potential damage of the vascular wall) [12] allowed the use of CHS as a cross-linking modifier of the enzyme subunits [13]. The water-soluble form of the obtained exogenic enzymatic conjugate SOD-CHS-CAT can be administered intravenously and per orally. It should be noted that the dimensions of a CAT molecule are 10.5 × 10.5 × 5.0 nm [14], and they are 6.7 × 3.6 × 3.3 nm for a SOD molecule [15]. The polymer chain of CHS winds around the enzymatic subunits and connects them, thus forming a covalent conjugate, which was confirmed by means of electrophoresis under denaturing conditions [13, 16]. Moreover, the conjugate is more active *in vivo* than combinations of individual components, which are accounted for by its optimal intravascular distribution and higher action efficiency [17]. Based on its dimensions – (17–20) × (14–18) × (8–12) – the conjugate occupies the lower zone of the nanoscale and, therefore, can be considered a nanoparticle. It is thought that the physical, chemical, and biological properties of molecular objects with nano-dimensions would surprisingly differ from the properties of their individual components, in part due to their quantum-mechanical effects.

Considering the supramolecular SOD-CHS-CAT conjugate, a bienzymatic nano-device, in the current work we studied its interactions with platelets (occurring in the blood stream) and with a ring arterial fragment (occurring on the surface of the vascular wall). We also studied the properties of the supramolecular conjugate at the level of the organism in laboratory animals, both under conditions of oxidative stress as modeled by infusion of hydrogen peroxide, and under stress-free conditions.

EXPERIMENTAL PROCEDURES

Materials

In the present study, we used the following reagents: Cu,Zn-superoxide dismutase (SOD), isolated from bovine red blood cells (specific activity 3,000 U/mg protein); catalase (CAT) from bovine liver (specific activity 11,000 U/mg protein); chondroitin-4-sulfate A (molecular weight 25-50 kDa) from bovine trachea; benzoquinone, dimethylformamide, β-galactosidase (from *Escherichia coli*), xantine, hydrogen peroxide, noradrenaline (NA), N_ω-nitro-*L*-arginine (L-NNA), acetylcholine, sodium nitroprusside (SNP) were purchased from Sigma, USA. Xantinoxidase was purchased from Calbiochem (USA), nitrotetrazolium blue - from Reanal (Hungary), sephadex G-25 and sephacryle S-300 - from Pharmacia (Sweden). The other reagents were of analytical grade purity and manufactured in Russia.

Bienzyme SOD-CHS-CAT conjugate was obtained as previously described [16]. The protein content in a SOD-CHS-CAT preparation was 4-6% weight, specific SOD activity was 60 U/mg solid, and CAT - 140 U/mg solid. For preparation of the SOD-CHS-CAT conjugate with irreversibly inactivated enzymes, SOD and CAT were preliminarily incubated in the presence of 0.3 M hydrogen peroxide (pH 7.0, 0.02 M phosphate buffer, 3 hours at room temperature) and at pH 11.8-12.0 (0.05 M NaOH, 2 hours at room temperature), respectively [16].

Methods

Biochemical measurements. The protein content in preparations was determined by the Bradford method. The enzymatic activity of SOD was measured as inhibition of reduction of nitrotetrasolium blue in the system xantine-xantinoxidase, pH 7.8 [13]. CAT activity was determined spectrophotometrically as a decrease in absorption at λ=240 nm (disappearance of hydrogen peroxide; pH 7.0, room temperature) [16].

Studies on platelet aggregation. In order to investigate the effects of hydrogen peroxide and SOD-CHS-CAT on platelet aggregation, we used the blood of volunteers, which was taken from the ulnar vein and stored in plastic tubes containing 0.13 M sodium citrate (pH 7.3). Platelet-rich plasma (PRP) was obtained by centrifugation of blood samples at 180g for 15 min. Platelet aggregation was estimated by means of the laser two-channel aggregation analyzer BIOLA (LA 230-2 model, NPF Biola, Russia). Apart from the customary approach based on registration of light transmission (Born's method), platelet aggregation was measured based on the analysis of fluctuations of the light transmission. The relative value of these fluctuations is proportional to the average radius of aggregates and allows to study the formation of micro-aggregates; i. e., those containing less than 100 platelets. It also allows the omission of the possible uncertainties that arise from light absorption by plasma and changes in platelet shape, which is of primary importance in studies of spontaneous aggregation.

The ability to form small-sized aggregates (3-100 platelets) was estimated by measuring both spontaneous aggregation and aggregation induced by the addi-

tion of 0.5 μM ADP, 0.5 μM serotonine and 1.0 μM TRAP (thrombin receptor-activation peptide) by the method of registration of the avarage size of the aggregates (in relative units). The formation of large aggregates (more than 100 platelets) in response to the addition of 5.0 μM ADP and 6.0 μM TRAP was estimated by the Born's method in % light transmission. Measurements were carried out during 2 hours after blood samples were obtained in 0.3 ml cuvettes.

Platelet adhesion was estimated by means of electronic microscopy. A 15 μM sample was mixed with a saline solution, H_2O_2 and/or CAT preparations and loaded onto an adhesive surface (glass), incubated 15 minutes at room temperature in an enclosed space for preventing drying-out, rinsed with saline to wash the unattached platelets off and fixed by 2.5% glutaraldehyde for 1.5 hours. After fixation, the samples were dehydrated and prepared for electron microscopy.

Platelets of various shapes were quantified on 25 scanning fields at a magnification of 2500x (PHILLIPS PSEM 550x scanning electron microscope). Adhesion was expressed as a percentage of platelets stuck to the surface.

Investigation of changes in tone of rat ring arterial fragment. After the decapitation of Wistar rats (males, 350-400 g), the abdominal cavity was opened and abdominal aorta was isolated. The isolated aortal segment was cleared from connective tissue and sliced into ring fragments 3 mm in length. Aortal fragments were then put on stainless needles connected to the tensosensor (which measures the mN forces produced by the arterial fragment) and placed into a Krebs-Henseleit buffer insufflated with carbogenum at 37°C and pH 7.4 [18]. Oxidative stress was modeled by the addition of hydrogen peroxide to the preparation previously subjected to NA preconstriction (preliminary constriction of the arterial fragment induced by the addition of NA prior to another stimulus). Alteration of vascular tone was estimated relative to the level of contraction induced by 0.1 μM NA taken as 100%. Various concentrations of antioxidant enzyme derivatives were added 10 minutes before the addition of hydrogen peroxide against a background of NA-preconstriction. The production of endogenous NO and its influence on the tone of the ring arterial fragment was initiated by the addition of the exogenous inhibitor of NO-syntase L-NNA (0.1 μM).

Experiments in vivo. Tolerance and the protective action of the SOD-CHS-CAT conjugate under conditions of H_2O_2-induced oxidative stress were studied in male rabbits (n=29) 3.65±0.10 kg in weight and male Wistar rats (n=13) 427±7g in weight. All animals were anaesthetized prior to the experiments by infusion of ketamine.

In rabbits undergoing ketamine narcosis (50-60 mg/kg weight), catheters (PE-50 diameter) were placed in the central artery of one ear and the marginal veins of both ears. After the initial narcosis, (55 mg/kg) 5% ketamine was infused with a syringe pump (SAGE Instruments, USA) at a rate of 36-54 μl/hour per 1 kg of weight. The vein of the other ear was bolus-infused with saline or SOD-CHS-CAT solution, and 0.8% hydrogen peroxide solution at a rate of 0.4 ml/min for 3 min, performed twice with a 20 min interval, which was necessary for full recovery of parameters after the first injection of hydrogen peroxide and distribution of the conjugate in the body. In the acute experiment, the infusion of reagents was done in the following order: hydrogen peroxide - saline solution (control group) or SOD-CHS-CAT (experiment group) - hydrogen peroxide. Registration of the mean blood pressure (BP), heart rate (HR) and electrocardiogram (ECG) in the second lead was carried out on BIOGRAF-4 (Saint-Petersburg State University of Aerospace Instrument Making, Russia) equipped with an ADC board NI 6210 (National Instruments, USA). Physiological signals were processed using corresponding software (Dr. E.V. Lukoshkova). After the registration of initial BP, HR and ECG (during 15 min) animals were infused with hydrogen peroxide (3 min) and the aforementioned physiological parameters were measured again during 10 minutes. The total dose of hydrogen peroxide injected throughout the experiment was 0.31 μM/kg. The number of animals in the control group was 12; in the experimental group - 15.

In Wistar rats under ketamine narcosis (100 mg/kg), catheters (PE-50) were placed in the carotid artery and jugular vein. The protocol of the experiment was essentially similar to the abovementioned, but with a minor difference: the time of hydrogen peroxide infusion was extended by 2 minutes in order to obtain similar effects on hemodynamic. The total dose of hydrogen peroxide injected in rats was 4.5 μM/kg. No animals died in the course of these experiments.

A separate series of experiments on rabbits (n=8) was focused on the investigation of tolerance of various doses of the SOD-CHS-CAT conjugate (therapeutic – 1.5 mg/kg and also 7.5 and 15 mg/kg). HR, BP, and ECG parameters were measured in these experiments. After a 15-minute control registration of BP, HR (taken as 100%), and ECG, the first dose of SOD-CHS-CAT (1.5 mg/kg) was injected and HR, BP, and ECG were registered during 15 min. Afterward, the next dose of SOD-CHS-CAT (7.5 mg/kg) was injected, followed by the registration of hemodynamic parameters. 20 minutes later, another (15 mg/kg) dose, ten times greater

than the therapeutic dose, was administered and hemodynamic parameters were measured. The animals that received a total dose of SOD-CHS-CAT 16 times greater in 72 hours were used to estimate the preventive effects of the bienzyme conjugate during oxidative stress (as described above). This contrasts with the effects obtained in both acute and control experiments on animals that had not been administered anything beforehand.

Statistical treatment of obtained results. The results obtained are presented as a mean value±standard error, where n is the number of animals. The two groups were compared using the two-sample Student's t-test (the statistical significance of the differences was estimated at $p < 0.01$). In the case of more than two comparison groups, the statistical calculations were carried out using the ANOVA ($p < 0.01$) method.

RESULTS AND DISCUSSION

The effects of SOD-CHS-CAT on platelets

Under normal conditions, hydrogen peroxide acts as an intra- and inter-cellular signaling molecule. In the concentration range of 20-50 μM, hydrogen peroxide has limited cytotoxicity for many cell types. Under physiological conditions, 50 μM concentration is considered high [19]. The addition of 50-2000 μM hydrogen peroxide *in vitro* led to an aggregation of platelets (Fig. 1). Scanning electronic microscopy showed that at the moment of maximum aggregation the aggregates consisted of tightly packed platelets in the core. However, the platelets were bound rather loosely on the periphery (Fig. 1B). By the 5th minute in the process, the average size of the aggregates had decreased and their structure had become so dense that it was impossible to distinguish discrete platelets (Fig. 1C). The decrease in the size of the aggregates could be explained by a

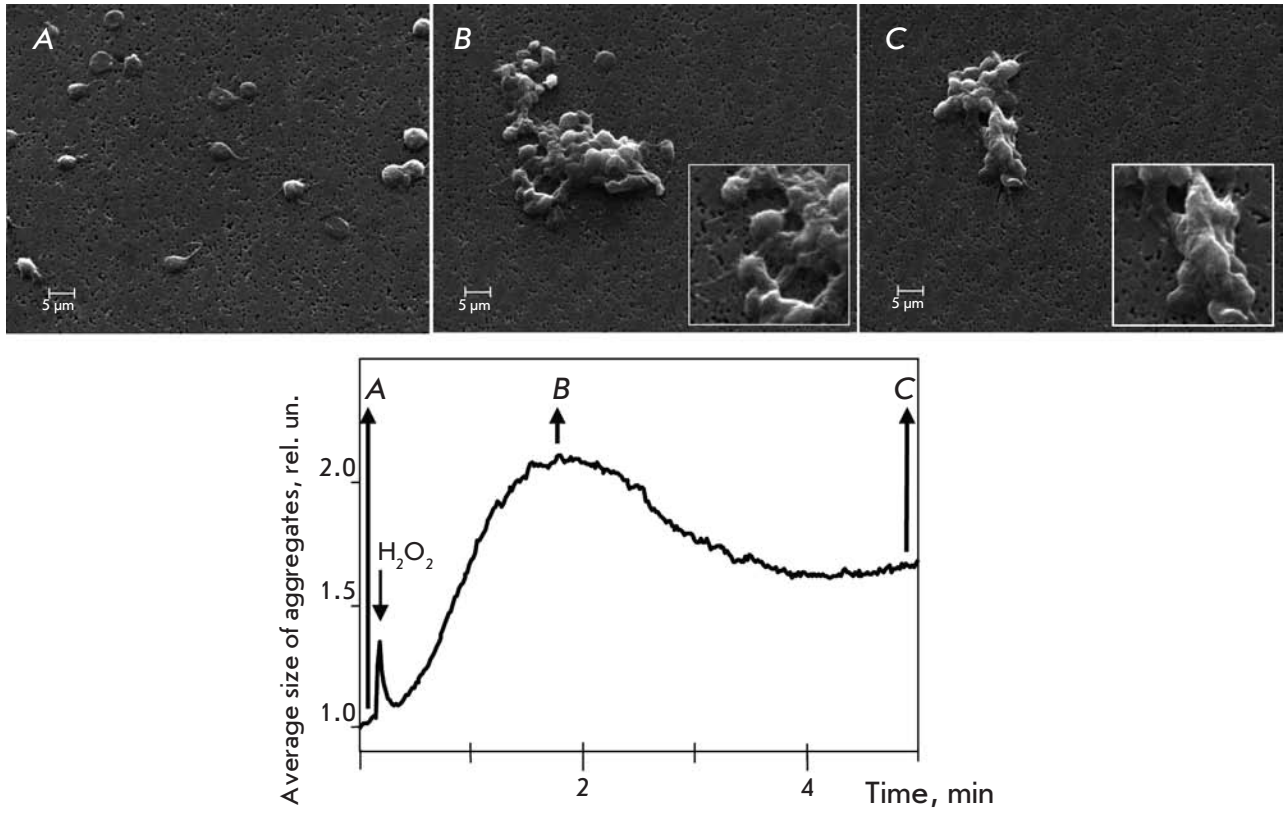


Fig. 1. Typical platelet aggregation curve induced by 50 μM – 2.0 mM hydrogen peroxide and composition of platelet aggregates in different stages of their formation. **A**: segregate platelets before aggregation, mainly, in discoid forms; **B**: sample of aggregation peak – aggregates consist of tight core in their center and weakly attached platelets on their periphery (friable structure of peripheral area has been shown on insert in magnified view); **C**: sample after five minutes of platelet aggregation – the size of aggregates is decreased, weakly attached platelets are absent on periphery, strong fusion of platelets in aggregates. Magnification x2500, on insert x5000.

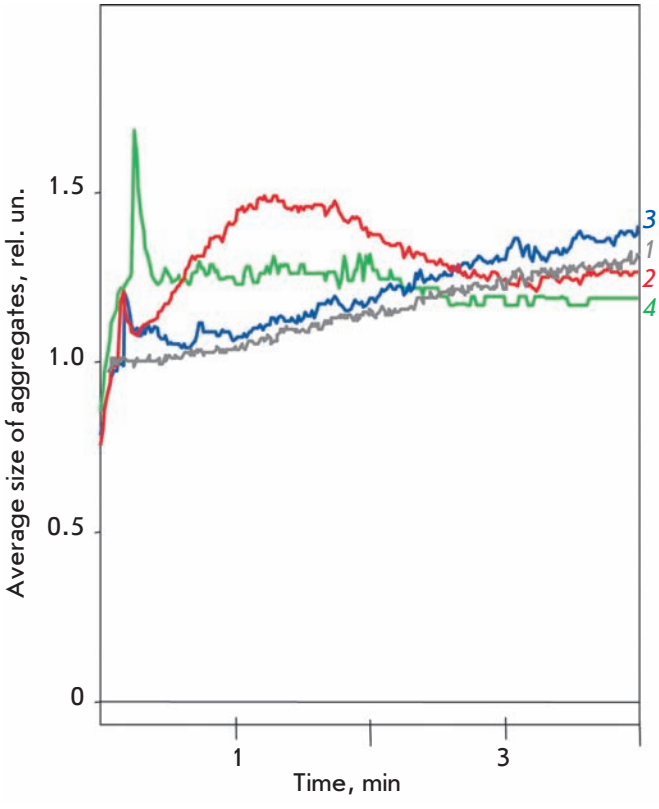


Fig. 2. Influence of hydrogen peroxide, CAT and SOD-CHS-CAT on platelet aggregation. Curve of spontaneous platelet aggregation (1), aggregation in presence of 300 μM hydrogen peroxide (2), addition of 3,000 U native CAT first, followed by 300 μM hydrogen peroxide (3), addition of 400 U CAT activity of SOD-CHS-CAT first, followed by 300 μM hydrogen peroxide (4).

dissociation of weakly bound platelets from aggregate clusters, as well as the consolidation of aggregate cores. The addition of 3,000 U native catalase to the cuvette of the aggregometer substantially lowered the aggregation induced by 300 μM hydrogen peroxide (Fig. 2). The SOD-CHS-CAT conjugate was shown to suppress aggregation in a dose as low as 400 U CAT activity, showing an increased dose-dependent antioxidant activity compared to CAT alone. Neither CAT nor CHS affects platelet aggregation.

The presence of 300 μM H_2O_2 increased platelet activation, in terms of their aggregation, induced by various inducers (with varying mechanisms of action, i.e., ADP, serotonin and TRAP (Fig. 3)). The addition of 3,000 U CAT or 400 U CAT-activity of the SOD-CHS-CAT conjugate to PRP suppressed the activating effects of H_2O_2 (curves 4 and 5 in Figs. 3A, 3B and 3C, respectively). Application of the conjugate proves its increased antioxidant activity, exceeding that of CAT. The effect demonstrated dose-dependence and reached a maximum (which did not substantially change when increased doses were added to the cuvette) at 3,000 and 400 U CAT activity for CAT and SOD-CHS-CAT, respectively.

Due to the catalase activity, the CAT and SOD-CHS-CAT derivates suppressed the effects of H_2O_2 on platelets (Figs. 2 and 3). However, it was shown that the platelets themselves can produce ROS [20]. This is in line with the inhibiting effect of the SOD-CHS-CAT on ADP-induced platelet aggregation (Fig. 4). At the same time, the effect of the CAT-CHS derivate was quite moderate, and CAT did not possess it at all (data not shown). The anti-aggregate activity of the bienzyme conjugate was due to the presence of enzymatic activities (curves 1 and 2 in Fig. 4) and also to its unique supramolecular structure, mediated by CHS (curves 1 and 3 in Fig. 4) [11, 17]. Indeed, β-galactosidase (similar in molecular size to SOD-CHS-CAT) used in the same experimental scheme in equimolecular protein concentrations did not inhibit the ADP-induced platelet aggregation.

The action of SOD-CHS-CAT was expressed in inhibiting ADP-, serotonin-, and TRAP-induced platelet aggregation (all the inducers exploit different mechanisms of action and were used in different concentrations) (Fig. 5). This is a new quality for the SOD-CHS-CAT conjugate, since the individual components lack it.

Platelet spreading on the adhesive surface is one of the critical stages of hemostasis, which induces a sequence of reactions, leading to thrombus formation. Platelet adhesion and spreading takes place during transferral to glass (Fig. 6A). In the presence of H_2O_2, the amount of spread platelets increases (Fig. 6B), but when SOD-CHS-CAT conjugate is transferred beforehand to the glass (Fig. 6C) or the PRP (Fig. 6D), no spread platelets are found in the sample. A similar situation can be observed when PRP and SOD-CHS-CAT are transferred to glass which had H_2O_2 added to it beforehand (Fig. 6E). When PRP and CAT are added, the amount of spread platelets decreases significantly (Fig. 6F). We note that free CHS did not demonstrate anti-aggregate inhibition with regard to hydrogen peroxide. It is probable that this effect of the bienzyme conjugate has something to do with its adhesive and antioxidant qualities, which allow it, on the one hand, to protect the surface from platelet adhesion, and, on the other, to neutralize H_2O_2, which strengthens adhesion and platelet spreading on glass.

The results we obtained indicate that the bienzyme SOD-CHS-CAT conjugate expresses an antioxidant dose-dependent effect during induced platelet aggregation in the presence of hydrogen peroxide. The anti-aggregate activity of the SOD-CHS-CAT conjugate is

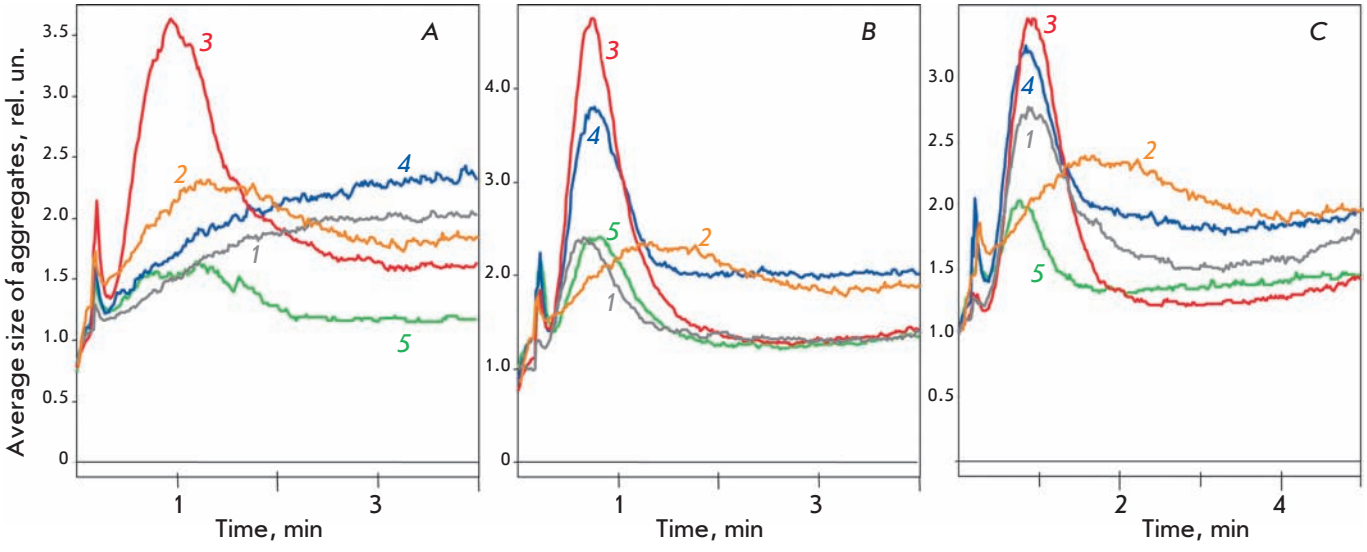


Fig. 3. Influence of CAT and SOD-CHS-CAT on platelet aggregation induced by ADP (**A**), serotonin (**B**), and TRAP (**C**) in the presence of hydrogen peroxide. **A**: Platelet aggregation curves in the presence of 0.5 μM ADP (1), 300 μM hydrogen peroxide (2), 0.5 μM ADP, and 300 μM hydrogen peroxide (3); preventive administration of 3,000 U CAT and then 0.5 μM ADP with 300 μM hydrogen peroxide (4); and preventive administration of 400 U of CAT activity of SOD-CHS-CAT and then 0.5 μM ADP with 300 μM hydrogen peroxide (5). **B**: Platelet aggregation curves in the presence of 0.5 μM serotonin (1), 300 μM hydrogen peroxide (2), 0.5 μM serotonin, and 300 μM hydrogen peroxide (3); preventive administration of 3,000 U CAT and then 0.5 μM serotonin with 300 μM hydrogen peroxide (4); and preventive administration of 400 U of CAT activity of SOD-CHS-CAT and then 0.5 μM serotonin with 300 μM hydrogen peroxide (5). **C**: Platelet aggregation curves in the presence of 1 μM TRAP (1), 300 μM hydrogen peroxide (2), 1 μM TRAP, and 300 μM hydrogen peroxide (3); preventive administration of 3,000 U CAT and then 1 μM TRAP with 300 μM hydrogen peroxide (4); and preventive administration of 400 U of CAT activity of SOD-CHS-CAT and then 0.5 μM TRAP with 300 μM hydrogen peroxide (5). Typical aggregation curves were representative of 4-5 experiments.

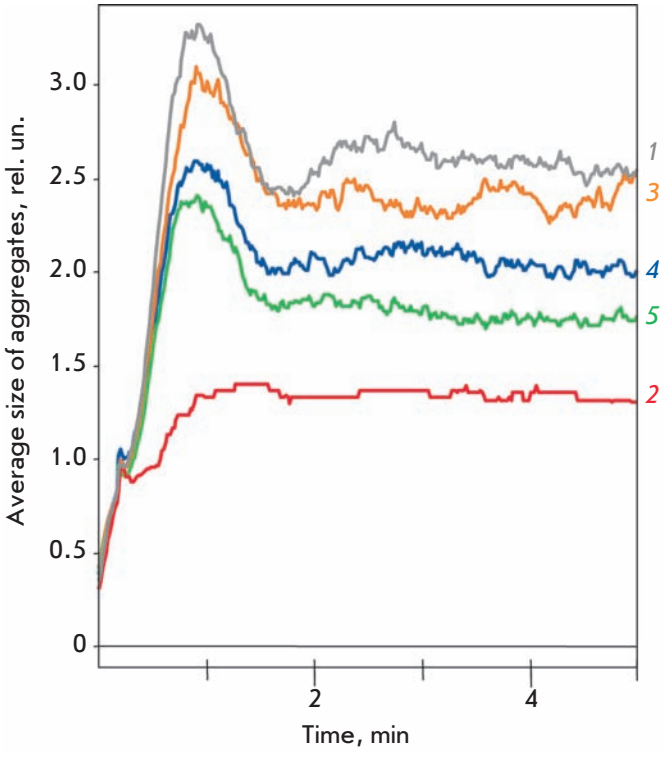


Fig. 4. Influence of SOD-CHS-CAT on platelet aggregation induced by 0.5 μM ADP. Aggregation curves with 0.5 μM ADP (1), and preventive administration (in equimolar concentration) of corresponding 400 U CAT activity of SOD-CHS-CAT forms: SOD-CHS-CAT (2), SOD_{inact}-CHS-CAT_{inact} (3), SOD-CHS-CAT_{inact} (4), SOD_{inact}-CHS-CAT (5). Typical aggregation curves were representative of 3-5 experiments.

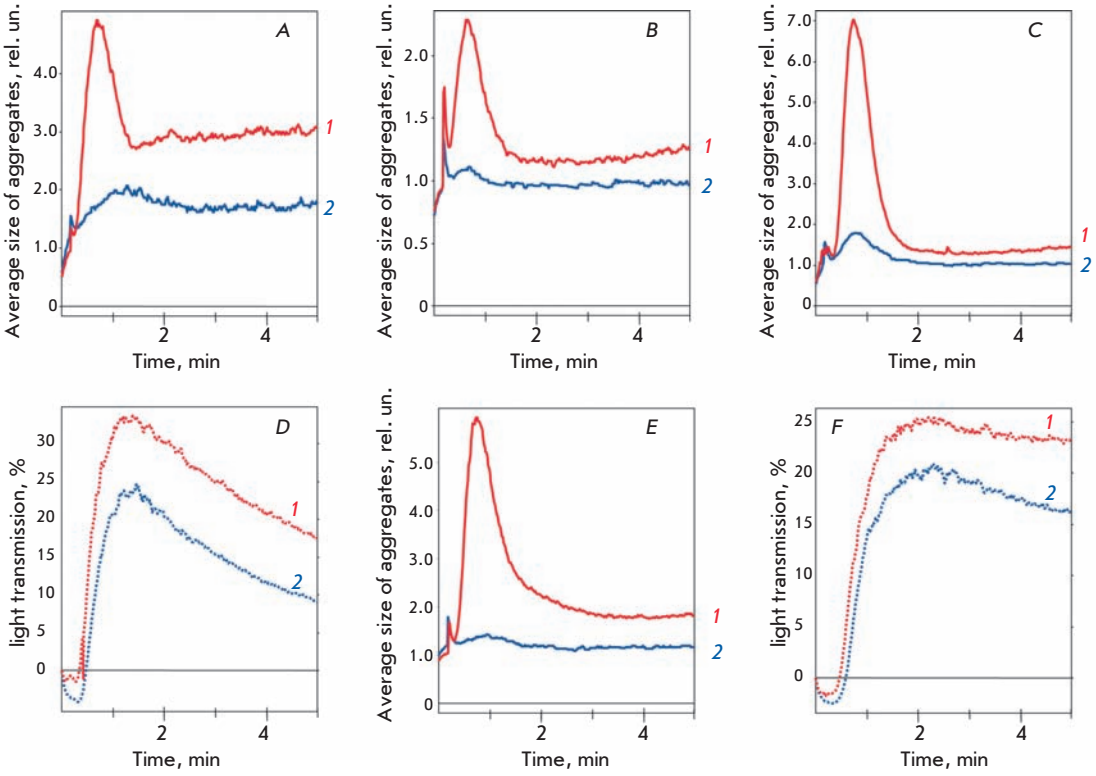


Fig. 5. Influence of SOD-CHS-CAT conjugate (400 U of CAT activity, curve 2) on platelet aggregation induced (curve 1) by: 0.5 μM (**A**) and 5.0 μM (**D**) ADP; 0.5 μM (**B**) and 5.0 μM (**E**) serotonin; 1 μM (**C**) and 6 μM (**F**) TRAP. Typical aggregation curves were representative of 3-6 experiments. Abscissa is time (minutes), ordinate is average size of platelet aggregates (in rel. units, **A-C**, **E**) or light transmission (%, **D** and **F**).

reliably higher than that of other CAT derivates and is demonstrated in a wide range of conditions when platelet aggregation is stimulated by various inducers (ADP, serotonin, TRAP – all of which differ in terms of mechanisms of action), both with and without hydrogen peroxide. The latter is evidence of a new aspect of the anti-aggregate potential of SOD-CHS-CAT, absent in the native enzymes and free CHS, and determined by its molecular composition and size. These qualities of the bienzyme SOD-CHS-CAT nanoconjugate show that it has a promising future in biopharmaceutical development for purposes of antioxidant therapy. They are also proof of the promises in using enzyme conjugates as medicinal agent.

The change in the tone of a ring fragment of rat aorta

The effects of oxidative stress on the tone of a ring fragment of rat abdominal aorta were modeled by the addition of H_2O_2 against a background of NA preconstriction. The latter made up 50-60% of the maximum possible contraction of the arterial fragment, which made it possible to measure both contraction and relaxation. The change in vascular tone was estimated relative to the contraction exhibited in response to 0.1 μM NA, taken at a 100% range of interval (for the gradation of the size of tone change), and at the baseline of the experiment as 0% (to indicate the direction of tone change, contraction or relaxation). Hydrogen peroxide caused a dose-dependent vascular contraction. At

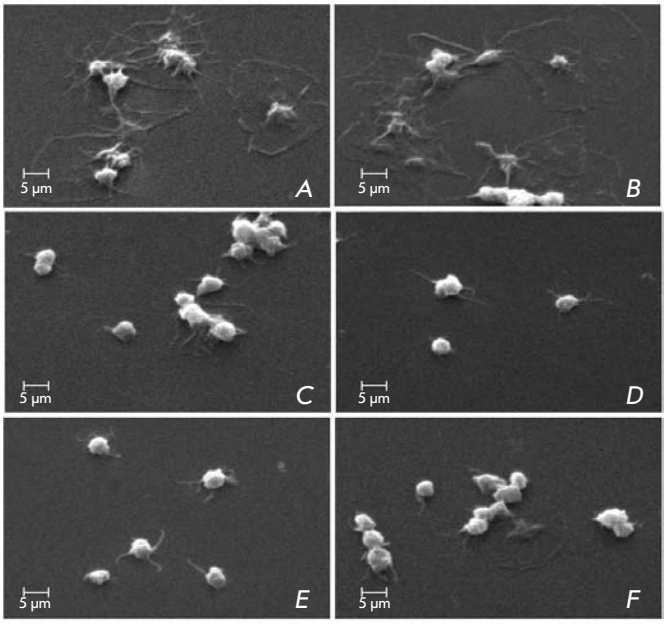


Fig. 6. Electron microscopy picture of platelet adhesion on glass surface (Scanning Electron Microscope Phillips PSEM 550x, magnification x2500). Platelet adhesion on glass with addition to PRP: saline (physiological solution, A), hydrogen peroxide (B), SOD-CHS-CAT (D). Adhesion of platelets on glass preventively treated by SOD-CHS-CAT and then the addition on the glass surface of the PRP with hydrogen peroxide (C); by hydrogen peroxide and then the addition on the glass surface of PRP with SOD-CHS-CAT (E) or with native CAT (F).

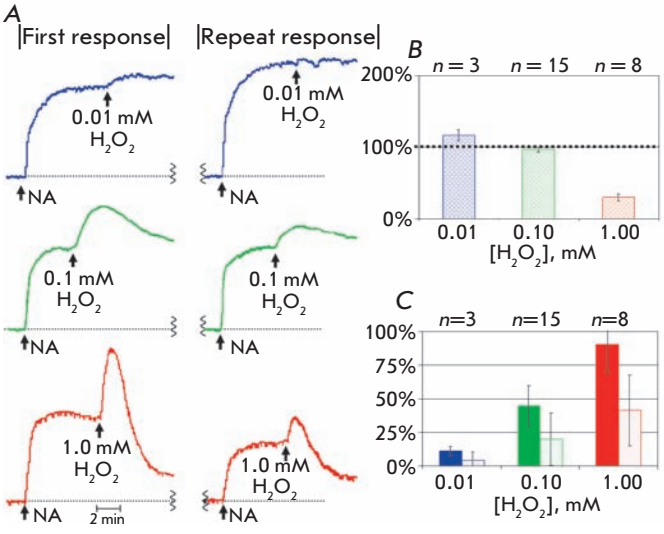


Fig. 7. First and repeat responses to NA and hydrogen peroxide. **A**: the original tracings of the first and repeat responses. **B**: the size of repeat response to NA, the size of first contraction induced by NA is equal to 100 %. The size of first (painted bar graphs) and repeat (empty bar graphs) contraction (**C**) of arterial fragment after administration of different concentrations of hydrogen peroxide.

a 0.01 mM concentration of H_2O_2, we observed a small (10-12%) tone increase. At a 0.1 mM H_2O_2, this increase was more (48-50%) with a subsequent decrease to the initial level of NA preconstriction (0-3% relaxation). At a 1.0 mM concentration of H_2O_2, there was a rapid contraction of the arterial fragment (88-90%), which was followed by a relaxation phase (68-70%). Thus, the model we presented revealed a dose-dependent effect of hydrogen peroxide and was suitable for an experimental study of the effect of ROS on the tone of the vascular fragment.

After triple washing and a 15-minute rest period, the vascular fragment was exposed to NA and hydrogen peroxide again at the concentrations used earlier (0.01-0.1 mM). The level of functional activity of the vascular fragment during the second response decreased with the increase in hydrogen peroxide concentration, both in relation to the second addition of NA, as well as to the amount of contraction during the second addition of H_2O_2 (Fig. 7). The reliable decrease in the second contraction of the arterial fragment after the second addition of large concentrations of hydrogen peroxide serves as a measure of the invariability of vascular reactions.

The addition of native CAT against the background of NA-induced preconstriction did not influence the vascular tone in any of the concentrations used. The reaction to the addition of 1.0 mM H_2O_2 in the presence of CAT decreased substantially in terms of contraction

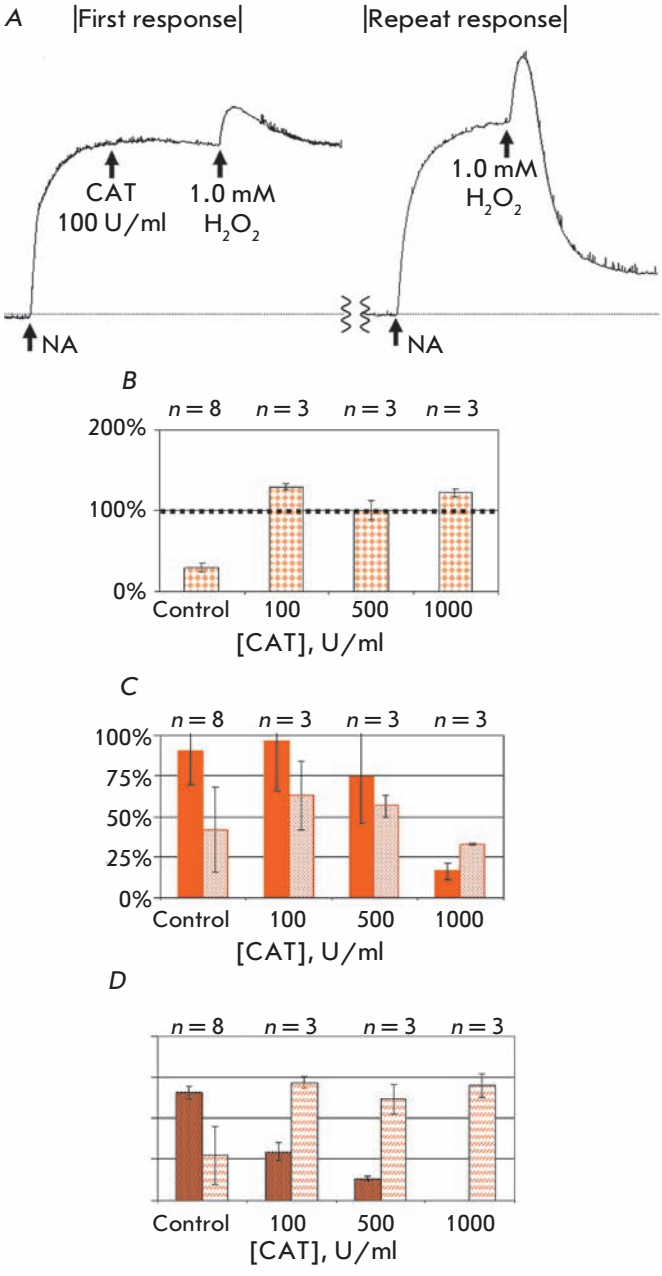


Fig. 8. Dose-dependent influence of native CAT on tone of arterial fragment after repeat addition of NA and 1.0 mM hydrogen peroxide. **A**: the original tracings. **B**: the size of repeat response to NA after addition of different concentrations of CAT (painted bar-graphs), CAT activity is equal in twain of comparison. The size of first (painted bar-graphs) and second (empty bar-grafs) contraction (**C**) and relaxation (**D**) of ring arterial fragment to 0.1 mM hydrogen peroxide according to CAT activity, (U/ml).

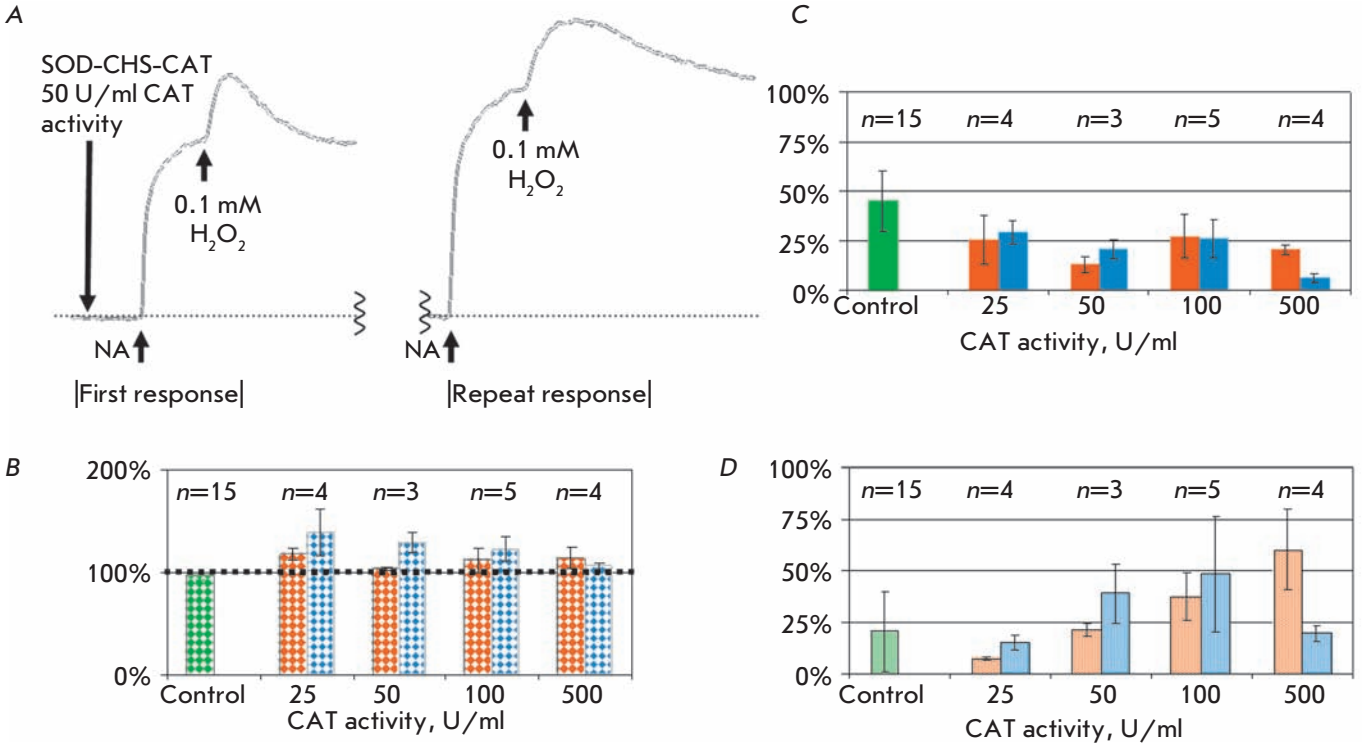


Fig. 9. Comparative effects of protective action of CAT and SOD-CHS-CAT against 0.1 mM hydrogen peroxide. **A**: the original tracings. **B**: the size of repeat response to NA after the addition of different concentrations of CAT (dark bar-graphs) and SOD-CHS-CAT (gray bar-graphs), CAT activity is equal in twain of comparison. **C** and **D**: the contraction of rat aorta fragment for first (**C**) and repeat (**D**) responses to 0.1 mM hydrogen peroxide in the presence (in first response) of noted concentration of CAT (left in twain) and SOD-CHS-CAT (right in twain) according to CAT activity (U/ml).

and relaxation. At a concentration of 1,000 U/ml the CAT effect of 1.0 mM H_2O_2 decreased to 16% contraction. CAT displayed a dose-dependent protective effect against the stimulation of the vascular fragment by hydrogen peroxide.

Prior incubation of the arterial fragment with CAT substantially influenced the invariability of subsequent responses to NA and hydrogen peroxide (Fig. 8). At a CAT concentration of 100 U/ml and higher the size of the second response to NA did not change, whereas in the absence of the enzyme it decreased three- or four-fold (Fig. 8B). The presence of CAT during the first addition of H_2O_2 increased the subsequent contractive and dilative response (in comparison with the control), and when 1,000 U/ml CAT was used the second response reliably exceeded the first (Figs. 8C and 8D). Insofar as vascular reactions to the repeated effects of NA and H_2O_2 in the absence of CAT decrease significantly, the protective function of CAT is obvious. The data received after using 0.1 mM H_2O_2 for 25-500 U/ml CAT confirm this conclusion.

The preventive character of antioxidants implies their defensive effect as soon as the physiological level of ROS is exceeded. Therefore, we used 0.1 mM H_2O_2 to estimate the comparative effectiveness of the protective effect of CAT and the SOD-CHS-CAT conjugate in our experimental model when testing the invariability of vascular functionality (Fig. 9). CAT derivates were used in concentrations that were identical in terms of CAT activity (U/ml). Native CAT and the SOD-CHS-CAT conjugate maintained their level of response to the secondary addition of NA (Fig. 9B); this addition is comparable to the control indicators. In a concentration range of 25-100 U/ml of catalase activity, the CAT and SOD-CHS-CAT derivates demonstrated similar (in terms of size) protective effects against 0.1 mM H_2O_2. At a concentration of 500 U/ml, the SOD-CHS-CAT conjugate lowered the amplitude response to H_2O_2 more effectively than CAT. The protective effect expressed by SOD-CHS-CAT may be connected to its affinity to the vascular wall (due to the conjugation of SOD with CAT via glycosaminoglycan of endothelial glycocalyx [11, 17]) and/or to the presence of SOD and endogenous NO in the model system. In the system described above, it is difficult to estimate the sorption of enzyme derivates because of the flow of vascular reactions not only from

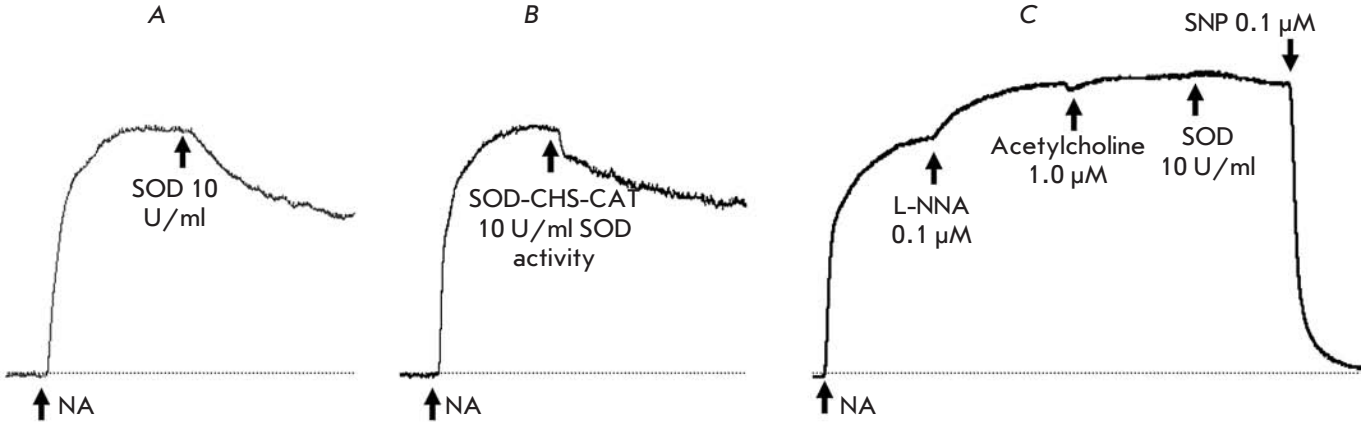


Fig. 10. Influence of SOD activity on the tone of arterial fragment. The original tracings after administration of 10 U/ml native SOD (**A**) or SOD-CHS-CAT (**B**) with NA preconstruction. **C**: the alteration of vascular tone after administration of NA, N-nitro-L-arginine (L-NNA), acetylcholine, native SOD and sodium nitroprusside (SNP), respectively. Alike curves were obtained after 3-4 experiments.

within outwards, but also from without inwards (i.e. on both the intima and the adventitia) [21]. The defensive effect of SOD with regard to NO, however, is possible to estimate.

This effect of native SOD and SOD-CHS-CAT was quite pronounced when studied on an NA-preconstricted arterial fragment (Fig. 10). Both substances caused reliable vascular relaxation. This suggests that the observed dilation was determined by the conservation of endogenous NO. This is because SOD neutralized the superoxide radical capable of turning NO into peroxynitrite, which does not have vasodilative properties. The experimental evidence of this is the inhibiting effect of NO-synthase with L-NNA, which caused an increase in vascular tone (Fig. 10C). This effect expresses the inhibition of NO-synthase activity and the development of a contractive response because of the lack of a dilative effect in endogenous NO. The introduction of acetylcholine (as an inducer of vascular relaxation by means of its effect on the endothelial receptors which start off NO-synthase) gave no response, which indicates the effective inhibition of NO-synthase. Against the background of L-NNA, the introduction of 10 U/ml SOD-activity gave no response. This confirms the role of endogenous NO in the vascular relaxation we observed earlier. The introduction of sodium nitroprusside (SNP), which acts as an NO donor, caused dilation (Fig. 10C); i.e., the vascular fragment did not lose its ability to relax under the influence of NO. The increase in the bioavailability of NO when the SOD-CHS-CAT conjugate was used, similar in size to the effect of native SOD (Fig. 10A), demonstrates the effectiveness of the vasoprotective action of the conjugate, due to the activity of its SOD component.

The bienzyme SOD-CHS-CAT overall reliably demonstrates its vasoprotective qualities that are connected to the activity of both of its enzyme components (SOD and CAT). These are no less effective than the native forms of biocatalysts. This is convincing evidence of the SOD-CHS-CAT conjugate's potential for medicinal development. However, its safety and effectiveness need to be assessed and proven *in vivo*.

Vasoactivity of the SOD-CHS-CAT conjugate *in vivo*

Intravenous bolus injection of varying doses of the SOD-CHS-CAT conjugate in rabbits showed that BP and HR were changed by no more than 4% relative to their average means in intact narcotized animals. ECG did not reveal any alterations of the ST-interval, rhythm or conductivity disturbances and other abnormalities, even when doses of the conjugate 10 times greater were administered. Singular intra-abdominal injections of SOD-CHS-CAT into BALB/c mice and CBAxC57B16 F1-hybrid mice showed a low acute toxicity of the conjugate and the absence of mutagenic properties, as confirmed through the Ames test. These data, together with the high tolerance of the SOD-CHS-CAT conjugate in animals as was mentioned earlier, as well as its pronounced antithrombotic activity [11, 17], justifies the significance of its further investigation.

The first injection of hydrogen peroxide into rabbits (Fig. 11A, curve 1) caused a sharp decrease in BP (up to 60% of its initial level), which was restored during 10 minutes in the control group to 90% of its initial level. Similar changes were observed in the experimental group (Fig. 11A, curve 2). It is noteworthy that the administration of the SOD-CHS-CAT derivate 72 hours

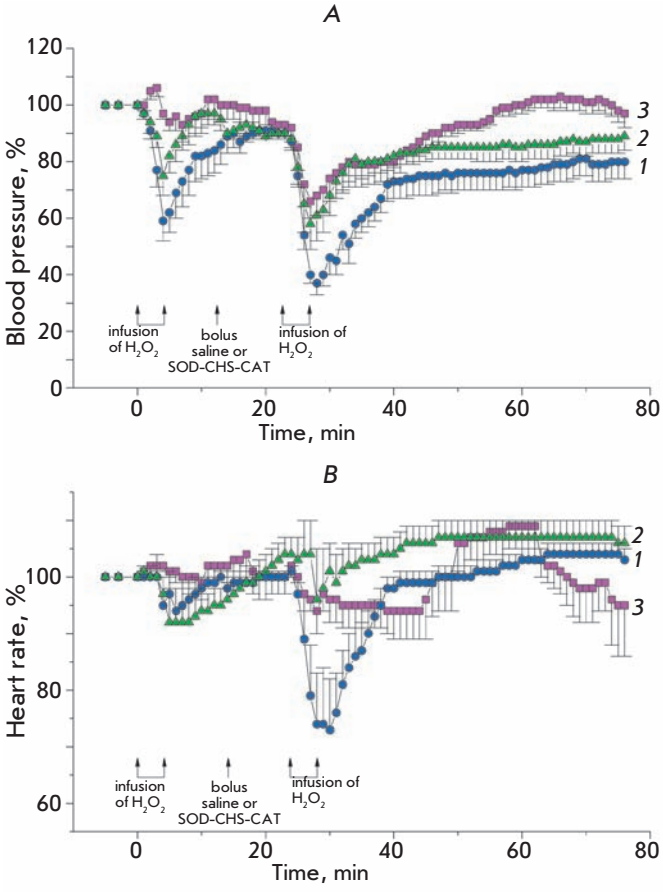


Fig. 11. Alteration of BP (**A**) and HR (**B**, both in % from initial) of anesthetized rabbits for oxidative stress (infusion of 0.8 % hydrogen peroxide) in control (n=12, bolus administration of saline, curve 1) and test (n=9, bolus administration of SOD-CHS-CAT conjugate, curve 2) groups. The curve 3 is appertained to prophylaxis test group (n=6) of rabbits obtaining the SOD-CHS-CAT conjugate earlier (three days earlier than this experiment) with hydrogen peroxide and after its first administration. The arrows show the moment beginning and ending the infusion of hydrogen peroxide and bolus administration of saline or SOD-CHS-CAT conjugate (1.5 mg/kg).

prior to the experiments reliably prevented a drop in BP. In that case, we observed only minor initial oscillations in the BP level followed by smoothening, and the stabilization of BP at its initial level (Fig. 11A, curve 3). Preventive administration of SOD-CHS-CAT accounted for its prophylactic effect in respect to HR in animals (Fig. 11B, curve 3) in comparison with data obtained in the control (Fig. 11B, curve 1) and the experimental (Fig. 11B, curve 2) groups.

Ten minutes after the injection of the conjugate (time required for its distribution in the rabbit's organism), which did not affect the characteristics of central hemodynamics, hydrogen peroxide was injected for the second time in a similar concentration and at the same rate. The second infusion caused more profound changes in BP and HR (Fig. 11). The course of BP recovery had two stages – a fast one that took 5 minutes and a slow one.

After the administration of SOD-CHS-CAT in the animals of the experimental group, the changes in BP and HR in response to the injection of the second dose of hydrogen peroxide were significantly less pronounced and recovery was faster (Fig. 11).

After the first injection of hydrogen peroxide, we did not detect any substantial changes in ECG in rabbits, although in the control group the ST-segment was lowered. The second injection of hydrogen peroxide caused the ST-interval to rise, especially in the cases when rabbits experienced breathing problems. At 5-8 minutes, we observed both isolated and group extrasistoles. The first injection of hydrogen peroxide was often accompanied by dyspnea, while after the second injection we observed hindered breathing and bronchospasms, which were more pronounced in the control group.

The first injection of hydrogen peroxide in rats caused an initial short-term (no more than 1 minute) 3-5% increase in BP with a subsequent decrease by 15% and full recovery during 10 minutes (Fig. 12A). Under these conditions, HR decreased by 3-5% and recovered to up to 96-98% of its initial level during 10 minutes (Fig. 12B, curves 1 and 2). Injection of SOD-CHS-CAT did not affect hemodynamics and ECG in rats. In ten minute's time after the injection of the conjugate (time required for its distribution in the rat's organism), which did not affect the characteristics of central hemodynamics, hydrogen peroxide was injected for the second time. The second infusion caused more profound changes in BP and HR as compared to the first one. The decrease in BP reached 52% of its initial level in the control group and 73% in the experimental group (Fig. 12A, curves 1 and 2, respectively, $p < 0.05$). BP recovery in the control group was significantly slower. A statistically significant difference in BP recovery in both groups was observed in the first minutes.

The shape of the ECG signal was not significantly changed, although the number of ventricular extrasistols increased in both groups of animals, especially after the second injection of hydrogen peroxide.

Thus, the bienzyme SOD-CHS-CAT conjugate was shown to significantly prevent changes in the hemodynamics caused by hydrogen peroxide in two animal species. The conjugate effectively neglected hydrogen peroxide's direct influence on smooth muscle and cardiac cells [22], presumably due to the neutralisation of both superoxide and H_2O_2. The effects of SOD-CHS-CAT were reliably detected before the second injection

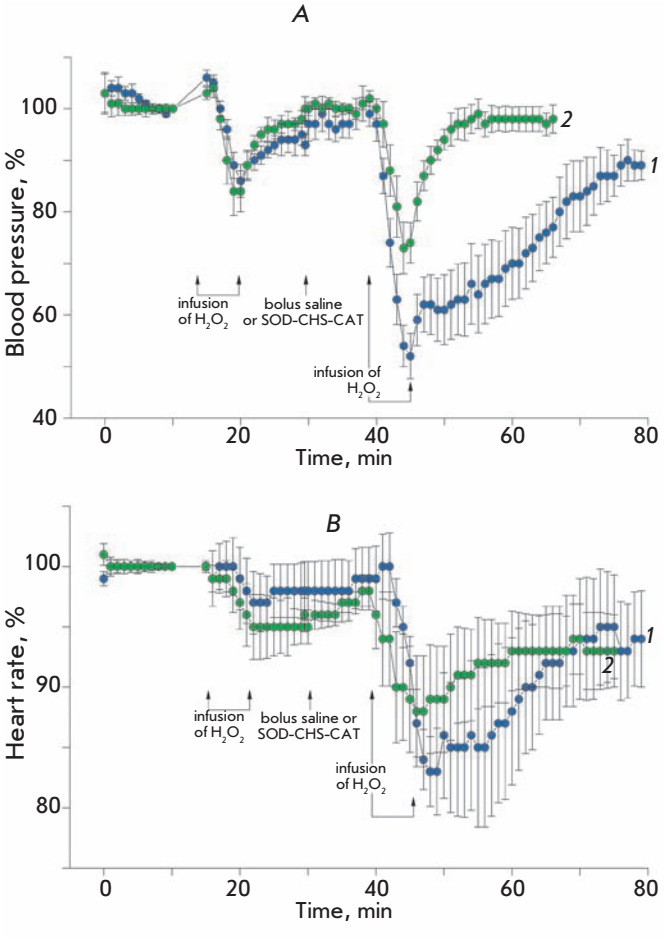


Fig. 12. Alteration of BP (**A**) and HR (**B**, both in % from initial) of anesthetized rats. The arrows show the moment of the beginning and end of the intravenous infusion of 0.8 % hydrogen peroxide (oxidative stress) and bolus administration of saline (n=6, control group, curve 1) or SOD-CHS-CAT conjugate (n=7, test group, curve 2).

of hydrogen peroxide, when the antioxidant systems were weakened by the first injection of H_2O_2.

The preventive effect of the SOD-CHS-CAT conjugate was more pronounced in animals that were administered high doses of the conjugate 3 days before the experiment. In this case, we observed a full recovery of BP and HR after the first injection of H_2O_2. This is the most pronounced antioxidant activity in comparison with the other cases considered. It is possible that preventive administration of the conjugates accounts for a more efficient resistance to the action of hydrogen peroxide through direct and/or indirect pathways. The importance of CAT activity is confirmed by its necessary presence as a component of lecitinized SOD-based therapy of bleomycin-induced pulmonary fibrosis in mice [23]. The dose-dependence of the therapeutic effect of lecitinized SOD is described as a bell-shaped curve. In the range of high SOD concentrations, the introduction of CAT restores the effect of SOD due to the prevention of peroxide accumulation. The presence of both SOD and CAT in the enzyme conjugate accounts for its antithrombotic activity [9, 17] and normalization of BP and HR, which were distorted by hydrogen peroxide infusion (Figs. 11 and 12). The effectiveness of oral and inhalation administration of a modified form of SOD, described in [24] and [23], indicates the importance of developing oral forms of SOD-CHS-CAT as means of preventive antioxidant therapy [25].

CONCLUSIONS

The results of the present study indicate that supramolecular enzymatic conjugates are potentially effective and safe and that they can be used in practical medicine. Covalent cross-linking of enzymatic subunits accounts for a mutual stabilization of biocatalyst activity. Augmented molecular weight and the inclusion of chondroitin sulphate (glycosoaminoglycan of the endothelial glycocalyx) in the conjugate results in a vasoprotective activity of the SOD-CHS-CAT derivate on the inner surface of the vascular wall. The therapeutical effect of the SOD-CHS-CAT is determined by covalent coupling of SOD and CAT activities, which underlies their combined action and formation of harmless products of enzymatic conversion. The supramolecular structure of the nanoconjugate accounts for its previously unknown quality – the capacity to prevent induced platelet aggregation. The activity of the components of the SOD-CHS-CAT conjugate is very pronounced, since the influence of the SOD and CAT activities of the conjugate on the tone of the vascular fragment is similar to that of native enzymes. The contribution of chondroitin sulphate to the antiaggregational effect is also very clear. The described qualities determine the higher antioxidant activity of the SOD-CHS-CAT conjugate, as compared to the other combinations of its individual components. The SOD-CHS-CAT conjugate is well-tolerated, it possesses satisfactorily acute toxicity, normalizing action in relation to hemodynamics under oxidative stress, and a pronounced therapeutical effect. All these factors make the conjugate a prospective drug candidate. This study represents a new approach to the development of therapeutically significant enzymatic conjugates. ●

This work is supported by the Federal targeted scientific program "Research and Development of Priority Directions of Science and Technology", sections "Biocatalytic Technologies for Chemical Synthesis, Analytical Systems and Medicine"

and "Development of Means and Approaches of Diagnostics and Treatment of Ischaemis Heart Disease", RFBR grants 06-08-00011, 06-04-08002-ofi, 09-08-00023, Roszdrav, Rosmedthekhnologii and Russian Ministry of Health and Social Development.

The authors are grateful to academician E.I. Chazov, corresponding member of RAS V.N. Smirnov, corresponding member of RAMS V. V. Kukharchuk, and professor S. A. Boytsov for their attention and support of the present study.

REFERENCES:

1. Marder V.J., Sherry S. // New Engl. J. Med. 1988. V. 318. P. 1512–1520.
2. Anderson A.V., Willerson J.T. // New Engl. J. Med. 1993. V. 329. P. 703–709.
3. Maksimenko A.V. // Mol. Biol. 1995. V. 29. № 1. P. 20–31.
4. Collen D., Lijnen H.R. // Arterioscler. Thromb. Vasc. Biol. 2009. V. 29. P. 1151–1155.
5. Ishihara T., Tanaka K., Tasaka Y., Namba T., Suzuki J., Ishihara T., Okamoto S., Hibi T., Takenaga M., Igarashi R., Sato K., Mizushima Y., Mizushima T. // J. Pharmacol. Exp. Ther. 2009. V. 328. P. 152–164.
6. Taylor J.K., Gardner T.B., Waljee A.K., Dinagno M.J., Schoenfeld P.S. // Aliment. Pharmacol. 2010. V. 31. P. 57–72.
7. Menshikova E.B., Zenkov N.K., Lankin V.Z., Bondar I.A., Trufakin V.A. Okislitelniy stress. Patologicheskie sostoyania I zabolevania (Oxidative Stress. Pathological Conditions and Diseases). Novosibirsk: ARTA, 2008. 248 p.
8. Dubinina E.E. Producti metabolisma kisloroda v funkcionalnoy aktivnosti kletok (Products of Oxygen Metabolism in the Functional Activity of Cells). M.: Medicinskaya pressa, 2006. 400 p.
9. Maksimenko A.V. // Khim-farm. jurn. (Pharmaceutical Chemistry Journal). 2007. V. 4. № 5. P. 3-12.
10. Menshikova E.B., Lankin V.Z., Zenkov N.K., Bondar I.A., Krugovih N.F., Trufakin V.A. Okislitelniy stress. Prooksidanti i antioksidanti (Oxidative Stress. Prooxidants and Antioxidants). M.: Slovo, 2006. 556 p.
11. Maksimenko A.V. // Curr. Pharm. Design. 2005. V. 11. P. 2007–2016.
12. Wight T.N., Merrilees M.J. // Circ. Res. 2004. V. 94. P. 1158–1167.
13. Maksimenko A.V., Tischenko E.G. // Biochemistry (Moscow). 1997. V. 62. P. 1167-1170.
14. Murthy M.R.N., Reid T.J.III, Sicignano A., Tanaka N., Rossmann M.G. // J. Mol. Biol. 1981. V. 152. P. 465–499.
15. Tainer J.A., Getroff E.D., Beem K.M., Richardson J.S., Richardson D.C. // J. Mol. Biol. 1982. V. 160. P. 181–217.
16. Maksimenko A.V., Tischenko E.G. // Biochemistry (Moscow). 1997. Vol. 62. № 10. P. 1163-1166.
17. Maksimenko A.V., Golubykh V.L., Tischenko E.G. // J. Pharmacy Pharmacol. 2004. V. 56. P. 1463–1468.
18. Vavaev A.V., Tischenko E.G., Mokh V.P., Maksimenko A.V. // Tehnol. zhiv. sistem (Life Sciences' Technology). 2009. V. 6. № 3. P. 26-32.
19. Halliwell B., Clement M.V., Long L.H. // FEBS Lett. 2000. V. 486. P. 10–13.
20. Krotz F., Sohn H.-Y., Pohl U. // Arterioscler. Thromb. Vasc. Biol. 2004. V. 24. P. 1988–1996.
21. Maiellaro K., Taylor W.R. // Cardiovasc. Res. 2007. V. 75. P. 640–648.
22. Kapelko V.I. // Ros. fiziol. jurn. (Russian Physiological Journal). 2004. V. 62. № 10. P. 1364–1368.
23. Tanaka K.I., Ishichera T., Azuma A., Kudoh S., Ebina M., Nukiwa T., Sugiyama Y., Tasaka Y., Namba T., Ishihera T., Sato K., Mizushima Y., Mizushima T. // Am. J. Physiol. Lung. Cell. Mol. Physiol. 2010. V. 298. P. L348–L360.
24. Suzuki Y., Matsumoto T., Okamoto S., Hibi T. // Colorectal. Dis. 2008. V. 10. P. 931–934.
25. Maksimenko A.V., Vavaev A.V. // Mol. med. (Molecular Medicine). 2010. № 2. P. 9–14.

"Prostate Cancer Proteomics" Database

S. S. Shishkin[1*], L. I. Kovalyov [1], M. A. Kovalyova [1], K. V. Lisitskaya [1], L. S. Eremina[1], A. V. Ivanov [1], E. V. Gerasimov[1], E. G. Sadykhov[1], N. Y. Ulasova[2], O. S. Sokolova[2], I. Y. Toropygin[3], V. O. Popov[1]

[1] Bach Institute of Biochemistry, Russian Academy of Sciences

[2] Lomonosov Moscow State University

[3] Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences

[*]E-mail: shishkin@inbi.ras.ru

Received 28.09.2010

ABSTRACT **A database of Prostate Cancer Proteomics has been created by using the results of a proteomic study of human prostate carcinoma and benign hyperplasia tissues, and of some human-cultured cell lines (PCP, http://ef.inbi.ras.ru). PCP consists of 7 interrelated modules, each containing four levels of proteomic and biomedical data on the proteins in corresponding tissues or cells. The first data level, onto which each module is based, is a 2DE proteomic reference map where proteins separated by 2D electrophoresis, and subsequently identified by mass-spectrometry, are marked. The results of proteomic experiments form the second data level. The third level contains protein data from published articles and existing databases. The fourth level is formed with direct Internet links to the information on corresponding proteins in the NCBI and UniProt databases. PCP contains data on 359 proteins in total, including 17 potential biomarkers of prostate cancer, particularly AGR2, annexins, S100 proteins, PRO2675, and PRO2044. The database will be useful in a wide range of applications, including studies of molecular mechanisms of the aetiology and pathogenesis of prostate diseases, finding new diagnostic markers, etc.**

KEYWORDS **proteomics, prostate cancer, digital database**

ABBREVIATIONS **2DE—two-dimensional gel electrophoresis, BPH—benign prostate hyperplasia, PCa—prostate cancer, PCP—prostate cancer proteomics**

INTRODUCTION

The first decade of the post-genome era was marked by a rapid development in the field of bioinformatics, the extension of major databases (such as NCBI and UniProt), and the creation of specialised information resources for biomedical research in many countries [1–4]. The impressive resources created in Ireland (UCD-2DPAGE, http://proteomics-portal.ucd.ie:8082/cgi-bin/2d/2d.cgi) [2] and India (Human Proteinpedia, www.humanproteinpedia.org) [3] make the state of things in Russia pale in comparison.

Currently, one of the most important tasks for biomedical research is to find efficient prostate cancer (PCa) biomarkers which would enable new diagnostic methods [5–8]. The fact that in recent years the PCa incidence rate has dramatically increased worldwide [9, 10], and particularly in Russia, making PCa the most frequent male oncological disease in some countries [11, 12], is reason enough to pay close attention to this disease. In early diagnostics of PCa at the moment it is important to establish the presence of one of the most studied biomarkers, the so-called Prostate-Specific Antigens (PSA), in the blood. The test, however, is known to produce a significant number of false-positive and false-negative results, leading to the wrong clinical and financial outcomes [5, 7]. Therefore, in the U.S. and in other Western countries, new PCa biomarkers are being actively sought, an initiative recently stimulated by the development of proteomic and other post-genome technologies [6, 8, 13].

Since 2005, the Bach Institute of Biochemistry, in collaboration with other research and medical institutions, has been researching new PCa biomarkers by utilising various proteomic technologies [14, 15]. In 2009, the "Prostate Cancer Proteomics" (PCP, http://ef.inbi.ras.ru/) national database was created in order to facilitate this research, summarising experimental and referenced published data and providing links to several other biomedical Internet databases. This paper describes the structure and capabilities of the new, extended PCP version.

MATERIALS AND METHODS

Biomaterials— Biopsy and surgical samples of prostate tissues from patients with PCa (n = 72) and benign prostatic hyperplasia (BPH, n = 69) were provided by staff members of the Urology Department of the Botkin Clinical Hospital (Moscow). Diagnosis

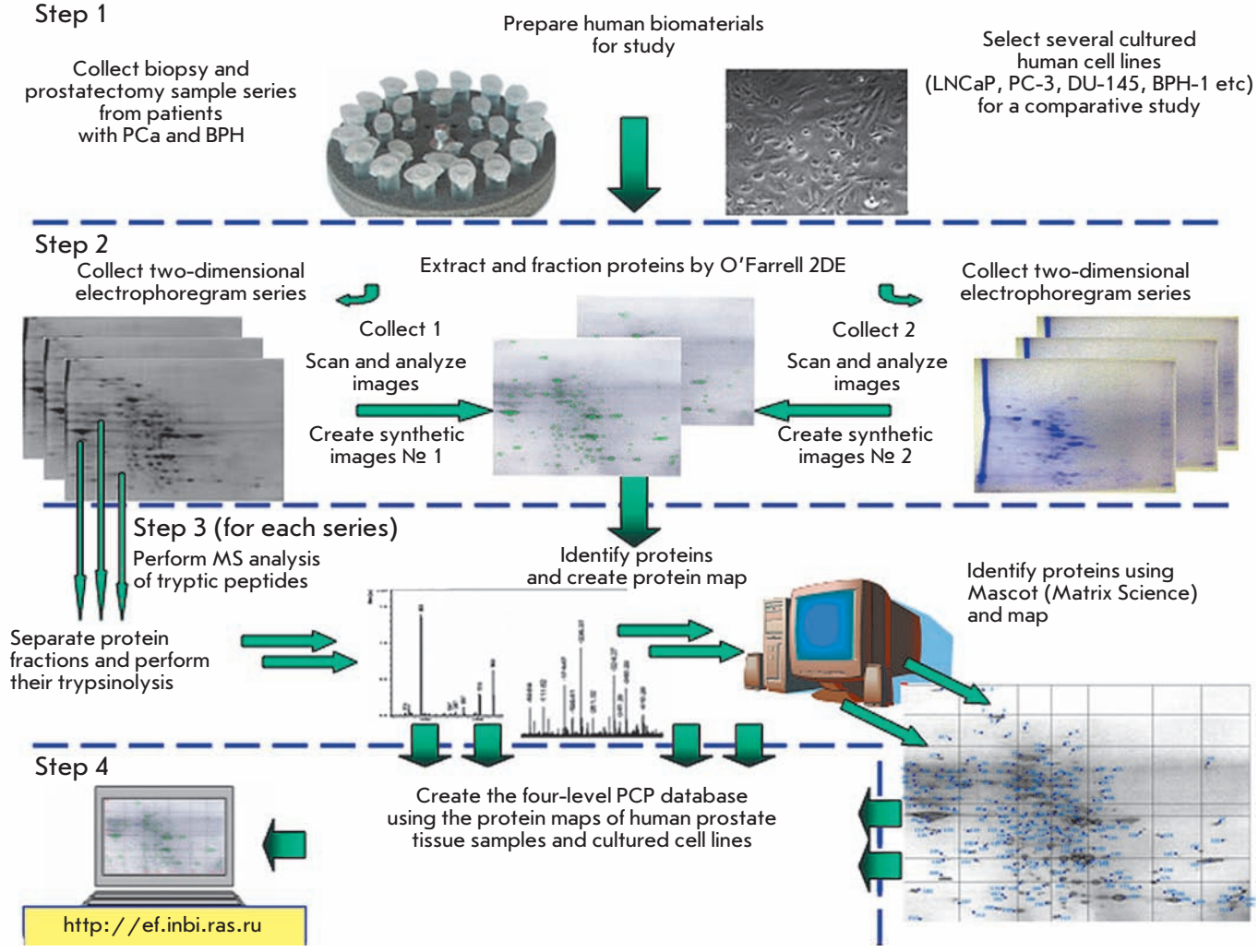


Fig. 1. Main steps in the proteomic study of prostate tissue samples from patients with malignant and benign tumors, as well as from several cultured human cell lines.

was performed using clinical, histological, and immunochemical (PSA level) tests. Histological verification was performed via U.S.-controlled transrectal multifocal needle biopsy; up to 18 tissue samples from various prostate zones per patient were taken [16, 17]. All PCa cases were found to be adenocarcinoma. Gleason score was determined by following the standard procedure [16, 17].

In parallel tests, we analysed the proteins of the PC-3 (ACC 465), DU-145 (ACC 261), and BPH-1 (ACC 143) cell cultures purchased from the German Collection of Microorganisms and Cell Cultures, as well as the proteins of cultured cells of the LNCaP line provided by Dr. I. G. Shemyakin (Obolensk National Science Centre for Applied Microbiology and Biotechnology). The cells were cultured in the RPMI-1640 medium with HEPES, sodium pyruvate, gentamicine and 20% fetal bovine serum (FBS) [18], using cell culture plastic (Costar, USA and Nunc, Denmark) in a CO_2-incubator (Sanyo, Japan). In addition, we studied proteins from the cultured cells of two lines of human rhabdomyosarcoma (A-204 and RD) purchased from the Ivanovsky Virology Institute, RAMS, and proteins from the cultured normal human myoblasts kindly provided by Dr T. B. Krohina [19].

The preparation of protein extracts, their O'Farrell 2DE fractioning, Coomassie Blue R-250 and silver nitrate staining, and 2DE analysis were performed following the techniques described in [20, 21]. In addition, we used a 2DE procedure with isoelectric focusing using IPG-PAGE and Ettan IPGphor 3 kit (GE Healthcare), according to the manufacturer's protocol. Pro-

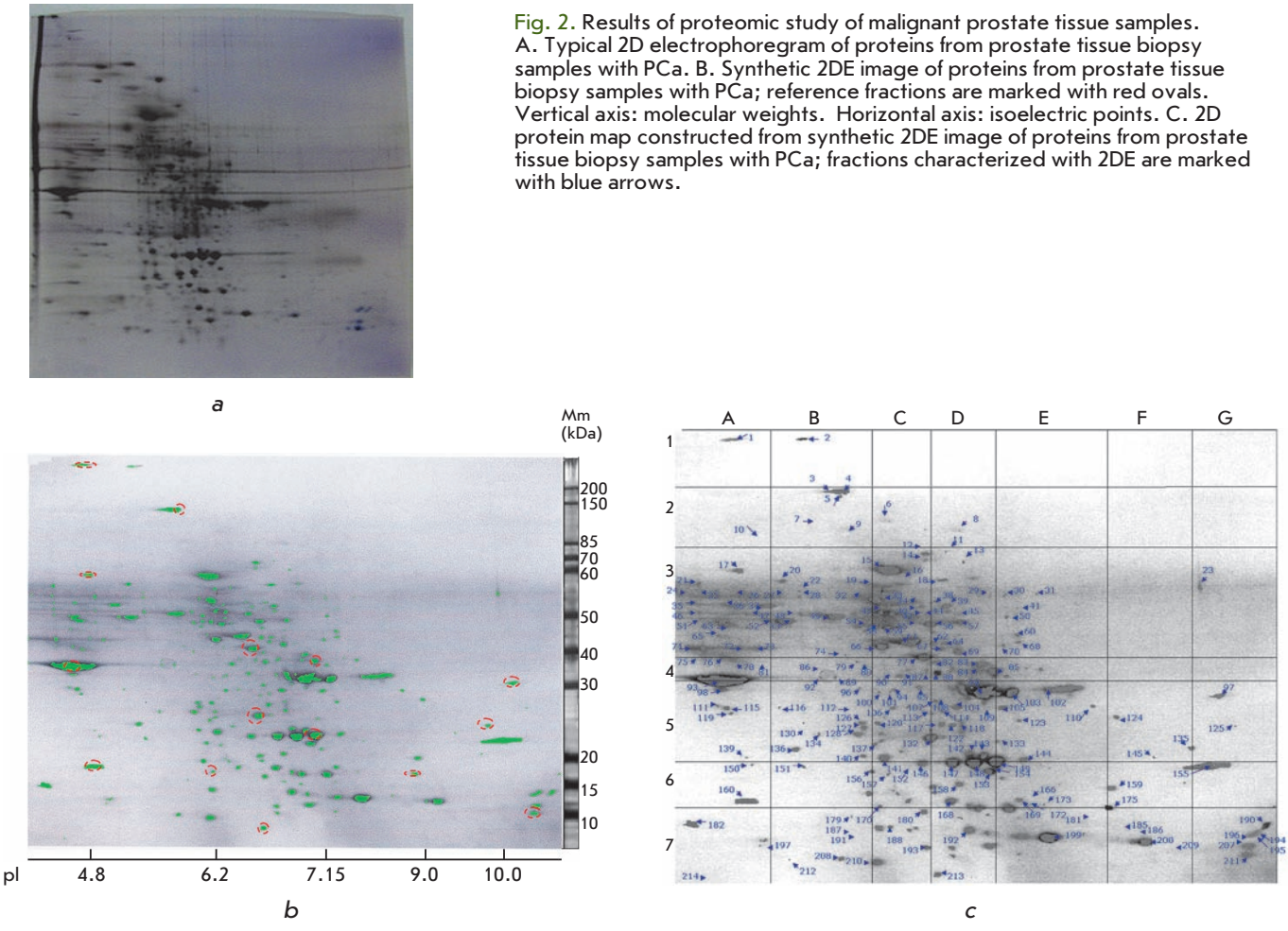


Fig. 2. Results of proteomic study of malignant prostate tissue samples. A. Typical 2D electrophoregram of proteins from prostate tissue biopsy samples with PCa. B. Synthetic 2DE image of proteins from prostate tissue biopsy samples with PCa; reference fractions are marked with red ovals. Vertical axis: molecular weights. Horizontal axis: isoelectric points. C. 2D protein map constructed from synthetic 2DE image of proteins from prostate tissue biopsy samples with PCa; fractions characterized with 2DE are marked with blue arrows.

teins were identified with MALDI-TOF MS and MS/MS using an Ultraflex instrument (Bruker) at a 336-nm UV laser beam in a 500–8000 Da cation mode calibrated using reference trypsin autolysis peaks and processed with Mascot software, Peptide Fingerprint option (Matrix Science, USA) [21, 22]. The proteins were identified by matching experimental masses with the masses of proteins listed in the NCBI Protein and SwissProt/TrEMBL databases. The accuracy of monoisotopic masses measured in the reflection mode calibrated with autolytic trypsin peaks was 0.005%, and the accuracy of the fragment masses was ±1 Da. Hypothetical proteins identified with MALDI-TOF MS corresponding to fragments of the full-size proteins, which are products of corresponding genes, were revealed with MS/MS. The molecular masses of protein fractions were determined using the ultrapure recombinant protein sets SM0661 (10–200 kDa) and SM0671 (10–170 kDa) (Fermentas). The measurement of the optical density of 2DE images and/or their fragments was performed following scanning (Epson expression 1680) or digital photography (Nikon 2500 or Canon PowerShot A1000 IS). Digital image processing with densitometry of the protein fractions was performed with Melanie ImageMaster, versions 6 and 7 (Genebio).

Data logging and processing for the Prostate Cancer Proteomics multilevel database were done with various software packages, including MapThis!, Molly Penguin Software, Mozilla Firefox, and some Microsoft Office applications. A MySQL-based interactive database was used which could be updated and modified online using any computer with Internet connection. The BIOSTAT and Microsoft Office Excel 2003 software packages were used for statistical analysis.

RESULTS AND DISCUSSION

According to the conventional proteomics strategy developed in the late 20th century, the national PCP database was created in several consecutive steps which involved systematic characterisation of proteins in prostate tissue samples obtained from benign and malignant tumors (Fig. 1) ([23, 24]). Proteins from several

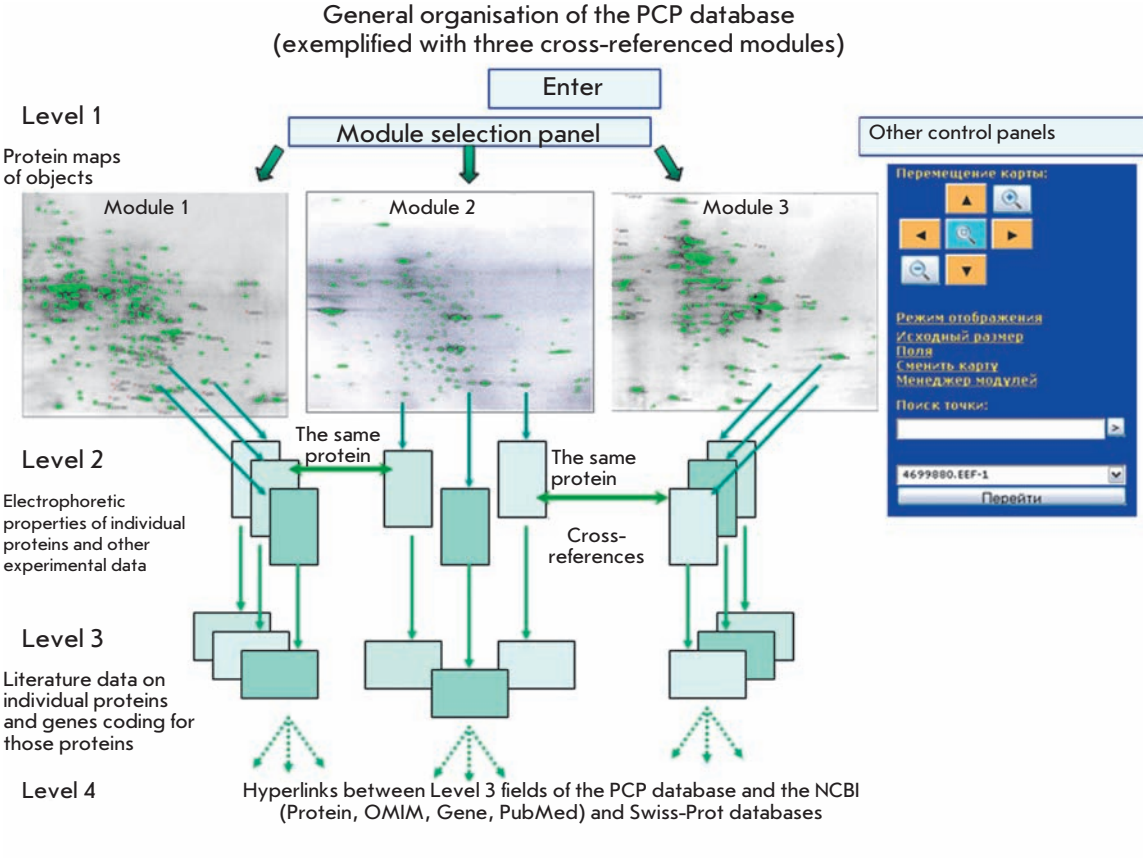


Fig. 3. General organisation of the "Prostate Cancer Proteomics" (PCP) database.

cultured human cell lines were studied in parallel experiments (Fig. 1).

The first step was to make series of 2DE protein samples (50 or more) by fractioning dozens of bioptates or prostate tissue samples (from 30 or more patients). Figure 2A illustrates a typical 2DE of the PCa prostate tissue proteins. The 2DE series for cell line proteins were created with 20 2DE, taking into account the homogeneity of the analyte.

The distribution of 2DE protein fractions was registered and stored as graphic *.tif files. The images of the entire 2DE and (in some cases) their segments were produced by scanning and/or digital photography. The relevance of the selected 2DE was assessed by comparing protein fractioning results using digital image matching [23, 24].

The second step was to construct synthetic 2D maps of the proteins. 2DE images from each series were standardised with Melanie ImageMaster software using 15 selected reference points corresponding to easily identifiable major protein fractions. Figure 2B shows reference points on the 2DE images of proteins from prostate tissue samples.

Each image was analysed using the Cummings technique [25] with some modifications [20, 24]. The analysis was based on dividing the images into 49 rectangular fragments, the sides of which formed six horizontal and six vertical standard lines and the sides of the 2DE image itself. The points for plotting the horizontal lines were determined by special molecular-weight-marker proteins, which were placed on each gel plate before fractioning in the second direction (SDS-PAGE). Thus, the protein fractions located on the corresponding horizontal lines have identical molecular weights. For plotting the vertical lines, protein markers with previously measured pI values were used [20, 24]. As a result, each image analyzed consisted of 49 rectangular fragments, each usually containing no more than 10 protein fractions (only 4 fragments contained more than 20 fractions). Image fragmenting significantly simplified the image comparison and construction of synthetic 2D maps.

The described procedure was performed with the 60 best 2DE images of proteins from BHP samples and with 70 2DE images of proteins from PCa samples. The comparison of standardised BHP and PCa 2DE images showed that the coordinates of more than 95% of protein fractions were constant. Quantitative or qualitative variations in the coordinates were observed for less than 5% of the fractions. The variations could be caused

Table 1. Description of the PCP modules and the number of identified proteins by modules.

Modules—synthetic maps of proteins in specific objects (fractioning technique)	Proteins identified
Prostate biopsy samples, PCa and BPH (IEF-PAGE*)	165
LNCaP (IEF-PAGE*)	60
LNCaP (IPG-PAGE)	18
PC-3 (IEF-PAGE*)	25
BPH-1 (IEF-PAGE*)	24
Rhabdomyosarcoma (IEF-PAGE*)	29
Normal human fibroblasts (IEF-PAGE*)	38
* modification [20]	

by genetic factors (e.g. single nucleotide polymorphism) or a different expression of corresponding genes, as well as differences in tissue composition of the samples and the pathology's intensity.

Having made sure that the positions of the majority of the protein fractions were constant, we were able to construct the 2D maps of prostate tissue proteins from patients with BHP and PCa. After that, we performed a fragment-by-fragment comparison of the 2D maps constructed. The fraction patterns in BHP and PCa maps were found to be quite similar, the difference being that about 20 fractions present in the PCa map were either present in a much smaller amount in the BHP map or absent altogether. We paid particular attention to those fractions in further study, as described below. In general, as a result of our analysis, an integrated synthetic 2D map of human prostate proteins was constructed that contains more than 200 protein fractions in the ranges of Mw 8.5–450 kDa and pI 4.5–11.5 (Fig. 2C). Each fraction was assigned a unique seven-digit number; the first four digits representing the logarithm of the fraction's molecular weight, and the next three digits representing the value of the isoelectric point expressed in units described in [20, 24].

The same procedure was applied to construct other synthetic maps of proteins from cultured human cell lines, although much fewer 2DE images were used for those maps.

Thus, each of the constructed maps contained data on the electrophoretic properties of the protein fractions (represented by their coordinates) in the corresponding object. These maps were in *.jpg format (with a resolution of at least 300 dpi), constituting Level 1 of the database. Further studies and analysis of data on proteins were based on the information contained in the Level 1 maps. Therefore, the synthetic maps represented original modules enabling the characterizing and formalizing of the biochemical properties of the proteins studied. There are currently seven modules in PCP (Table 1). There is a special panel enabling navigation among the modules. The 2D maps are scaleable, and the user can mark certain proteins on the maps and create links (buttons) for accessing other levels–second, third, and fourth–that contain data on the proteins. The database also automatically displays 2D coordinates (along the two fractioning axes) as the cursor moves around the map. The general organisation of the PCP database is presented in Fig. 3.

The third step in the proteomic study was to identify individual protein fractions. The proteins were mainly identified by mass-spectrometry: the results are presented in Table 1.

As Table 1 shows, there is a total of 359 identified proteins in PCP. Among them there are many well-known proteins, such as the enzymes responsible for glycolysis (glyceraldehyde-3-phosphate dehydrogenase, triose-phosphate isomerase, etc.) and other metabolic processes, as well as cytoskeletal (actin, transgelins, etc.) and mitochondrial (porins, superoxide dismutase, etc.) proteins. Some of the identified proteins, for instance transgelins [21, 22], were represented with several isoforms.

We paid particular attention to the identification of the protein fractions which differed qualitatively or quantitatively in the prostate tissue samples from patients with BHP and PCa. We previously reported on the results of identification of two potential PCa biomarkers, the proteins AGR2 [14] and Dj-1 [26]. In total, we succeeded in identifying 17 potential PCa biomarkers, some of which are new. Table 2 provides a short description of the potential PCa biomarkers. For example, Fig.4 shows the results of MS identification of one of the new biomarkers, protein PRO2675, which contains an albumin domain in its primary structure.

For each protein identified (and marked with a "button" on the 2D map), the second information Level was formed, comprising a standardized system of 15 fields for the entry of text and graphical data obtained during characterization of the corresponding protein fraction. In four fields, general information about the protein is entered, in the next six fields the identification results are entered, and in the other five fields additional in-

Table 2. Potential PCa biomarkers in the "Prostate biopsy samples, PCa and BPH" module and other PCP modules

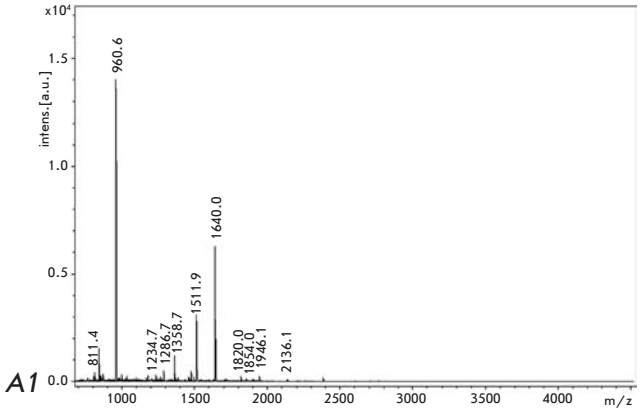
Unique identifier*	Protein (synonyms and symbol in PCP)	Numbers in NCBI** and *Swiss-Prot*	Additional information in PCP and references***
5653580	**Ferritin light chain complex (K-(L)F)**	182516, *P02792*	[15]
4785508 (4799550)	Chaperonin (HSPD1)	31542947, NP_002147, 118190, *P10809*	Found in rhabdomyosarcoma cells; {Bindukumar B. et al. 2008, 18646040}
4716560 (4756612)	Protein-disulphide isomerase (ER60)	7437388, *P30101*	[15]
4531685	N-acetylneuraminat phosphate synthase (NANS)	12652539, AAH00008, NP_061819, 605202, Q9NR45	Found in rhabdomyosarcoma cells; [15]
4502675	Annexin 2, isophorm 2 (ANXA2-i2)	4757756, NP_004030, 151740, *P07355*	{Shiozawa Y. et al. 2008, 18636554; Hastie C. et al. 2008, 18211896}
4454692	**Unknown protein PRO2675 containing the albumin domain (PRO2675)**	7770217	[15]
4447605	Protein 29 of endoplasmic reticulum, isophorm 1 (Erp29)	5803013, NP_006808, 602287, *P30040*	{Myung J.K. et al. 2004, 15598346}
4352630 (4342620)	Dj-1 protein (Dj-1)	50513593, 1SOA_A, 606324, *Q99497*	{Bindukumar B. et al. 2008, 18646040}
4356607 (4344615)	**Dj-1 protein, electrophoretic isophorm (Dj-1-ei)**	31543380	
4336712 (4301795)	Prostatic binding protein (neuropolypeptide h3, PEBP1)	21410340, AAH31102, 604591, *P30086*	[15]; {Li et al. 2008, 18161940; Woods Ignatoski K.M. et al. 2008, 18722266}
4286750 (4290620)	NM23B-protein, nucleoside phosphate kinase B	4505409, NP_002503, 156491, *P22392*	{Johansson B. et al. 2006, 16705742}
4255880	**Unnamed protein (NEDO human cDNA sequencing project, tissue type = "testis") (NEDO)**	21758704, BAC05360	
4204630	Fatty acid binding protein, isophorm 5 (E-FABP)	30583737, AAP36117, 605168, *Q01469*	[15]; {Morgan E.A. et al. 2008, 18360704}
4279900	AGR2 (AGR2)	37183136, AAQ89368, 606358. *Q4JM47*	[14, 15]; {Zhang J.S. et al. 2005, 15834940; weitzig D.R. et al. 2007, 17694278}
41811130	Hystone H3, 3A family (H3f3a)	55665435	[15]
4161675	**Unknown protein PRO2044 containing the albumin domain (PRO2044)**	6650826	[15]
4021610	S100 calcium binding protein A11 (S100A11)	12655117, AAH01410, 603114, *P31949*	{Rehman I. et al. 2004, 15668896; Schaefer K.L. et al. 2004, 15150091}

* Numbers in the "LNCaP (IEF-2DE Modification)" module are given in parenthesis
** Numbers from the NCBI databases are given in the following order: Protein, Genbank and/or Nucleotide, OMIM
*** Publications listed in the References section of this article are given in square brackets; publications from the PubMed database are given in curly brackets.
Note: a new potential biomarkers are shown in bold

formation is entered. The filled Level 2 fields for one of the potential PCa biomarkers, protein NANS (N-acetylneuraminic acid phosphate synthase), are shown in Fig. 5.

The same protein could be present in more than one object. Therefore, within Level 2, one can use the control panel to create cross-links between identical proteins in different modules. An example of such cross-referencing for protein Dj-1 is presented in Fig. 6.

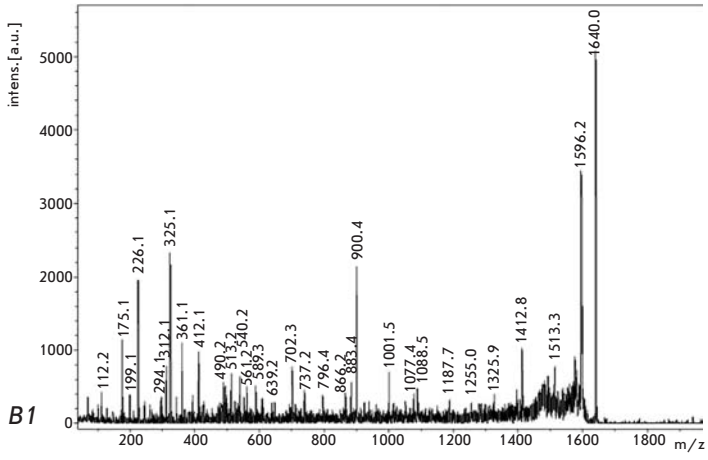
The majority of the 359 identified fractions were well-known proteins (and/or their electrophoretic isophorms) described in the literature and various databases. Some information on those proteins, relating to the PCP scope, constituted the third information Level of the database. Level 3 is a standardized system of 23 fields for the entry of text and graphical data. In twelve fields information about the protein is entered, in the next six fields information about the gene coding for



A1

A2

Start-end	Experimental Mw	Calculated Mw	Peptide sequence
1 - 16	1706.77	1705.82	MPADLPSLAADFVESK + Oxidation (M)
27 - 39	1639.82	1638.79	DVFLGMFLYEYAR + Oxidation (M)
76 - 92	2045.05	2044.09	VFDEFKPLVEEPQNLIK
106 - 113	960.52	959.56	FQNALLVR
118 - 131	1511.74	1510.84	VPQVSTPTLVEVSR
249 - 263	1763.73	1762.77	AVMDDFAAFVEKCCK + Oxidation (M)



B1

#	Immon.	a	a*	a°	b	b*	b°	Seq	v	w	w*	y	y*	y°	#
1	101.11	101.11	84.08		129.10	112.08		K							15
2	72.08	200.18	183.15		228.17	211.14		V	1467.78	1480.80		1511.84	1494.82	1493.83	14
3	70.07	297.23	280.20		325.22	308.20		P	1370.73	1369.73		1412.77	1395.75	1394.76	13
4	101.07	425.29	408.26		453.28	436.26		Q	1242.67	1241.67		1315.72	1298.70	1297.71	12
5	72.08	524.36	507.33		552.35	535.32		V	1143.60	1156.62		1187.66	1170.64	1169.65	11
6	60.04	611.39	594.36	593.38	639.38	622.36	621.37	S	1056.57	1055.57		1088.59	1071.57	1070.58	10
7	74.06	712.44	695.41	694.42	740.43	723.40	722.42	T	955.52	968.54	970.52	1001.56	984.54	983.55	9
8	70.07	809.49	792.46	791.48	837.48	820.46	819.47	P	858.47	857.47		900.51	883.49	882.50	8
9	74.06	910.54	893.51	892.53	938.53	921.50	920.52	T	757.42	770.44	772.42	803.46	786.44	785.45	7
10	86.10	1023.62	1006.59	1005.61	1051.61	1034.59	1033.60	L	644.34	643.34		702.41	685.39	684.40	6
11	72.08	1122.69	1105.66	1104.68	1150.68	1133.66	1132.67	V	545.27	558.29		589.33	572.30	571.32	5
12	102.05	1251.73	1234.70	1233.72	1279.73	1262.70	1261.72	E	416.23	415.23		490.26	473.24	472.25	4
13	72.08	1350.80	1333.77	1332.79	1378.79	1361.77	1360.78	V	317.16	330.18		361.22	344.19	343.21	3
14	60.04	1437.83	1420.80	1419.82	1465.83	1448.80	1447.82	S	230.12	229.13		262.15	245.12	244.14	2
15	129.11							R	74.02	73.03		175.12	158.09		1

B2

C

```
  1 MPADLPSLAA DFVESKDVCK NYAEAKDVFL GMFLYEYARR HPDYSVVLLL RLAKTYETTL
 61 EKCCAAADPH ECYAKVFDEF KPLVEEPQNL IKQNCELFEQ LGEYKFQNAL LVRYTKKVPQ
121 VSTPTLVEVS RNLGKVGSKC CKHPEAKRMP CAEDYLSVVL NQLCVLHEKT PVSDRVTKCC
181 TESLVNRRPC FSALEVDETY VPKEFNAETF TFHADICTLS EKERQIKKQT ALVELVKHKP
241 KATKEQLKAV MDDFAAFVEK CCKADDKETC FAEEGKKLVA ASQAALGL
```

Fig. 4. Results of mass-spectrometric identification of the PRO 2675 protein.
A. Mass-spectrum of tryptic peptides. (1) MALDI-TOF MS data. (2) Peptide identification by Mascot.
B. Mass-spectrum of one of the tryptic peptides (1) MALDI-TOF MS data. (2) Peptide identification by Mascot.
C. Amino acid sequence of the PRO 2675 protein (records AAF69644.1 GI:7770217 in the NCBI Protein database); amino acid residues of the tryptic peptides are printed in red; the peptide's whose sequence was identified by MALDI-TOF MS/MS and is highlighted in grey; the sites corresponding to the albumin domain are underlined.

Experimental data (Level 2)

General properties	
Fraction (Protein)	N-acetylneuraminic acid phosphate synthase, NANS Edit
Localization, number	D4, 4531685 Edit
Mw, kDa	38.0 Edit
pI (experimental)	6.85 Edit
Identification	
Method	MALDI TOF MS Edit
Sequence coverage, %	58% Edit
Mascot (Matrix Science, USA) search results	Start-End Observed Mr(expt) Mr(calc) ppm Miss Peptide 66 - 74 1064.56 1063.55 1063.57 -0.02 0 ALERPYTSK 87 - 96 1331.59 1330.58 1330.61 -0.03 0 HLEFSHDQYR 132 - 145 1540.68 1539.68 1539.72 -0.04 0 VGSGDTNNFPYLEK 150 - 166 1919.79 1918.78 1918.83 -0.04 0 GRPMVISSGMQSMDTMK 4 Oxidation (M) 199 - 205 866.44 865.43 865.45 -0.03 0 VISEYQK 206 - 233 2767.42 2766.41 2766.50 -0.08 0 LFPDIPIGYSGHETGIAISVAAVALGAK 247 - 264 1866.88 1865.87 1865.91 -0.04 0 GSDHSASLEPGELAELVR 301 - 315 1693.83 1692.82 1692.87 -0.05 0 IPEGTILTMDMLTVK 2 Oxidation (M) 321 - 333 1448.71 1447.70 1447.73 -0.04 0 GYPPEDIFNLVGK 335 - 355 2401.08 2400.07 2400.14 -0.06 0 VLVTVEEDDTIMEELVDNHGK Oxidation (M) Edit
Figure	Edit
Amino acid sequence (without signal peptides)	0001 MPLELELCPG RWVGGQHPCF IIAEIGQNHQ GDLDVAKRMI RMAKECGADC 0051 AKFQKSELEF KFNRKALERP YTSKHSWGKT YGEHKRHLEF SHDQYRELQR 0101 YAEEVGIFFT ASGMDEMAVE FLHELNVPFF KVGSGDTNNF PYLEKTAKKG 0151 RPMVISSGMQ SMDTMKQVYQ IVKPLNPNFC FLQCTSAYPL QPEDVNLRVI 0201 SEYQKLFPDI PIGYSGHETG IAISVAAVAL GAKVLERHIT LDKTWKGSDH 0251 SASLEPGELA ELVRSVRLVE RALGSPTKQL LPCEMACNEK LGKSVVAKVK 0301 IPEGTILTMD MLTVKVGEPK GYPPEDIFNL VGKKVLVTVE EDDTIMEELV 0351 DNHGKKIKS Edit
Comments to amino acid sequence	Amino acid sequence of the NANS protein. Matched peptides shown in Red. Edit
Supplementary identification information	
Supplementary information	MS/MS Fragmentation of VGSGDTNNFPYLEK Found in gi|8453156, N-acetylneuraminic acid phosphate synthase [Homo sapiens] Match to Query 16: 1539.695159 from(1540.702435,1+) Edit
Tandem mass-spectrometry	Monoisotopic mass of neutral peptide Mr(calc): 1539.72 Ions Score: 22 Expect: 8.7

Fig. 5. Level 2 fields for the NANS protein.

Fig. 6. Control panel with cross-references for the Dj-1 protein from the "Prostate biopsy samples, PCa and BPH" module.

Identical proteins in other modules			
Module	Point	Link	Disconnect
LNCaP (IPG-2DE)	4301630.Dj-1	Link	Disconnect
LNCaP (IEF-2DE modification)	4342635.Dj-1	Link	Disconnect
Rhabdomyosarcoma	4342690.DJ-1	Link	Disconnect
PC3	4342630.DJ-1	Link	Disconnect
Add			

the protein is entered, and in the next three fields information about the protein's polymorphism is entered. In the other fields, selected references to publications about the general properties of the proteins, as well as its oncological properties, are entered (Fig. 7).

The Level 3 text fields can contain hyperlinks to various Internet databases, such as NCBI's Protein, OMIM and PubMed, and SwissProt. This feature allowed the creation of the fourth information Level, providing the user with prompt access to international databases containing, in particular, the results of human genome sequencing.

The PCP database is an interactive MySQL-based web resource located at http://ef.inbi.ras.ru and can

References	
General	PMID: 9790916 1. Thompson DA, Weigel RJ. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell. Biochem. Biophys. Res. Commun. 1998. V.251(1), P.111-116. PMID: 12975309 2. Clark HF, Gurney AL, Abaya E, Baker K, Baldwin D, Brush J, Chen J, Chow B, Chui C, Crowley C, Currell B, Deuel B, Dowd P, Eaton D, Foster J, Grimaldi C, Gu Q, Hass PE, Heldens S, Huang A, Kim HS, Klimowski L, Jin Y, Johnson S, Lee J, Lewis L, Liao D, Mark M, Robbie E, Sanchez C, Schoenfeld J, Seshagiri S, Simmons L, Singh J, Smith V, Stinson J, Vagts A, Vandlen R, Watanabe C, Wieand D, Woods K, Xie MH, Yansura D, Yi S, Yu G, Yuan J, Zhang M, Zhang Z, Goddard A, Wood WI, Godowski P, Gray A. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res. 2003. V.13(10), P.2265-2270. PMID: 15340161 3. Zhang Z, Henzel WJ. Signal peptide prediction based on analysis of experimentally verified cleavage sites. Protein Sci. 2004. V.13(10). P.2819-2824. Edit
Oncological	PMID: 12592373 1. Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, Daviet L, Legrain P, Parekh R, Harris AL, Terrett JA. hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. Br. J. Cancer. 2003. V.88(4). P.579-585. PMID: 15532095 2. Kristiansen G, Pilarsky C, Wissmann C, Kaiser S, Bruemmendorf T, Roepcke S, Dahl E, Hinzmann B, Specht T, Pervan J, Stephan C, Loening S, Dietel M, Rosenthal A. Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol. 2005. V.205(3). P.359-376. PMID: 15867376 3. Liu D, Rudland P.S, Sibson DR, Platt-Higgins A, Barraclough R. Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. Cancer Res. 2005. V.65(9). P.3796-3805. PMID: 15958538 4. Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, Meropol NJ, Weiner LM, Cohen SJ, Moreno JG, Connelly MC, Terstappen LW, O'Hara SM. Global Gene Expression Profiling of Circulating Tumor Cells. Cancer Res. 2005. V.65. P.4993-4997. PMID: 15834940 5. Zhang JS, Gong A, Cheville JC, Smith DI, Young CY. AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. Genes Chromosomes Cancer. 2005. V.43(3) P.249-259. PMID: 17022460 6. Kovalev LI, Shishkin SS, Khasigov PZ, Dzeranov NK, Kazachenko AV, Toropygin Ilu, Mamykina SV. Identification of AGR2 protein, a novel potential cancer marker, using proteomics technologies. Prikl Biokhim Mikrobiol. 2006. V.42(4). P.480-484. PMID: 18829536 7. Ramachandran V, Arumugam T, Wang H, Logsdon CD. Anterior gradient 2 is expressed and secreted during the development of pancreatic cancer and promotes cancer cell survival. Cancer Res. 2008. V.68(19). P.7811-7818. Edit

Fig. 7. Level 3 fields for specially selected references to publications on the AGR2 protein.

be accessed from any computer connected to the Internet using the Mozilla Firefox and Microsoft Internet Explorer browsers. There are three access permission categories: "Guest," "Manager," and "Administrator," each giving certain rights for working with PCP. In particular, users with "Manager" access permission can make entries into and correct the Level 2 and 3 fields, while users with the "Administrator" category access permission, have the ability to expand the database by creating new modules and new functional elements. Users with "Guest" access can browse all fields but cannot edit them.

In conclusion, our work resulted in the creation of an original multi-module national database entitled "Prostate Cancer Proteomics," which summarizes data on the proteins in prostate tissue collected from patients with BHP and PCa, as well as on proteins in several human cell lines. This is very promising in the further use of proteomic and other biochemical data. We are hopeful that the PCP database will be useful to biochemists and other biomedical scientists, making their research on PCa more efficient. ●

This work was supported by the Moscow Department of Science and Industrial Politics (Government contracts № 8/3-373n-08 and 8/3-375n-08).

REFERENCES

1. Gottlieb B., Beitel L.K., Wu J.H., Trifiro M. // Hum. Mutat. 2004. V. 23. P. 527–533.
2. Westbrook J.A., Wheeler J.X., Wait R., Welson S.Y., Dunn M.J. // Electrophoresis. 2006. V. 27. P. 1547–1555.
3. Kandasamy K., Keerthikumar S., Goel R., Mathivanan S., Patankar N., Shafreen B., Renuse S., Pawar H., Ramachandra Y.L., Acharya P.K., Ranganathan P., Chaerkady R., Keshava Prasad T.S., Pandey A. // Nucleic Acids Res. 2009. V. 37 (Database issue). P. D773–D781.
4. Vizcaino J.A., Cote R., Reisinger F., Barsnes H., Foster J.M., Rameseder J., Hermjakob H., Martens L. // Nucleic Acids Res. 2010. V. 38 (Database issue). P. D736–742.
5. Stamey T.A., Caldwell M., McNeal J.E., Nolley R., Hemenez M., Downs J. // J. Urol. 2004. V. 172. P. 1297–1301.
6. Zhang J.S., Gong A., Cheville J.C., Smith D.I., Young C.Y. // Genes Chromosomes Cancer. 2005. V. 43. № 3. P. 249–259.
7. Lim L.S., Sherin K. // Am. J. Prev. Med. 2008. V. 34. № 2. P. 164–170.
8. Leman E.S., Getzenberg R.H. // J. Cell Biochem. 2009. V. 108. № 1. P. 3–9.
9. Zlokachestvennie novoobrazovania v Rossii v 2008. Zabolevaemost i smertnost (Malignant Tumors in Russia in 2008. Morbidity and Mortality) / ed. Chissov V.I., Starinsky V.V., Petrova G.V. M.: «MEDpress-inform», 2008. P. 18.
10. Zlokachestvennie novoobrazovania v Rossii i SNG v 2002 (Malignant Tumors in Russia and CIS in 2002). N. N. Blokhin Russian Cancer Research Center of Russian Academy of Medical Sciences / ed. Davidov M.I., Aksel E.M. M.: «MIA», 2004. 256 p.
11. Jemal A., Siegel R., Ward E., Murray T., Xu J., Smigal C., Thun M.J. // CA Cancer J. Clin. 2007. V. 57. № 1. P. 43–66.
12. Maddams J., Brewster D., Gavin A., Steward J., Elliott J., Utley M,. Muller H. // Br. J. Cancer. 2009. V. 101. № 3. P. 541–547.
13. Primrose S.B., Twyman R.M. Genomika. Rol v Medicine (Genomics. Applications in Human Biology). M.: «BINOM. Laboratoria znaniy, 2008. 277 p.
14. Kovalyov L.I., Shishkin S.S., Hasigov P.Z., Dzeranov N.K., Kazachenko A.V., Kovalyova M.A., Toropygin I.Y., Mamikina S.V. // Prikladnaya biokhimia i mikrobiologia (Applied Biochemistry and Microbiology). 2006. V.42. №4. P.480-484.
15. Shishkin S.S., Dzeranov N.K., Totrov K.I., Kazachenko A.V., Kovalyov L.I.., Eremina L.S., Kovalyova M.A., Toropygin I.Y. // Urologiia (Urology). 2009. V. 1. P. 56-58.
16. Kogan M.I., Loran O.B., Petrov S.B. Radikalnaya khirurgia raka predstatelnoy zhelezi (Radical Surgery of Prostate Cancer). M.: «Geotar-Media», 2006. 392 p.
17. Shishkin S.S., Kovalyov L.I.., Kovalyova M.A., Krakhmaleva I.N., Eremina L.S., Makarov A.A., Lisitskaya K.V., Loran O.B., Veliev E.I., Okhrits V.E. Problemi ranney diagnostiki raka prostati i vosmozhnosti primenenia novih potencialnih biomarkerov (The Problems of Early Diagnostics of Prostate Cancer and Possibilities of Using of New Potential Biomarkers. (Informational-methodical letter)). M.: OOO "Originalnaya compania", 2009. 45 p.
18. Chernikov V.G., Terehov S.M., Krohina T.B., Shishkin S.S., Smirnova T.D., Lunga I.N., Adnoral N.V., Rebrov L.B., Denisov-Nikolsky Y.I., Bikov V.A. // Bull. eksp. biol. med. (Bulletin of Experimental Biology and Medicine). 2001. V. 131. № 6. P. 680-682.
19. Krohina T.B., Shishkin S.S., Raevskaya G.B., Kovalyov L.I.., Ershova E.S., Chernikov V.G, Mirochnik V.V., Bubnova E.N., Kucharenko V.I. // Bull. eksp. biol. med. (Bulletin of Experimental Biology and Medicine). 1996. V. 122. № 9. P.314-317.
20. Kovalyov L.I., Shishkin S. S., Efimochkin A.S., Kovalyova M.A., Ershova E.S., Egorov T.A., Musalyamov A.K. // Electrophoresis. 1995. V. 16. P. 1160–1169.
21. Eremina L.S., Kovalyov L.I., Shishkin S.S., Toropygin I.Y., Burakova M.I., Kovalyova M.A., Makarov A.A., Dzeranov N.K., Kazachenko A.V., Totrov K.I., Kononkov I.V., Loran O.B. // Vopr. biol. med. farm. khimii (Problems of Biological, Medical, and Pharmaceutical Chemistry). 2007. V.3. P.49-52.
22. Kovalyova M.A., Kovalyov L.I., Eremina L.S., Makarov A.A., Burakova M.I., Toropygin I.Y., Serebryakova M.V., Shishkin S.S., Archakov A.I. // Biomedicinskaya khimia (Biomedical Chemestry). 2008. V. 54. № 4. P. 420-434.
23. Anderson N.G., Anderson L. // Electrophoresis. 1996. V. 17. P. 443–453.
24. Shishkin S.S., Kovalyov L.I., Gromov P.S. / Mnogolikost sovremennoy genetiki cheloveka (The Variety of Contemporary Human Genetics). M.-Ufa: «Gilem», 2000. P. 17-50.
25. Cumings D. // Clin. Chem. 1982. V. 28. P. 782–789.
26. Loran O.B., Veliev E.I., Okhrizts V.E., Lisitskaya K.V., Eremina L.S., Kovalyov L.I., Kovalyova M.A., Shishkin S.S. // Eur. Urol. Suppl. 2010. V. 9. № 2. P. 309.

A Novel Approach to the Development of Anticarcinogenic Vaccines

A. N. Glushkov[1], S. V. Apalko[1]*, M. L. Filipenko[1,2], V. A. Matveeva[1,2], A. Yu. Bakulina[1], V. G. Lunin[3], M. V. Kostyanko[1,4]

[1] Institute of Human Ecology, Siberian Branch, Russian Academy of Sciences

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

[3] Gamaleya Research Institute of Epidemiology and Microbiology

[4] Kemerovo State University

* E-mail: apalko@ngs.ru

Received 07. 10. 2011

ABSTRACT **Human exposure to chemical carcinogens is an important etiological factor in cancer diseases. In this article, we will discuss a new approach to the development of anticarcinogenic vaccines. The main task in our research was to select a benzo[a]pyrene immunomimetic peptide considered as a hapten-specific component. For this purpose, we synthesized carcinogen-protein conjugates and prepared mono- and polyclonal antibodies to benzo[a]pyrene. Phage display technology was used to select the benzo[a]pyrene immunomimetic peptide, followed by an evaluation of the immunological properties of the obtained peptide. The obtained benzo[a]pyrene immunomimetic peptide could only simulate chemical carcinogens in the frame of the pIII protein. As a result, we prepared a recombinant protein composed of the benzo[a]pyrene immunomimetic peptide and pIII-encoding sequences. Using ELISA, we demonstrated that the recombinant protein specifically interacts with the anti-benzo[a]pyrene monoclonal antibody (mAB B2). Using molecular modeling, we predicted the 3-D structure of the mAB B2 active center and analyzed the characteristics of its interaction with different polycyclic aromatic hydrocarbons, as well as with the benzo[a]pyrene immunomimetic peptide. Thus, a comprehensive analysis of the results of the obtainment of hapten-specific components of anticarcinogenic vaccines allowed us to outline a strategy for future development in this direction.**

KEY WORDS **benzo[a]pyrene, anticarcinogenic vaccine, immunomimetic peptide, phage display, molecular modeling**

ABBREVIATIONS **CBD – cellulose-binding domain; OD – optical density; BP – benzo[a]pyrene; BSA – bovine serum albumin; AB – antibody; mAB – monoclonal antibody; ELISA – enzyme-linked immunosorbent assay; PAH – polycyclic aromatic hydrocarbon**

INTRODUCTION

The UN Health Agency has reported that more than 8 million people die from cancer every year. This reinforces the need to develop a novel therapeutic strategy based on antitumor vaccines. Unfortunately, such vaccines commonly target the existing disease rather than its cause.

Data from the World Health Organization (WHO) indicate that 90% of cancer cases are a result of the action of environmental carcinogenic agents. The bulk (70–80%) of such agents is made up of chemicals, including widely circulating polycyclic aromatic hydrocarbons (PAHs). It is an easy guess to assume that the identification of carcinogenic substances and their elimination from the sphere of human activity could serve as effective cancer prophylactic. However, such an approach is virtually impossible to pursue because of many factors. Therefore, the creation of antitumor defense by anticarcinogenic vaccines buttressing the immunological barrier in animals (including humans) against carcinogenic chemicals, seems necessary.

Chemical carcinogens are low-molecular substances that cannot, in themselves, induce an immune response. In 1937, Creech and Franks first synthesized conjugates of carcinogens with high-molecular carriers - blood serum proteins. They found that immunization of these conjugates leads to synthesis of specific anticarcinogenic antibodies (ABs). At the same time, some inhibition of carcinogen-induced tumor progression was noted following pre-immunization, and the idea that the approach could be applied to prevent tumor development in humans was first put forward [1].

In 1981, Moolten and associates took the next step in the development of anticarcinogenic vaccines. They prepared protein conjugates with a structural analog of carcinogen, which in itself cannot induce a tumor. The pre-immunization of animals with this conjugate

essentially decreased the probability of tumorigenesis caused by an actual carcinogen [2].

A novel approach was used by Chagnaud and associates. In 1992, they reported the preparation of an anti-idiotypic monoclonal antibody (mAB) against benzo[a] pyrene (BP). The second anti-idiotypic mAB carries the internal image of a carcinogen and can induce the synthesis of the first ABs against the carcinogen without the use of carcinogen-protein conjugates. The inhibitory effect of the second mABs on the development of chemically induced tumors was described in 1993 [3].

In the mid-1980s and afterward, Silbart and associates focused their efforts on the induction of specific secretory ABs in the gastrointestinal and respiratory mucosa by combining carcinogen-protein conjugates with various adjuvants to create a barrier preventing carcinogen transport between the environment and the body. In a review published in 1997, Silbart directly raised the issue of the future use of anticarcinogenic vaccines in humans [4].

Since conjugates of carcinogens or their analogs with carrier proteins can lead to iatrogenic induction of tumors, and the introduction of anti-idiotypic ABs – to allergic and autoimmune diseases, the proposed approaches are inapplicable in the development of anticarcinogenic vaccines for animals, including humans.

We offer a fundamentally novel approach in anticarcinogenic vaccine development implying the use of a peptide as a hapten-specific component that can induce specific anticarcinogenic ABs. Since BP is one of the most active and widely distributed PAHs and an absolute carcinogen for humans, we set out to prepare a peptide immunomimetic of BP.

MATERIALS AND METHODS

Synthesis of conjugates

PAH-protein conjugates of BP, benz[a]anthracene, anthracene, chrysene, and pyrene (Aldrich, Germany) were synthesized by means of covalent binding of haptene aldehyde groups with the amino groups of the carrier protein (bovine serum albumin [BSA] or hexokinase) [5].

Peptide-cBSA conjugates: 2 mg of the synthetic peptide and 10 mg of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) were added to 700 μL of solution containing 2 mg of cationized BSA (cBSA) [6], incubated for 2 hours, and dialyzed six times against 1 L of H_2O.

Immunization of laboratory animals

Preparation of hybridomas producing mABs specific to BP: hybridomas were prepared by the fusion of murine Sp2/0 myeloma cells and female Balb/c mouse splenocytes following immunization with the BP-BSA conjugate [7], as described by Kohler G. and Milstein C. [8].

Preparation of polyclonal ABs against BP: rabbits were immunized with 2 mg of the BP-BSA conjugate by weekly intramuscular injections for three weeks. The first injection was performed with the mixture of the antigen with a complete Freund's adjuvant (CFA) (Sigma, USA); the second – with the mixture of the antigen with a incomplete Freund's adjuvant; and the third – with the antigen in PBS. Then, supporting injections were introduced once in two weeks. Blood was taken two months after the beginning of immunization – every other week.

Immunization of animals with a chimeric protein containing a BP immunomimetic peptide: Balb/c mice were immunized by intraperitoneal injections of a chimeric protein four times once every two weeks. For the first injection, the antigen was mixed with CFA. Other injections were performed using the antigen mixture with an incomplete Freund's adjuvant. The amount of the antigen was 100–150 mg. The blood serum was tested for specific ABs against PAH, beginning from the first injection.

Purification of anti-BP ABs was carried out by affinity chromatography on columns filled with PAH-hexokinase-Sepharose 4B [9]. The replacement of the carrier protein enabled to avoid a preliminary purification of the antiserum from admixing ABs against the protein carrier used for immunization (anti-BSA ABs).

ELISA

Identification of specific anti-PAH-BSA ABs by ELISA: PAH-BSA conjugate (5 μg/mL, 100 μL in each well) was sorbed into wells of a polystyrene 96-well plates (Medpolymer, Russia) for 12 hours at 4°C. Nonspecific binding sites were blocked with 0.5% BSA in PBS, pH 7.2–7.4, containing 8 g of NaCl, 0.2 g of KCl, 2.68 g of $Na_2HPO_4 \times 7H_2O$, 0.24 g of KH_2PO in 1 L of water, then100 μL of blood serum samples serial dilutions in PBS containing 0.05% Tween-20 (PBST) and 0.5% BSA were added into the wells. Unbound material was removed by washing with PBST and PBS, and bound ABs were detected by treatment with an anti-mouse IgG horseradish peroxidase (Biosan, Russia) conjugate, followed by staining with TMB (Fluka, Switzerland). The optical density was determined on a microplate reader (FFM, Russia) at 450 nm.

To detect a specific binding of the chimeric protein with anti-BP ABs, The mono- or polyclonal AB to BP (5 μg/mL) was sorbed into wells of a polystyrene 96-well plate. Following blocking, serial dilutions of the chimeric protein in PBST containing 0.5% BSA (100 μL per a well) were added and incubated. The plates were then thoroughly washed, and 100 μL of rabbit serum

against a cellulose-binding domain (CBD) was added into each well followed by incubation, while the bound recombinant proteins were detected with an anti-rabbit IgG horseradish peroxidase conjugate as described above. The experiment was triplicated for estimation of reproducibility.

Competitive ELISA: conjugate BP-BSA (5 μg/mL) was sorbed into wells of a polystyrene 96-well plate. Following blocking, a mixture of mAB B2 of equal concentration with varied amounts of a competitor (PAH or synthetic peptide) was added into the wells. The mAB B2-competitor mixtures (total volume 100 μL, each) were preincubated for 30 min at 37°C and shaken gently. All tested samples were diluted with PBST containing BSA. The plates were incubated for 1 hour at 37°C and shaken gently. Then, following thorough washing with PBST, the bound mABs were detected with an anti-mouse IgG horseradish peroxidase conjugate as described above. The experiment was triplicated for estimation of reproducibility.

Affinity selection of phage display peptide library was carried out according to the recommended protocol to the Ph.D-12™ kit (New England BioLabs), with additional modifications [10].

Chemical synthesis of peptides by the method of activated esters in solution was performed in the Laboratory of Organic Synthesis, Institute of Chemical Biology and Fundamental Medicine, SB, RAS.

Molecular modeling.

Optimum patterns for the AB structure modeling by homology were matched using the BLAST server. Modeling was performed using the Modeller9v1 software. Molecular docking was performed using AutoDock version 4.0. Construction of the model for a peptide comprising the pIII protein was performed using the Rosetta program for *de novo* modeling [11].

RESULTS AND DISCUSSION

Synthesis of PAH-protein conjugates for preparation and analysis of antibodies

A major shortcoming of known methods of PAH-protein conjugate synthesis including the preparation of BP conjugates with proteins is the formation of a polymer, which considerably decreases the yield of the soluble fraction and, as a result, makes this conjugate unsuitable for immunoassay. So, the main task in this set of studies was to prepare well-soluble hapten-protein conjugates that contain a minimum of the polymeric admixture and are stable without the need for specific stabilizers.

We applied a method of covalent hapten-protein binding via the formation of an azomethine bond between an aldehyde group of hapten and amine groups of protein [5]. The use of BP aldehyde for the synthesis of conjugates with proteins enabled good results: animal blood sera with high titers of anti-BP ABs were obtained. Immunization with a hapten bound with one carrier protein (for example BSA) followed by the detection of ABs against this hapten based on another carrier protein (hexokinase) was found to be highly effective in both the analysis of anti-BP ABs in direct and competitive ELISA and the one-step preparation of affinity-purified ABs against PAH [9].

Preparation and immunochemical characterization of a monoclonal antibody against benzo[a]pyrene

There are several anti-BP mABs available in the world currently (USA, Czech Republic, Japan). They were raised mainly for the development of ELIZA-based test-systems to detect PAH pollutants in the environment, as well as their metabolites and DNA-adducts in the biological fluids of animals, including humans [12–14]. The main shortcoming of these mABs – in the case of hapten-specific vaccine component preparation – is the insufficient specificity of their binding with BP, compared with noncarcinogenic PAH. Besides, the capability of the abovementioned mABs to bind with hydrophobic endobiotics (steroid hormones) and aromatic aminoacids hasn't been studied yet . So, the key stage in our work was the preparation of a highly specific anti-BP mAB and the analysis of its cross-reactions with other PAHs, steroid hormones, and aromatic aminoacids.

Among the obtained murine hybridoma clones, the clone B2 was chosen producing IgG mAB, which had no affinity to the anthracene-BSA conjugate and had low affinity to chrysene-BSA and pyrene-BSA conjugates. The mAB B2 most effectively binds BP and benz[a]anthracene, a putative human carcinogen [7].

We checked for the possibility of cross-reaction of the mAB B2 with aminoacids, such as tryptophan and phenylalanine, based on the belief that the presence of an aromatic ring is one of the requirements of the interaction of anti-PAH AB with other substances. It is known that one aryl hydrocarbon receptor is implicated in signal transduction from PAH and endogenous substrates (in particular, estrogens); so, we also studied the cross-reaction of the mAB B2 with these substances and found no binding. This excludes the probability of preparing an anticarcinogenic vaccine with a side effect such as inducing autoimmune reactions against endogenous ligands.

Preparation and characterization of a benzo[a]pyrene immunomimetic peptide

We applied the phage display technique in the search

for a BP immunomimetic peptide. The procedure of affinity selection included incubation of the Ph.D-12™ initial library with mono- and polyclonal ABs to BP and washing from AB-unbound and elution of bound bacteriophages. The preliminary procedure of bacteriophage exhaustion with intact murine or rabbit IgG was substituted with cross-mapping of ABs in the third round of selection; i.e., the first two rounds were carried out on single species ABs, for instance mAb B2, and polyclonal ABs against BP were mapped with the obtained bacteriophage population in the third round, and vice versa. The proposed approach should favor the selection of high-affinity bacteriophage clones.

Five resulting bacteriophage clones were produced that specifically interact with mAB B2. Four clones resulted from cross-selection, when the two first rounds were carried out on mAB B2, and the last round – on polyclonal ABs to BP. One clone was produced when all selection rounds were carried out on mAB B2. DNA sequencing of these clones, followed by translation, demonstrated that all five clones had an identical aminoacid sequence of the recombinant peptide: LeuHisLeuProHisHisAspGlyValGlyTrpGly [10, 15].

The BP immunomimetic peptide (named PiP) was synthesized for further study of its immunochemical properties. Since two halves of the PiP, LeuHisLeuProHisHis (LH-peptide) and AspGlyValGlyTrpGly (DG-peptide), were synthesized first and then linked, we also evaluated LH- and DG-peptides for specific interaction with mAB B2.

Since the PAH structural mimicry is supposed to depend on the presence of a tryptophan residue within the peptide sequence, tryptophan was used as a negative control.

Synthetic peptides were found to compete with the BP-BSA conjugate for the binding with mAB B2. However, their binding force is substantially weaker than that of BP. Tryptophan (Trp) did not demonstrate any significant competition for binding with mAB B2 (Fig. 1). This fact suggests that Trp can specifically bind with ABs against the PAH group of chemical carcinogens, only when in context with other amino acid residues of these peptides.

The nature of LH-peptide binding with mAB B2 remains an enigma. One can speculate that a very complex interaction takes places between the PiP peptide and mAB B2, which cannot be explained by the fact that Trp or other hydrophobic residues structurally mimic BP.

Analysis of blood sera from mice immunized with peptide–cBSA conjugates revealed the presence of ABs to benz[a]anthracene and anthracene. However, their level is an order of magnitude lower than that of anti-PAH ABs induced by the immunization of mice with BP-BSA [11].

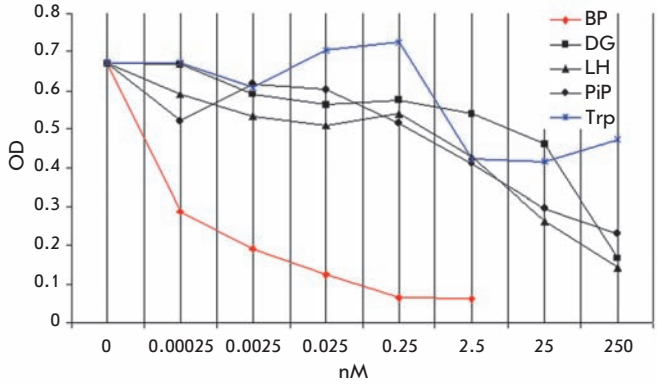


Fig. 1. Competitive inhibition of mAb B2 binding with immobilized conjugate BP-BSA by BP, DG, LH, PiP, and Trp.

Several reports on the production of the peptide mimetic of low-molecular compounds have shown that the initial conformation of peptides present on the surface of a bacteriophage carrier can change when the peptide is disengaged, or undergo additional modification. These alterations are crucial for ABs to recognize a peptide [16].

Thus, using the Rosetta software, we constructed a model for the peptide portion within the bacteriophage M13 pIII protein. The model assumes that the Trp side radical is localized on the surface of the protein [11]. It is likely that the structure of a peptide immunomimetic enabling mimicry of PAH-type carcinogens is possible only within the context of the pIII protein. In this context, our efforts therefore focused on the production of a recombinant protein composed of the BP immunomimetic peptide and bacteriophage pIII protein.

Preparation and characterization of a recombinant protein containing the benzo[a]pyrene immunomimetic peptide

Several approaches in gene engineering are known to enable an increase in the expression level and stability of transgenic proteins in a bacterial system, the facilitation of the testing procedure, and to enhance efficiency in protein purification. One of these approaches is a fusion technique directed toward the synthesis of chimeric proteins. It is based on the linkage of two genes (a gene for the antigene component and that encoding a carrier protein) in one reading frame, which leads to the synthesis of a chimeric protein in the bacterial system [17].

Using this technology, we produced and characterized the chimeric protein whose antigen component comprises amino acid sequences of a BP immunomimetic peptide and the pIII protein of bacteriophage

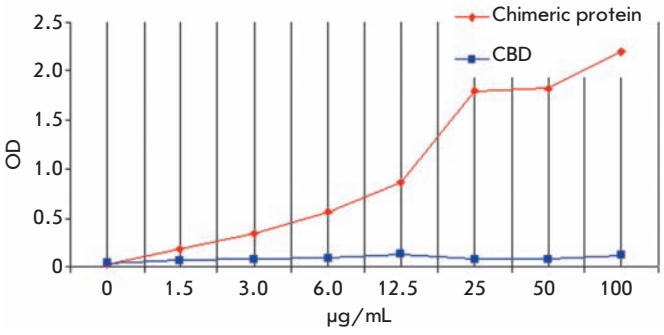


Fig. 2. Binding of chimeric protein containing benzo[a] pyrene immunomimetic peptide with mAB B2.

M13, on the basis of which the Ph.D-12™ library was constructed. A CBD-domain of *Anaerocellum thermophilum* endoglucanase was used as a carrier. It can interact with high affinity with a cellulose sorbent, thus enabling the isolation and purification of the recombinant protein containing the immunomimetic peptide.

Noncompetitive ELISA was used in the study of the interaction between mAB B2 and the produced chimeric protein containing the antigen component (BP immunomimetic peptide within the pIII protein) and the carrier protein (CBD). CBD was used as the negative control. The ability of the chimeric protein containing the immunomimetic peptide to bind with mAB B2 sorbed on a plastic was found to be dose-dependent (Fig. 2). The chimeric protein did not bind polyclonal murine and rabbit ABs to BP.

Balb/c mice were intraperitoneally immunized with the chimeric protein to test for the ability of the immunomimetic peptide within the bacteriophage pIII to induce ABs against PAH. A low level of anti-PAH ABs was detected in the blood serum of immunized mice. The most prominent binding was detected between the ABs and anthracene. At the same time, the blood serum of mice immunized with the bacteriophage recombinant clone containing the BP immunomimetic peptide within the pIII protein contains a mAB against BP, wherein the titers are comparable with those in the positive control – immunization with the BP-BSA conjugate [10, 15].

With the aim to find approaches to enhancing chimeric protein immunogenicity in relation to BP, we studied – using molecular modeling – the spatial structure of the mAB B2 active center and the features characterizing its interaction with PAHs and the immunomimetic peptide.

Peculiarities of the interaction between mAB B2 and benzo[a]pyrene immunomimetic peptide

A model for the mAB B2 Fab-fragment was created by the method of homology using the determined primary structures of heavy and light chains. The average binding energy between the mAB B2 Fab-fragment and several PAHs calculated using molecular docking software correlated with the experimental data on cross-reactivity between mAB B2 and these PAHs, thus confirming the validity of the created model [11].

Two pockets for PAH binding possibly exist in the active center of mAB B2, as was determined by molecular docking. The best position for BP and other PAH binding was determined to be between the third loops of the light and heavy chains of mAB B2 (this pocket was named P1). Two variants of BP docking in the P1 pocket were determined: the vertical and the horizontal, the first and the second one, respectively. The second pocket between the second loop of the light chain and the third loop of the heavy chain was less profound and less preferable for PAH binding, judging from the higher binding energy predicted by the docking (this pocket was conditionally named P2) [11].

A number of molecular docking calculations for the mAB B2 Fab-fragment with tripeptides comprising PiP have been carried out with the aim of modeling the interaction between the LeuHisLeuProHisHisAspGlyValGlyTrpGly peptide and mAB B2. None of the tripeptides binds with ABs in the region of the first pocket. Several tripeptides (HisLeuPro, LeuProHis, ProHisHis, and GlyTrpGly) have bound with Ab in the region of the second pocket (Fig. 3). Three tripeptides are convergent in the presence of a histidine residue, which is not shielded by other amino acid residues. This explains the LH-peptide's ability to compete with the BP-BSA conjugate for mAB B2.

At the same time, tryptophan, being within the BP immunomimetic peptide, obviously plays the key role in the binding of mAB B2, if the latter is exposed to the protein's surface. This immunomimetic peptide structure is possible within the pIII protein structure.

When the presence of the second binding pocket in the active center of mAB B2 is taken into account, one can explain the fact that the chimeric protein containing the immunomimetic peptide actively binds only to mAb B2, but not other polyclonal ABs against BP.

It is likely that in the process of recombinant bacteriophage selection on mAb B2, the peptides were selected by their binding with the second pocket as the most desired one. It is possible that the initial library contained no peptide capable of specifically binding with the deeper first pocket. The fact that no clones capable of specifically binding with the mAB B2 were found among the recombinant bacteriophages that re-

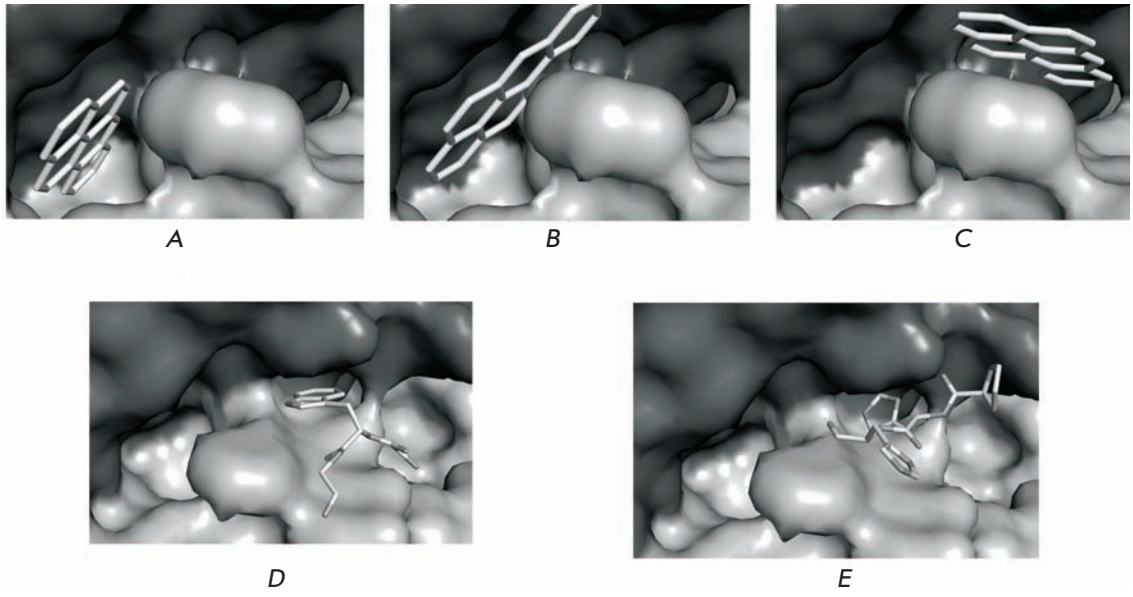


Fig. 3. The structure models of BP, GlyTrpGly, and ProHisHis complexes with the active center of the mAB B2 Fab-fragment. BP binding in positions P1P1 (*A*), P1P2 (*B*), and P2 (*C*); GlyTrpGly (*D*) and ProHisHis (*E*) binding in position P2.

sulted from the affinity selection on polyclonal ABs, as well as the fact that all five clones had an identical amino acid sequence, confirms the abovementioned hypothesis.

The experimentally revealed weak reverse-mimicry *in vivo*, i.e. weak immune reaction between AB and PAH, when mice were immunized with chimeric protein can be explained as follows.

As mentioned above, tryptophan has the closest structural similarity to BP among all side radicals. Superposition of the tryptophan and BP structures at the first position of the P1 pocket demonstrates that tryptophan, being within the polypeptide chain, cannot bind to such a deep cavity, because the length of BP exceeds that of the tryptophan side radical. At the same time, BP structural mimicry by tryptophan, in combination with some other side aminoacid radicals, is possible in the second position of the P1 pocket. It is likely that some part of AB induced by immunization with the immunomimetic peptide has a cavity for binding which is similar to the P2 pocket: therefore, tryptophan binding to AB does not generally bind to BP. Other parts of AB can possess a cavity for the binding of the P1 pocket's second position, enabling them to bind PAHs, including BP.

CONCLUSION

The use of conjugates of chemical carcinogens (or their structural analogs) with macromolecular carriers as vaccines is completely unacceptable in the immune prophylaxis of malignant tumors in humans because of the risk that the vaccine itself can induce a tumor. The hybridoma technique for the production of anti-idiotypic ABs has limitations in the optimization of the immunogenic properties of target vaccines. In addition, it is complex and expensive.

The proposed approach, i.e. the production of immune peptides that mimic chemical carcinogens using phage display technology, is preferable.

The mAB B2 obtained in our experiments possesses high specificity to BP, low cross-reactivity with non-carcinogenic PAHs, and does not react with endobiotics. Moreover, molecular docking showed that the predicted average energy for dibenzo[a]pyrene binding to the mAB B2 Fab-fragment is −8.91 kcal/mol, which is higher than the values for BP and other PAHs. We have established a direct correlation between the predicted binding energy and the experimentally measured PAH cross-reactivity. Based on this, one can surmise that the usage of mAB B2 in phage display technology could be effective in the search for immune peptides that mimic not only BP, but other PAHs with higher carcinogenic activity as well.

It is important to note that anti-idiotypic mABs are produced when animals are immunized with polyclonal ABs to BP [3]. In this context, the right strategy going forward will be to use new molecular targets in the search for PAH immunomimetic peptides. In our opinion, the use of the recombinant Fab-fragment of the mAB B2 with its second pocket removed via site-directed mutagenesis seems to be the most successful avenue. The recombinant bacteriophages resulting from the selection on such AB must be tested for binding with polyclonal ABs against BP.

The second method in enhancing the immunogenicity of target vaccines against carcinogenic PAHs would be the use of other phage libraries and/or optimization of the recombinant peptide structure via point muta-

tions. This approach could enable the production of a peptide inducing a highly specific immune response to BP and more carcinogenic PAHs. ●

The authors are indebted to Drs. E.A. Khrapov, L.E. Matveev, A.M. Lyashchuk, V.N. Sil'nikov, E.A. Sherina, and A.V. Averianov for their assistance in the work. The authors express especial acknowledgment to Academician D. G. Knorre for the support and assistance in the development of this direction.

The work is supported by the Government Contract (02.512.12.2044) in frames of Federal Target Programme "R&D in Priority Fields of the S&T Complex of Russia (2007-2012)" and the Russian Foundation for Basic Research (grant № 10-04-98003).

REFERENCES

1. Creech H.J., Oginsky E.L., Tryon M. // Cancer Res. 1947. V 7. P. 301-304.
2. Moolten F.L., Schreiber B., Rizzone A. // Cancer Res. 1981. V. 41. P. 452-459.
3. Chagnaud J.L., Faiderbe S., Geffard M. // Acad.Sci. Paris, Sciences de la vie. 1993. V. 316. P. 1266-1269.
4. Silbart L.K., Rusmussen M.V., Oliver A.R. // Vet. Hum. Toxical. 1997. V. 39 (1). P. 37-43.
5. Kostyanko M.V., Glushkov A.N. Patent № 2141114, RF, 6 G 01 N 33/5, 1998.
6. Muckerheide A., Apple Raimond J., Pesce Amadeo J., Michale Gabriel J. // The Journal of Immunology. 1987. V. 138. P. 833-837.
7. Glushkov A.N., Apalko S.V., Matveeva V.A., Kostyanko M.V., Cherno S.V. // Russian J. Immunology. 2009. V. 3(12). P. 30–38.
8. Kohler G., Milstein C. // Nature. 1975. V. 256. P. 495-497.
9. Glushkov A.N., Kostyanko M.V., Cherno S.V., Vasilchenko I.L. // Russian J. Immunology. 2002. V. 7. P. 41-46.
10. Glushkov A.N., Apal'ko S.V., Filipenko M.L., Matveeva V.A., Khrapov E.A., Kostyanko M.V. // Molecular Genetics, Microbiology and Virology. 2008. V. 23. P. 147–152.
11. Glushkov A. N., Apal'ko S. V., Bakulina A. Yu., Matveeva V. A., Khrapov E. A., Kostyanko M. V., Sil'nikov V. N., Filipenko M. L. // Molecular Biology. 2010. V. 44. P. 699–707.
12. Gomes M., Santella R.M. // Chem. Res. Toxicol. 1990. V. 3. P. 307-310.
13. Scharnweber T., Fisher M., Suchanek M., Knopp D., Niessner R. // Fresenius J Anal Chem. 2001. V. 371. P. 578-585.
14. Yamashita. N. Nishama S. United State Patent № US 6,277,964 B1, 2001.
15. Glushkov A.N., Apalko S.V., Filipenko M.L., Matveeva V.A., Khrapov E.A., Kostyanko M.V. Patent № 2357975, RF, C07K 16/42, A61K 38/04, C12P 21/03, 2009
16. Böttger V., Peters L.-E., Micheel B. // J. Mol. Recognit. 1999. V. 12. P. 191–197.
17. Terpe K. // Appl Microbiol Biotechnol. 2003. V. 60. P. 523-533.

GENERAL RULES

Actae Naturae publishes experimental articles and reviews, as well as articles on topical issues, short reviews, and reports on the subjects of basic and applied life sciences and biotechnology.

The journal is published by the Park Media publishing house in both Russian and English.

The journal *Acta Naturae* is on the list of the leading periodicals of the Higher Attestation Commission of the Russian Ministry of Education and Science

The editors of *Actae Naturae* ask of the authors that they follow certain guidelines listed below. Articles which fail to conform to these guidelines will be rejected without review. The editors will not consider articles whose results have already been published or are being considered by other publications.

The maximum length of a review, together with tables and references, cannot exceed 60,000 symbols (approximately 40 pages, A4 format, 1.5 spacing, Times New Roman font, size 12) and cannot contain more than 16 figures.

Experimental articles should not exceed 30,000 symbols (20 pages in A4 format, including tables and references). They should contain no more than ten figures. Lengthier articles can only be accepted with the preliminary consent of the editors.

A short report must include the study's rationale, experimental material, and conclusions. A short report should not exceed 12,000 symbols (8 pages in A4 format including no more than 12 references). It should contain no more than four figures.

The manuscript should be sent to the editors in electronic form: the text should be in Windows Microsoft Word 2003 format, and the figures should be in TIFF format with each image in a separate file. In a separate file there should be a translation in English of: the article's title, the names and initials of the authors, the full name of the scientific organization and its departmental affiliation, the abstract, the references, and figure captions.

MANUSCRIPT FORMATTING

The manuscript should be formatted in the following manner:

- Article title. Bold font. The title should not be too long or too short and must be informative. The title should not exceed 100 characters. It should reflect the major result, the essence, and uniqueness of the work, names and initials of the authors.
- The corresponding author, who will also be working with the proofs, should be marked with a footnote *.
- Full name of the scientific organization and its departmental affiliation. If there are two or more scientific organizations involved, they should be linked by digital superscripts with the authors' names. Abstract. The structure of the abstract should be very clear and must reflect the following: it should introduce the reader to the main issue and describe the experimental approach, the possibility of practical use, and the possibility of further research in the field. The average length of an abstract is 20 lines (1,500 characters).
- Keywords (3 – 6). These should include the field of research, methods, experimental subject, and the specifics of the work. List of abbreviations.
- INTRODUCTION
- EXPERIMENTAL PROCEDURES
- RESULTS AND DISCUSSION
- CONCLUSION

  The organizations that funded the work should be listed at the end of this section with grant numbers in parenthesis.
- REFERENCES

  The in-text references should be in brackets, such as [1].

RECOMMENDATIONS ON THE TYPING AND FORMATTING OF THE TEXT

- We recommend the use of Microsoft Word 2003 for Windows text editing software.
- The Times New Roman font should be used. Standard font size is 12.
- The space between the lines is 1.5.
- Using more than one whole space between words is not recommended.
- We do not accept articles with automatic referencing; automatic word hyphenation; or automatic prohibition of hyphenation, listing, automatic indentation, etc.
- We recommend that tables be created using Word software options (Table → Insert Table) or MS Excel. Tables that were created manually (using lots of spaces without boxes) cannot be accepted.
- Initials and last names should always be separated by a whole space; for example, A. A. Ivanov.
- Throughout the text, all dates should appear in the "day.month.year" format, for example 02.05.1991, 26.12.1874, etc.
- There should be no periods after the title of the article, the authors' names, headings and subheadings, figure captions, units (s – second, g – gram, min – minute, h – hour, d – day, deg – degree).
- Periods should be used after footnotes (including those in tables), table comments, abstracts, and abbreviations (mon. – months, y. – years, m. temp. – melting temperature); however, they should not be used in subscripted indexes (T_m – melting temperature; $T_{p.t.}$ – temperature of phase transition). One exception is mln – million, which should be used without a period.
- Decimal numbers should always contain a period and not a comma (0.25 and not 0,25).
- The hyphen ("-") is surrounded by two whole spaces, while the "minus," "interval," or "chemical bond" symbols do not require a space.
- The only symbol used for multiplication is "×"; the "×" symbol can only be used if it has a number to its right. The "·" symbol is used for denoting complex compounds in chemical formulas and also noncovalent complexes (such as DNA·RNA, etc.).
- Formulas must use the letter of the Latin and Greek alphabets.

- Latin genera and species' names should be in italics, while the taxa of higher orders should be in regular font.
- Gene names (except for yeast genes) should be italicized, while names of proteins should be in regular font.
- Names of nucleotides (A, T, G, C, U), amino acids (Arg, Ile, Val, etc.), and phosphonucleotides (ATP, AMP, etc.) should be written with Latin letters in regular font.
- Numeration of bases in nucleic acids and amino acid residues should not be hyphenated (T34, Ala89).
- When choosing units of measurement, SI units are to be used.
- Molecular mass should be in Daltons (Da, KDa, MDa).
- The number of nucleotide pairs should be abbreviated (bp, kbp).
- The number of amino acids should be abbreviated to aa.
- Biochemical terms, such as the names of enzymes, should conform to IUPAC standards.
- The number of term and name abbreviations in the text should be kept to a minimum.
- Repeating the same data in the text, tables, and graphs is not allowed.

GUIDENESS FOR ILLUSTRATIONS

- Figures should be supplied in separate files. Only TIFF is accepted.
- Figures should have a resolution of no less than 300 dpi for color and half-tone images and no less than 500 dpi.
- Files should not have any additional layers.

REVIEW AND PREPARATION OF THE MANUSCRIPT FOR PRINT AND PUBLICATION

Articles are published on a first-come, first-served basis. The publication order is established by the date of acceptance of the article. The members of the editorial board have the right to recommend the expedited publishing of articles which are deemed to be a priority and have received good reviews.

Articles which have been received by the editorial board are assessed by the board members and then sent for external review, if needed. The choice of reviewers is up to the editorial board. The manuscript is sent on to reviewers who are experts in this field of research, and the editorial board makes its decisions based on the reviews of these experts. The article may be accepted as is, sent back for improvements, or rejected.

The editorial board can decide to reject an article if it does not conform to the guidelines set above.

A manuscript which has been sent back to the authors for improvements requested by the editors and/or reviewers is reviewed again, after which the editorial board makes another decision on whether the article can be accepted for publication. The published article has the submission and publication acceptance dates set at the beginning.

The return of an article to the authors for improvement does not mean that the article has been accepted for publication. After the revised text has been received, a decision is made by the editorial board. The author must return the improved text, together with the original text and responses to all comments. The date of acceptance is the day on which the final version of the article was received by the publisher.

A revised manuscript must be sent back to the publisher a week after the authors have received the comments; if not, the article is considered a resubmission.

E-mail is used at all the stages of communication between the author, editors, publishers, and reviewers, so it is of vital importance that the authors monitor the address that they list in the article and inform the publisher of any changes in due time.

After the layout for the relevant issue of the journal is ready, the publisher sends out PDF files to the authors for a final review.

Changes other than simple corrections in the text, figures, or tables are not allowed at the final review stage. If this is necessary, the issue is resolved by the editorial board.

FORMAT OF REFERENCES

The journal uses a numeric reference system, which means that references are denoted as numbers in the text (in brackets) which refer to the number in the reference list.

For books: the last name and initials of the author, full title of the book, location of publisher, publisher, year in which the work was published, and the volume or issue and the number of pages in the book.

For periodicals: the last name and initials of the author, title of the journal, year in which the work was published, volume, issue, first and last page of the article.

Bressanelli S., Tomei L., Roussel A., et al // Proc. Natl. Acad. Sci. USA. 1999. V. 96. P.13034–13039 (If there are more than five authors). If there are less than five authors, all the authors must be listed.

References to books which have Russian translations should be accompanied with references to the original material listing the required data.

References to doctoral thesis abstracts must include the last name and initials of the author, the title of the thesis, the location in which the work was performed, and the year of completion.

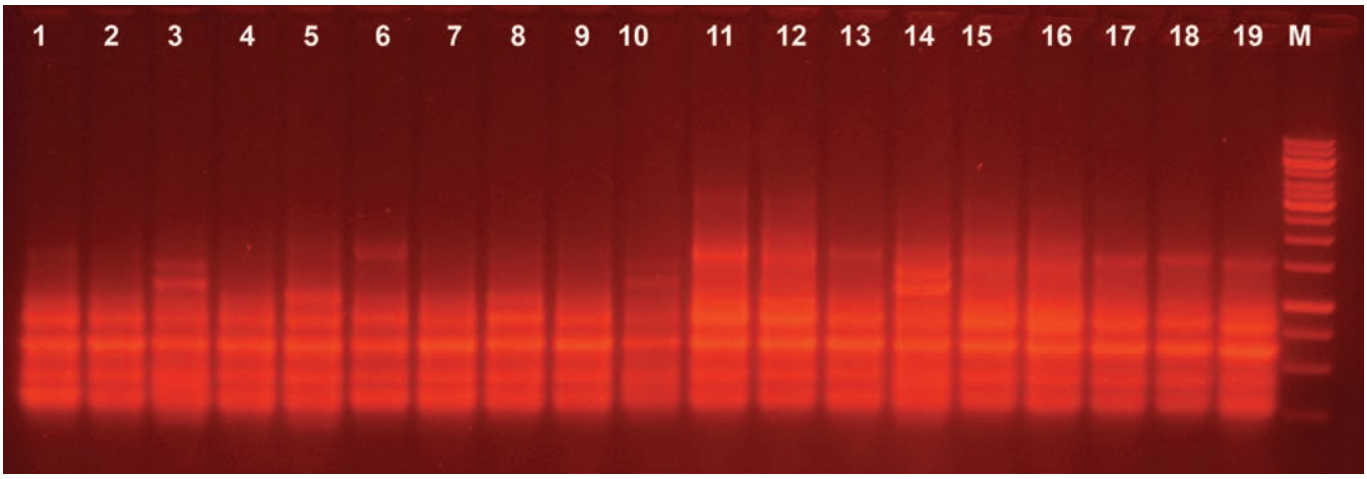
References to patents must include the last names and initials of the authors, the type of the patent document (the author's rights or patent), the patent number, the name of the country that issued the document, the international invention classification index, and the year of patent issue.

The list of references should be on a separate page. The tables should be on a separate page, and figure captions should also be on a separate page.

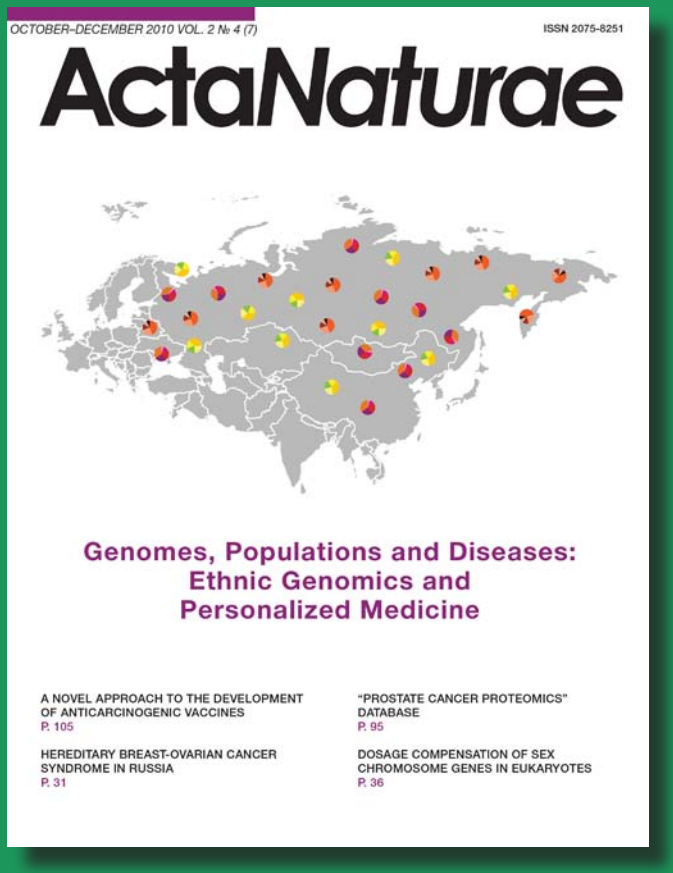
The following e-mail addresses can be used to contact the editorial staff: vera.knorre@gmail.com, actanaturae@gmail.com, tel.: (495) 727-38-60, (495) 930-80-05

ERRATUM

In the article A Novel High-Resolving Method for Genomic PCR-fingerprinting of Enterobacteria by A.S. Isaeva, E.E. Kulikov, K.K. Tarasyan and A.V. Letarov published in the 2010, Vol. 2, № 1 (4) issue of *Acta Naturae* (p.163-172), the same figure was erroneously printed twice (see Fig. 3 and Fig. 4). The publisher would like to apologize. The correct Figure 3 is provided below.



Acta Naturae is a new international journal on life sciences based in Moscow, Russia. Our goal is to present scientific work and discovery in molecular biology, biochemistry, biomedical disciplines and biotechnology. *Acta Naturae* is also a periodical for those who are curious in various aspects of biotechnological business, intellectual property protection and social consequences of scientific progress.



For more information and subscription please contact us at info@biorf.ru

NANOTECHNOLOGIES *in Russia*

Peer-review scientific journal

Nanotechnologies in Russia (Rossiiskie Nanotekhnologii) focuses on self-organizing structures and nanoassemblages, nanostructures including nanotubes, functional nanomaterials, structural nanomaterials, devices and facilities on the basis of nanomaterials and nanotechnologies, metrology, standardization, and testing in nanotechnologies, nanophotonics, nanobiology.

→ Russian edition: http://nanoru.ru

→ English edition: http://www.springer.com/materials/nanotechnology/journal/12201

Issued with support from:

The Ministry of Education and Science of the Russian Federation