

Letter from the Editors

Dear friends,
We are delighted to present the sixth volume of *Acta Naturae*. As has become the custom, we are continuing to publish review articles on promising aspects of life sciences, as well as experimental papers and the Forum section. The problem of achieving a balance between innate and adaptive immunity in the body's fight against infection has been of concern to the scientific community since the mid-1990s. In the past few years, it has become evident that the innate immunity factor has been underestimated. Many new factors highlighting the significance of the work of Russian Nobel Prize laureate Ilya Metchnikov have come to light. In this respect, we would like to note the review by D.V. Scheblyakov, dedicated to the role of innate immunity and, more specifically, that of Toll-like receptors in tumor progression. We also deemed it necessary to continue the discussion around pluripotent stem cells (see the review by S.P. Medvedev, A.I. Shevchenko, S.M. Zakiyan). It is indeed possible that in the foreseeable future medicine will become more personalized. High-throughput genome sequencing, detection of the distinct genes responsible for the expression of certain disease-related proteins will provide for efficient treatment and individual-targeted drugs. It will become possible, soon, to forecast an individual's response to therapeutic influence, which is of crucial importance in the case of drugs that can have negative side effects on an individual. The editors are also paying attention to biochemistry and enzymology; suffice to note the major importance of penicillin-specific enzymes. Their role in biotechnological and pharmaceutical advances can hardly be overestimated (the review by V.I. Tishkov, S.S. Savin, A.S. Yasnaya).

Experimental papers in the current issue cover a breadth of materials in the field of physico-chemical biology, from the physiological role of peptide structures (N.I. Minkevich, V.M. Lipkin, I.A. Kostanyan) to the structural peculiarities of "nano-antibodies" (S.V. Tillib, T.I. Ivanov, L.A. Vasiliev). The editors also offer readers an article on the molecular physiology of the cell (A.V. Shalyguin *et al.*),

which describes the role of Homer-proteins in the regulation of store-operated calcium channels. Some articles are dedicated to innovative approaches in diagnostics and drug design (O.B. Bekker *et al.*, M.M. Ulyashova *et al.*). The editors embrace the idea of publishing material on serious advances in biopharmaceutics.

In the Forum section, readers will find an interview with the laureates of state prizes in science and technology for 2009, which was announced in the fifth issue of *Acta Naturae*. In addition, we also want to inform you on the targeted actions of the Russian government in bolstering science at universities, a step that has sparked a lively debate in the professional community. After selecting the leading universities – both research and federal which, apart from gaining the status were also awarded extra funds – the Ministry of Education and Science announced a new competition meant to encourage cooperation with business, beef up the innovative infrastructure, and attract the world's leading scientists. In the Forum section readers will find a review article, as well as comments by officials of the Ministry on the results of the latter competition (RF Government decree №218 and №219). Sergey Ivanets, director of the Department of International Integration (Ministry of Education and Science), explains how grant applications are evaluated. These grants are aimed at attracting leading scientists, in accordance with the Decree of the Government of the Russian Federation, entitled On Measures for Attracting Leading Scientists to Russian Higher educational Institutions."

The publication is working on expanding its readership. We invite postgraduates and young researchers to send us their original research for publication in our pages. We guarantee objective reviews by leading experts in the specific field and in the shortest possible time. We welcome all applications. Please send your comments and suggestions to our editorial board at: actanaturae@gmail.com.

**We wish you every success
in your scientific endeavors!
The Editorial Board**

Save 10% on Subscription for 2011

Details at www.actanaturae.ru

DISCUSSION

Modeling approaches can be 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).

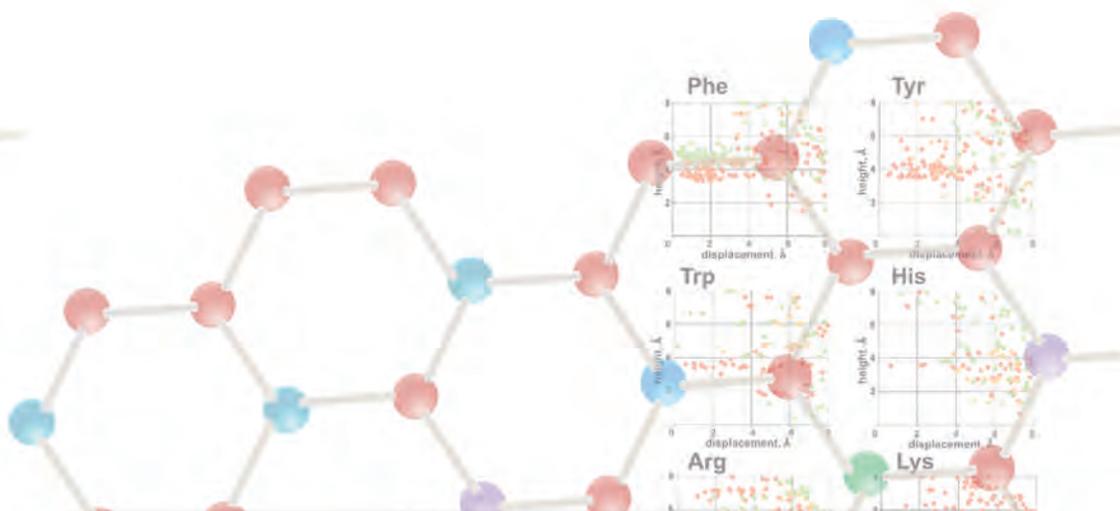
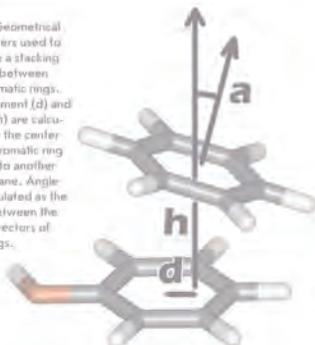
However, the range of these parameters, which corresponds to the presence or absence of a stacking contact, is still not very well defined and usually takes an arbitrary [6, 7]. Defining it more accurately would assist in developing more efficient scoring functions and should improve the prediction quality 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 aromatic moieties as a substructure.

The 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 naphthalene) [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 aromatic 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 guanidino 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. 3, shown in red and green, respectively). The displacement d lies in the same range (1–3 Å) for both types of contacts. Meanwhile, they differ in the value of height h , which is ≈ 3 Å for parallel and ≈ 1 Å for perpendicular orientation. Similar distributions were obtained for Tyr, Trp, and His, though the data are sparser in those 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.



APRIL/JUNE 2009, No. 1

Acta Naturae



SYNTHETIC ANTIBODIES

FOR CLINICAL USE

REGULATING TELOMERASE IN ONCOGENESIS
P. 51

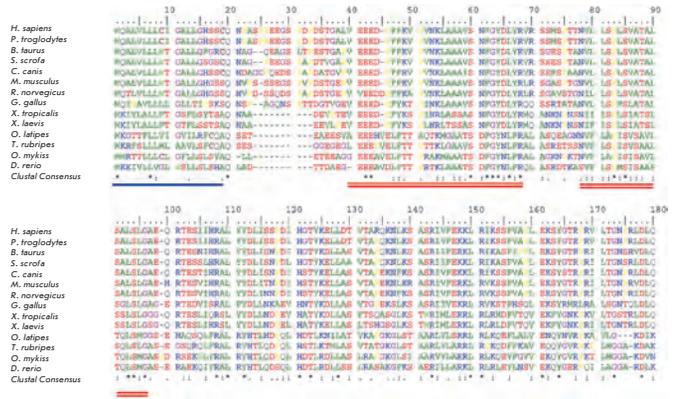
THE STRUCTURE OF THE MITOCHONDRIAL GENOME AS AN ACTIVATOR OF OPISTHORCHIASIS
P. 99

STACKING INTERACTIONS IN COMPLEXES OF FIBERS WITH ADENINE- AND GUANINE-CONTAINING LIGANDS

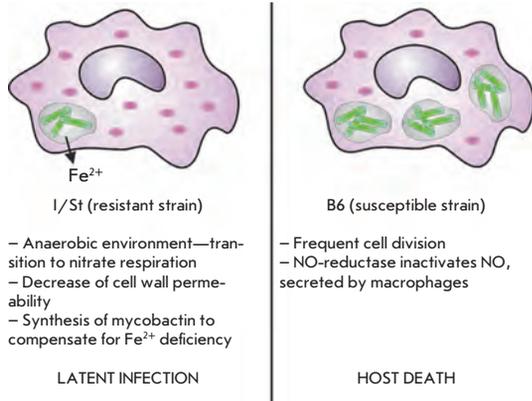
N.I. Minkevich, V.M. Lipkin, I.A.Kostanyan

PEDF – A Noninhibitory Serpin with Neurotrophic Activity

The pigment epithelium-derived factor (PEDF) is a glycoprotein with a molecular weight of 50 kDa belonging to the noninhibitory serpin family. It regulates several physiological processes. PEDF generates great interest as a promising drug for the therapy of a wide range of neurodegenerative, ophthalmological, and oncological diseases. This review is a summary of what is known today about the structural features, biochemical properties, and multimodal functions of PEDF.



Comparison of the amino acid sequences of PEDF proteins of different origins.



Mycobacterium avium metabolic state in the lungs of the I/St and B6 mice.

D. V. Ignatov, T. A. Skvortsov, K. B. Majorov, A. S. Apt, T. L. Azhikina

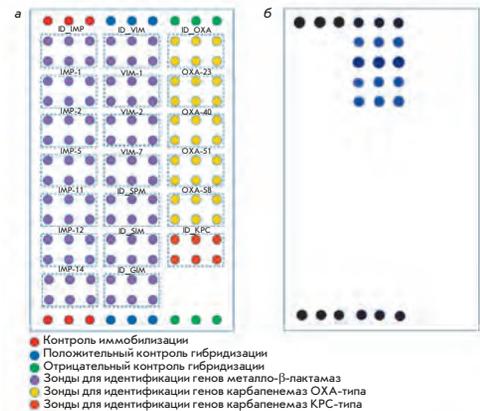
Adaptive Changes in *Mycobacterium avium* Gene Expression Profile Following Infection of Genetically Susceptible and Resistant Mice

We performed a comparative analysis of *Mycobacterium avium* transcriptomes (strain 724R) in infected mice of two different strains—resistant and susceptible to infection. Sets of mycobacterial genes transcribed in lung tissue were defined, and differentially transcribed genes were revealed.

M. M. Ulyashova, Yu. I. Khalilova, M. Yu. Rubtsova, M. V. Edelstein, I. A. Alexandrova, A. M. Egorov

Oligonucleotide Microarray for the Identification of Carbapenemase Genes of Molecular Classes A, B, and D

This work is a report on the development of a method of hybridization analysis on DNA microarrays for the simultaneous identification and typing of carbapenemase-encoding genes. These enzymes are produced by the microorganisms that are responsible for causing infectious diseases. The microarray method for the identification of carbapenemase genes is very accurate and highly productive. It can be employed in clinical microbiological laboratories for the identification and study of carbapenemase epidemiology.



Layout of specific and control oligonucleotide probes on the surface of the DNA microarray for the identification of the carbapenemase genes.

Founders

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

Editorial Council

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

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

Editorial Board

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

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

Project Head: E.A. Novoselova
Editor: N.U. Deyeva

Strategic Development Director: E.L. Pustovalova

Designer: K.K. Oparin

Photo Editor: I.A. Solovey

Art and Layout: K. Shnaider

Copy Chief: Daniel M. Medjo

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

Reprinting is by permission only.

© ACTA NATURAE, 2009

Номер подписан в печать 18 сентября 2010 г.

Тираж 300 экз. Цена свободная.

Отпечатано в типографии «МЕДИА-ГРАНД»

CONTENTS

Letter from the Editors 1

FORUM

Russian Federation State Prize in Science
and Technology for 2009 6

Alexander Potapov:
“Neurosurgery uses advancements
in modern biology in all of its aspects...” 10

Science in Universities: Ordered to Live 13

Universities and Business Cooperation
Contest Results 17

International Evaluation Procedure
for Supergrants 19

REVIEWS

D.V. Shchebliakov, D.Y. Logunov,
A.I. Tikhvatulin, M.M. Shmarov,
B.S. Naroditsky, A.L. Ginzburg
Toll-Like Receptors (TLRs):
The Role in Tumor Progression 21

S. P. Medvedev, A. I. Shevchenko, S. M. Zakian
Molecular Basis of Mammalian Embryonic
Stem Cell Pluripotency and Self-Renewal 30

V. I. Tishkov, S. S. Savin, A. S. Yasnaya
Protein Engineering of Penicillin Acylase 47

N.I. Minkevich, V.M. Lipkin, I.A.Kostyanan
**PDF – A Noninhibitory Serpin
 with Neurotrophic Activity62**

RESEARCH ARTICLES

I. V. Zvyagin, V. Yu. Dorodnykh,
 I. Z. Mamedov, D. B. Staroverov,
 A. G. Bochkova, D. V. Rebrikov, Y. B. Lebedev
**Association of ERAP1 Allelic Variants
 with Risk of Ankylosing Spondylitis72**

D. V. Ignatov, T. A. Skvortsov, K. B. Majorov,
 A. S. Apt, T. L. Azhikina
**Adaptive Changes in *Mycobacterium avium*
 Gene Expression Profile Following
 Infection of Genetically Susceptible
 and Resistant Mice78**

S. V. Tillib, T. I. Ivanova, L. A. Vasilev
**Fingerprint-like Analysis of “Nanoantibody”
 Selection by Phage Display Using
 Two Helper Phage Variants85**

A. V. Shalygin, M. A. Ryazantseva,
 L. N. Glushankova, I. B. Bezprozvanny,
 G. N. Mozhayeva, E. V. Kaznacheyeva
**Regulation of Store-Operated
 Channels by Scaffold Proteins
 in A431 Cells94**

M. M. Ulyashova, Yu. I. Khalilova,
 M. Yu. Rubtsova, M. V. Edelstein,
 I. A. Alexandrova, A. M. Egorov
**Oligonucleotide Microarray
 for the Identification
 of Carbapenemase Genes
 of Molecular Classes A, B, and D101**

O.B. Bekker, M.G. Alekseeva, D.I. Osolodkin,
 V.A. Palyulin, S.M. Elizarov, N.S. Zefirov,
 V.N. Danilenko
**New Test System for Serine/Threonine Protein
 Kinase Inhibitors Screening: *E. coli* APHVIII/
 Pk25 design.110**

N. N. Chekanov, E. S. Boulygina,
 A. V. Beletskiy, E. B. Prokhortchouk,
 K. G. Skryabin
**Individual Genome of the Russian Male:
 SNP Calling and a *de novo* Assembly
 of Unmapped Reads.122**

Guidelines for Authors.127

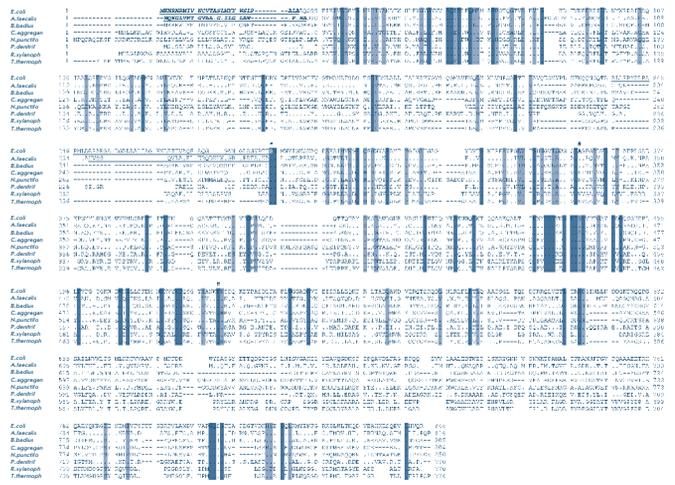


IMAGE ON THE COVER PAGE
 Alignment of amino acid sequences of penicillin acylases G
 from different sources.

Russian Federation State Prize in Science and Technology for 2009

The Russian Federation State Prize in the field of Science and Technology for 2009 for "A Set of Scientific Works on the Development of Laser and Information Technologies in Medicine" (Presidential Decree № 678, dated June 6th, 2010) was awarded jointly to Doctor of Physical and Mathematical Sciences, Academician of the Russian Academy of Sciences, Director of the Institute of Laser and Information Technologies, Russian Academy of Sciences, V. Ya. Panchenko; to Doctor of Medical Sciences, Academician of the Russian Academy of Medical Sciences, Deputy Director of the Burdenko Neurosurgery Research Institute, Russian Academy of Medical Science, A.A. Potapov; and to Doctor of Medical Sciences, Russian Academy of Medical Sciences, Academician of the Russian Academy of Sciences, Director of Herzen Cancer Research Institute V.I. Chissov.

In an interview with *Acta Naturae*, the winners talk about the developments that were honored as fundamental and of great importance to modern applied medicine.

State Prize Winner, Doctor of Physical and Mathematical Sciences, Academician, Director of the Institute of Laser and Information Technologies Vladislav Panchenko.

Vladislav Yakovlevich, please tell us about the work for which you received the State Prize?

– The Institute of Laser and Information Technologies is the first institute to build a system that helps remotely produce individual implants and biomodels based on the tomographic data of patients' pre-surgery examinations.

It was an interdisciplinary achievement which involved the participation of many of the Russian Academy of Sciences' institutes, as well as Lomonosov Moscow State University, and leading medical centers such as the Vladimirsky Moscow Regional Research Clinical Institute (MRRCI), the Burdenko Institute of Neurosurgery, the Herzen Moscow Can-

cer Research Institute (MCRI), the Central Research Institute of Dental and Maxillofacial Surgery of the Federal Agency of High-Tech Medical Care (CRIDMS), and the Blohin Russian Cancer Research Center (RCRC). At the Institute of Laser and Information Technologies of the Russian Academy of Sciences (ILIT RAS), the main work was done at laboratories run by V.V. Vasil'tsov, A.V. Evseev, and A.V. Ulyanov.

We began developing laser-information technologies of remote biomodeling more than 15 years ago. It is gratifying today to know that this technology has been included in the Health Ministry's approved list of high-tech procedures in oncology and neurosurgery.

The algorithm of this technology is as follows: a three-dimensional patient diagnostic is done (usually via tomography), then modern imaging is employed to diagnose both bone fragments and soft tissue with a resolution of about 1 millimeter. The obtained tomographic data is transmitted via the Internet to the center of rapid prototyping located in Shatura, in the greater Moscow area.

The tomographic data is then verified with a computer, and a working program is generated to reconstruct the model. This information is transmitted to a laser stereolithograph computer. A working chamber of a stereolithograph is filled with a complex composite polymer (developed in conjunction with the Institute of Chemical Physics and the Center of Photochemistry, Russian Academy of Science). Then, using programming, a laser beam with a preset wavelength, power, and frequency scans the surface of the liquid for any assigned program. Following that, a trace of the laser beam is hardened at a certain depth. All major equipment in the center (laser stereolithography machine, selective laser sintering of micro- and nano-powders) was designed and engineered at the Institute's laboratories.

How long did the selection of those conditions take?

Our institute conducted several experiments in photochemistry using different lasers. The results of these experiments led us to the synthesis of a new oligomer. Dozens of iterations were performed



Vladislav Panchenko

before we started to get hard, not easily breakable biomodels that exactly reproduce the data of the patient's tomographic examination.

For more than 5 years, several laboratories worked on developing the photo polymerization selection conditions. This is a very interesting, delicate, and time-consuming technology, which automatically forms biomodels layer after layer; for example, of a skull or bone fragment lost as a result of trauma.

Now ILIT RAS is investigating the possibility of multiphoton laser polymerization. Preliminary studies indicate the possibility of topologically complex structures

with a spatial resolution of up to 10 nm. Such a high resolution is required for the development of future technologies in neurological cancer surgery and cognitive studies.

Laser-Information biomodeling technology allows for very precise planning of future surgery. For example, its use in spinal surgery allows us to produce and "fit" the implant for biomodels without disturbing the patient. Summing up the experience, we can say that the use of rapid prototyping allows us to create individual biomodels based on tomographic data, and a script can be written for surgeries in almost every field of medicine.

As a result of this technology, the surgery itself is accelerated 2-3 times, while at the same time the rehabilitation time is reduced by the same rate. The number of surgeries planned and performed using this technology is approaching 4,000 now, and the technology is being used in more than 30 clinics throughout the country. The technology is mostly applied in neurosurgery, where almost everything is done under a microscope. Using this method in maxillofacial surgery has allowed to save dozens of children suffering from congenital diseases. This method has generated some interest on the part of many surgeons, including

oncological surgeons, and heart surgeons are also starting to show interest. But here I am intruding into someone else's territory. This is medicine. This would be better discussed by my colleagues, Academician A.A. Potapov and Academician V.I. Chissov, who are more knowledgeable on the subject.

Are the authors of this method Russian scientists?

– Yes. Based on currently available published data, our teams were among the first to create individual biomodels based on a patient's tomographic data, which have been found to be widely used in clinical practice. We were behind the concept of making individual "spare parts" for the human body in preliminary tomographic surveys, which are carried out anywhere in the world. In principle, you can scan the human skeleton, record the results, and, if necessary, refer to this information to reconstruct fragments, such as bones, in case of a fracture. Such surgery has already been successfully performed on animals.

Vladislav Yakovlevich, is government support necessary for the development of this technology?

Government support is always important, especially in large projects of international importance. The first problem is that such operations need to be standardized. When we speak of modern medical technology, we set specific actions that can be replicated by an accurate description in another location by another professional.

Development of a new technology is fundamental research and, to some extent, art. However, its application is routine. Russia often amazes the world by the unique surgeries performed here, but repeating it, making it serial, to help all those who are in need, is nearly impossible due to the lack of technology. It's 'high', but not 'tech', but if we're talking about high-

tech mass medical care, that is exactly what it should be - technical.

Did the successful establishment of biomodels inspire you and your staff to create implants from biocompatible materials?

– To some degree, yes. It is the next phase associated with tissue engineering. In conjunction with the Institute of Transplantation and Artificial Organs, we are looking for new biocompatible materials to use with laser stereolithography technology and for cleaning materials using supercritical fluids. This will create a complex configuration of implants and scaffolds – the vessels of a given shape for the directional growth of cells, which are used for tissue growth, eventually growing organs as well.

An intelligent laser surgical diagnostics system was included in the award. What is this system?

I will say a few words about one of these systems. As we know, laser is a good scalpel, and we thought - why not arm the surgeon with such a scalpel? That way it would be instantly clear what biological tissue is being cut. The fact is that during surgery various tissue particles can fly in various directions. Moreover, different particles fly with different distributions of velocity in space. A laser can adopt a scattered light and is much faster than human reaction; so it can help a surgeon identify the type of particles that evaporated at the moment. This information is transmitted to the surgeon in the form of certain signals. Thus, the surgeon knows exactly that he is removing a meningioma and not a healthy brain tissue.

So this system allows the surgeon to clearly define the boundaries of the unhealthy tissue, which is especially important in oncology. This idea has been discussed by Academician V.I. Chissov and corresponding member of the Academy, I.V. Reshetov. Intelligent laser

systems of this type were designed, built, and used in dozens of successful operations.

Corresponding member of Russian Academy of Medical Sciences, Medical doctor, Professor Igor Vladimirovich Reshetov:

– Research in the use of laser information technology began almost 10 years ago. In speaking to different audiences in academia, Vladislav Yakovlevich Panchenko awakened some interest in the medical applications of these developments. A research group was started focusing on the use of laser information technology in medicine. The group was headed by the following leaders: V.Ya. Panchenko from the Institute of Laser and Information Technologies, A.H. Konovalov and A.A. Potapov from the Research Institute of Neurosurgery, and V.I. Chissov from the Moscow Research Institute of Oncology.

Prototyping of biological objects was recently put into clinical application. Models are created at the stage of diagnosis clarification, evaluation of treatment, or follow-up corrective rehabilitation. The easiest prototypes to fabricate are those of body parts that have a support structure. So, there is no part of a human skeleton that was not considered. Naturally, we have sought not only to accumulate the observations, but also wanted to find new solutions that qualitatively improve the patient's life; for example, to produce an individual prosthetic bone, called an implant. The technology is already here: however, now we are working on making these implants not by making a cast of a symmetric part (i.e. through the mirror transport of the missing pieces), but in a way such that during the prototype model growth a biocompatible fragment is produced.

Why is the technology of rapid prototyping important to the surgeon and his patient?

- Thanks to this technology we now have an opportunity to predict the operation, optimally fit all “parts” and to assess the impact. Currently, the requirements for functional rehabilitation have increased considerably: people don’t want to feel impaired in any way, which requires technology to ensure very precise custom fitting. A surgeon has to be very aware of his actions. Procedures that previously had to be estimated are now performed accurately. This technology is not computer 3D animation, but a real, tangible thing.

As a general conclusion, laser technology in medicine is a reservoir of knowledge and technology that still requires lots of work. We are in the process of applying so-called intelligent laser devices that allow the surgeon to dissect tissue and to change the power mode, as well as the depth of dissection, that will contribute to the simplification of robotic surgery.

I have to note that the introduction of these technologies will require an increase in technical staff at hospitals, especially “medical physicists,” who are now in high demand. ●

**Interview by
Svetlana Sinyavskaya**



Igor Reshetov

Alexander Potapov: "Neurosurgery uses advancements in modern biology in all of its aspects..."

Alexander Potapov, laureate of the State Prize of the Russian Federation in Science and Technology for 2009, doctor of medical sciences, academician, Russian Academy of Medical Sciences, and deputy director of the N.N. Burdenko Scientific Research Institute of Neurosurgery of RAMS, talks to *Acta Naturae* about the achievements in modern neurosurgery, the latest developments in the field, and the close convergence between basic biology and clinical practice.

Alexander Alexandrovich, there has recently been such enormous progress in the field of neuroscience to the extent that some speak of a revolution in modern biology and medicine. What is the reason behind this trend?

The arsenal of the modern neurosurgeon has been supplemented by absolutely unique and constantly improving neuroimaging techniques. Above all, these include X-ray computed and magnetic resonance tomography. Forty years ago, the brain in the skull was in a kind of "black box," and the surgeons could see what was happening inside only after trepanation. Now we can see all the happenings inside the brain of a living human by using quick, non-invasive techniques that are absolutely safe for the patient. Thanks to new neuroimaging methods, we are able to not only study the anatomy of the brain and the pathological formations, but also see the blood-flow and obtain quantitative blood-flow data in milliliters per gram/minute for various structures and pathological formations in the brain. This allows to monitor the changes in the blood flow during the development of various pathological processes, as well as during their treatment. Methods allowing the mapping of metabolic changes in the brain are also gaining in popularity in clinical

practice; these methods include positron-emission tomography and, in recent years, magnetic-resonance spectroscopy (essentially molecular neuroimaging). Using MRI spectroscopy, we can see the changes in tissue metabolism, such as alterations in the level of N-acetylaspartate, which is a marker of neuron damage. These changes can be monitored not only in distinct areas, but also in whole sections of the brain at various levels. Neuroimaging methods allow doctors to distinguish cancerous processes from a disrupted blood-flow, as well as inflammatory or neurodegenerative processes. Modern neurovisualization methods are not only important for practical applications, but they are also a fundamental tool in research involving the study of the metabolic and structuro-functional changes during pathogenesis and neural tissue regeneration. These are fundamental issues in neurobiology.

Another possibility also recently developed and gaining wide application in practice is the ability to monitor a living brain in the process of various actions. This is called functional magnetic resonance tomography. For instance, we would ask a patient to do a task with his arm and then see what areas of the brain are activated, the cortical representation of the arm, and the location of any pathological formation, which can then be approached

so as not to damage the functional structures of the brain. The possibilities of modern Diffusion-tensor MRI allow us to see the pathways of the brain. For instance, we can see the corticospinal tract, which transfers signals from the pyramidal cells of the cortex to the motor neurons of the spine, via long axons.

In your practice, you use the developments in various basic sciences; not only biology, but also physics, for example. Your collaboration with Vladislav Panchenko from the Institute of Laser and Informational Problems and Valery Chissoy, the director of the P.A. Gertsin Moscow Oncological Scientific Research Institute, was awarded the Russian Federation State Prize in Science and Technology in 2009. Can you explain the essence of the new techniques that you developed and implemented?

Our collaboration with physicists proved very successful – our arsenal of tools was greatly improved through modern methods of rapid prototyping, which allow us to fully reproduce the exact copy of any human organ; body parts or the whole body. Currently, there are over a dozen prototyping methods. One of them allows the production of a full-scale copy of a virtual model from liquid polymers using a computer-controlled stereolithograph. There are other methods of rapid prototyping, such as powder



Alexander Potapov

metallurgy, which involves the use of metal powder, as well as the use of specialized laser technologies and a control computer to recreate complete copies of objects with very complex configurations.

Russian scientists have developed computer-controlled stereolithographs which provide the opportunity to recreate any object

that has been input and constructed in the computer. Suppose we perform a computer tomography of a human – any organ, brain or blood vessel. We obtain a three-dimensional image, construct the missing portion of the damaged organ in virtual space, and input the result into the controlling computer. As a result, the stereolithograph rapidly

builds a solid copy of the damaged object out of the liquid polymer.

How did all of this begin?

During the 1990's, when we were just starting this project, it wasn't clear if any of the technology would be useful. We read publications about the methods involved in the identification of the remains of the Russian Royal Family. Scientists from the Institute of Laser and Informational Problems used stereolithography to create precise replicas of the discovered skulls and their fragments. We believed that this method could be used for reconstructive neurosurgery and quickly joined the development process. Our first copy-models were a bit rough, but as we streamlined the techniques, we found that this technology had huge potential for clinical practice.

What are the advantages of the rapid prototyping technique?

Complex defects and deformations of bone structures require various implants. Several tons of metal (particularly titanium) are being implanted into humans every year. Reconstructive spine operations involve artificial discs and stabilizing systems, which are prepared in advance. Then, during the operation, the surgeon selects one of the several available sizes so as to accommodate the needs of the patient at hand. So, rapid prototyping makes it possible to produce implants that suit each patient individually according to his or her real anatomy.

Suppose that a patient needs a right pelvic joint replacement. We make a three dimensional image of the left (healthy) pelvic joint and then use the modern methods of rapid prototyping to manufacture a pelvic joint implant which retains all the dimensions, peculiarities, and other anatomical traits of the original. The same is true for neurosurgery – if a person damages part of his or her face, then the plastic surgeon could restore that part of the face by us-

ing the undamaged healthy region as a guide. We obtain a virtual 3D image of the skull, then take the undamaged portion, reflect it and place it into the damaged portion. In the end, we have a fully restored and completely symmetrical skull. If the facial damage is in the center of the face, we use a virtual 3D “donor” image and restore the missing central part. Over many years of using computer tomography, a huge database of images has been accumulated that can be used to virtually choose the appropriate configuration for the organ in question.

Rapid prototyping techniques are being developed in other countries, such as Japan, the U.S., and Germany. Yet your technology has already been successfully implemented in 50 medical centers. Why is our Russian technology competitive?

There are companies that manufacture similar custom-made implants. However, our technique is different in that it involves the surgeon in the implant manufacturing process. However, the surgeon’s main task is not only to restore the bone structure, but also to reconstruct the nervous tissue and individual neurons.

So the saying “nervous cells do not regenerate” is a misconception?

The thesis that a nervous cell cannot regenerate, postulated by the Nobel Prize laureate Santiago Ramón y Cajal, has been revised. Collaboration with the other biologists involved in the study of the problems of reconstructing the nerve cell structure will allow us to transfer the achievements in experimental neurobiology into clinical practice as soon as possible.

There are currently several approaches that can facilitate the regeneration of nervous cells. There are several known neurotrophic factors that can stimulate the cell’s regenerative processes. Lithium, one of the simplest elements, has been

shown to promote regeneration. There are also various compounds that inhibit the enzymes that block neural regeneration. Experiments have shown that neurons can be regenerated by neurotransplantation. However, for clinical use, we need more progress in the field of neurobiology. Then we will be able to use these discoveries in clinical practice.

So, you mean to say that medical practice and fundamental biology are integrated much more closely than we think?

Neurosurgery uses advancements in modern biology in all of its aspects. For instance, when we study the processes of brain damage caused by a stroke, trauma or cancerous damage, we are also studying the biology of these processes. How does this process progress in the neural tissue? Which mechanisms can lead to regeneration, and which can lead to cell death? Therefore, we need to know the biology of these processes. What action should we take to restore the damaged neural tissue, and all of its elements – neurons, the glial complex, and the vascular network? How do compounds in the blood reach their final target (brain cells) and overcome the unique hematoencephalic barrier? Today we can manage these processes in order to achieve a therapeutic effect.

Genetic research is also very important to us. Modern knowledge shows that the brain’s reaction to damage is determined not only by the type of damage *per se*, but also by the genetic traits of the individual, i.e. genetic profile. There is evidence that genetic polymorphism in a number of genes determines the strength of the response of neural tissue to damage. Any damage leads to a universal reaction – inflammation. In neural tissue, this is neuroinflammation, which is manifested by increased hematoencephalic barrier permeability, edema (intracellular, intercellular,

and mixed), disrupted circulation of liquor and blood, and other general metabolic disorders. The nature of these reactions and the following regenerative processes may be determined by the peculiarities of the human genome, or by the so-called genetic polymorphism. One of the issues of your journal published a review by Professor V.S. Baranov (*Editor’s note* – see *Acta Naturae*, v.1, №3, 2010) on genetic polymorphism and the prospects of medical genomics. This field is also very important for us, clinicians, since it means that our prognoses on the course of the pathological process and its outcomes should factor in not only the etiology, pathogenesis and clinical picture of the disease, but also the genetic profile of the patient, including his or her risk factors, which can lead to various scenarios of disease development.

And what about the reverse process – which methods of clinical medicine are used for studying the more basic problems?

The possibilities of modern neuroimaging techniques, which we discussed earlier, allow the study of more fundamental issues – the relationship between the structure and function of the brain. For instance, how do the brains of different people (lefties, righties, people with left or right-hemisphere dominance, or no dominance at all) function after the appearance of a pathological area? How does the brain adapt to aging, development, and pathological effects? These are all basic issues concerning the functioning of the nervous system (and the brain in particular) under normal and pathological conditions. Today these methods, which are becoming increasingly useful in clinical practice, are also successfully being used in basic neurobiological research – the study of the brain and the nervous system. ●

**Interview
by Elena Novoselova**

Science in Universities: Ordered to Live

M. Murav'eva, special for *Acta Naturae*

An unprecedented act of financial support, in the form of 90 billion rubles for leading Russian universities, has rocked the professional community. With stakes in the development of university science and research, the State has started to take action towards strengthening and amplifying science programs. After determining leading universities (including both research and federal centers) which, in addition to enjoying a high status also receive additional funds, the Ministry of Education and Science has unveiled a new competition for higher educational institutions. This competition focuses on cooperation with business, strengthening the infrastructure needed for innovation, and attracting leading scientists. As early as this fall, the winners are expected to begin receiving funding to the tune of millions of rubles, which should allow them to pursue the most daring and ambitious projects.

Information about the significant financial support for Russian universities first emerged late last year at the congress of the United Russia party when **Vladimir Putin** pledged to allocate an additional 90 billion rubles to universities for the next three years. Later, in January of the current year at the World Conference of Ministers of Education in London, the head of education and science in Russia **Andrei Fursenko** announced that Russia was embarking on modernization of its higher education, in order to bolster the scientific potential of its universities. He also expressed confidence that the best universities in the country would soon be able to compete not only with the Russian Academy of Sciences, but also with leading laboratories and institutions around the world.

To achieve these goals, the authorities are taking a number of measures. The Ministry of Education plans to spend some additional funds in the following areas:

- Support of development programs for the Moscow and St. Petersburg state universities;
- Support of the modernization ef-

fort of federal universities in the fields of scientific research and innovation;

- Support for national research universities;
- Development of cooperation between universities and industrial enterprises;
- Attracting leading researchers to Russian universities; and
- Developing innovative infrastructure at universities.

LOGICALLY AND CONSISTENTLY

The planned intention of the State to beef up science at universities has triggered a lively discussion in the scientific community. There was much opposition to the university-based approach to the development of research in Russia. As an argument, critics of the approach have usually referred to the particularity of the scientific organization in Russia and the traditionally strong position of the Academy. Many people cannot understand why the authorities have suddenly begun actively supporting university science. Yet in fact, there is nothing unexpected in the choice of such a policy. Since the mid 2000s, the Ministry of Edu-

cation has been grappling with the stranglehold the Russian Academy of Sciences has on scientific activity, trying to shift the center of gravity of research work towards universities. For many years, science has developed at universities for the most part on their personal initiative, without significant funding from the State. Now, strengthening university science has become a priority that is included in all key documents: in particular, in a document titled Development Strategy for Science and Innovation in Russia, which was passed in February of 2006 and will last until the year 2015. The document states the following: "Separating scientific and educational entities saps their potential for development, limits their contribution to the economy and society, and makes next to impossible full integration into the international scientific and educational community." A need to "integrate the education and research activities" was also mentioned back in 2004 in a presidential address to the Federal Assembly.

The strategy proposes a wide range of measures for the devel-

opment of university science, first and foremost the establishment of a network of leading research universities, “large, prominent scientific centers, which are equally effective in training specialists at all levels as well as carrying out research and development at an international level.” This approach, as well as other ideas, was used nearly five years later (2009-2010) as a basis for selecting research universities.

Among other measures, the strategy points to the need for developing and improving skills, including offering large grants to young scientists to conduct research, purchasing scientific equipment, and creating small startup firms, as well as furthering the mobility of personnel amongst research institutions, universities and various enterprises. Finally, another important stipulation is the need for a developed infrastructure. The State is currently pursuing all of these planned activities in a logical and consistent manner.

LEADERS EXIST

The reform of higher education, or “modernization” as it is usually called at the Ministry, began with an institutional reform, specifically, with the formation of a group of leading universities.

The Ministry of Education predicts that, over time, there will be no more than 50 universities left in the country, and in total, about 150-200 competitive institutions of higher education. This was mentioned for the first time by Andrei Fursenko two years ago at a meeting with the Russian president at the Moscow Engineering Physics Institute. Around the same time, work began on identifying and forming the country’s elite institutions of higher education - both federal and national research centers.

At the moment, there are two federal universities (the Southern and Siberian Universities);

there are also five emerging ones: the Volga, Urals, Northern (Arctic), Northeast and Far East. This project, one might say, is geopolitical. The strategic objective for federal universities, which are created by government mandate by combining several universities, is to provide quality education and an opportunity to engage in higher level science, so as to become the scientific and educational centers in their regions. This will allow the economy to develop, and allow qualified specialists to remain in the Far East, Siberia, the Urals, etc.

The mission of the national research universities (NIU) is, firstly, to fill the demand for workers in the high-tech sector, including the nuclear power industry, metallurgy, machine-building, shipbuilding, and others. Secondly, the aim is to develop applied science in order to commercialize research results.

Based on the results of the two contests that were held in late 2009 and early 2010, the status of an NIU was assigned to 29 universities. In addition to the prestigious status, the winners were promised sizable financial aid - 1.8 billion rubles each for five years. Over all, the NIU project will cost the state treasury nearly 50 billion rubles.

Another two schools that have historically been, de facto, the pillars of university education and science are Moscow State University and the St. Petersburg State University, who have also received special powers and very broad rights. By law, they have been given statuses of “the oldest universities in the country with great importance for the development of Russian society.”

But even for these “flagships,” state support will not be based simply on “giving” money. Moscow State University and St Petersburg State University will receive budget money for specific development programs, just like other research

and federal institutions of higher education.

GOVERNMENT STIMULATES

Selecting a group of leaders was the first serious step in the modernization of higher education in Russia. This step was then followed by additional measures to support research and innovation at universities. “Developing the university research component is one of the top priorities in the state’s policy,” Deputy Minister of Education and Science **Alexei Ponomarev** said in a speech: “It is a necessary element of the economy’s innovation infrastructure; a mechanism for generating new ideas and technologies, as well as their implementation and replication. Without the participation of universities in research, it is impossible to provide quality training in the areas of priority that are the scientific and technological modernization of Russia.”

In the spring of this year, the Russian government approved three resolutions, № 218, 219, and 220, to address the most pressing issues affecting domestic science - its aging and the deterioration of the quality of research. According to statistics, at Russian universities, only 15 percent of teachers engage in scientific research; therefore, students also remain remote from science, and without it, their education is essentially meaningless; it turns into daily grinding through outdated textbooks.

Competitions have been announced this summer for each decision. The winners can expect a significant financial infusion that will be aimed at strengthening cooperation between universities and businesses (12 billion rubles for three years), developing the innovation infrastructure (8 billion), and attracting leading scientists (12 billion).

Money is allocated from an anti-crisis fund, so that all of the three

resolutions are dedicated not only to research, but in general, to the modernization of higher education. Accordingly, the Ministry of Education prioritizes problems as follows:

- 1) Transfer of experience in implementing scientific research in universities to create a climate of healthy competition in Russia;
- 2) Mobilization of the scientific and management teams so that universities can achieve a qualitative leap in development; and
- 3) The actual scientific research.

ON COOPERATION BETWEEN UNIVERSITIES AND BUSINESS

The new tools proposed by the government are designed to make higher education central in applied science, by doing away with the skepticism business has towards the quality of their research. The State is trying to find ways to integrate higher education into the economy. This is true not only for the performance of individual economic agreements, as was the case before, but also for the implementation of strategic plans for the development of Russian companies. Decree № 218 contains a list of the 54 largest public and private companies that have to come up with programs of innovation development incorporating interaction with universities by November of this year. Deputy Minister Alexei Ponomarev has predicted that in the near future 20-30 companies out of 54 will begin cooperating with universities. This will be the basis for the establishment of so-called technology platforms - conglomerates of leading industrial firms, universities, and regional administrations. This tool, which has been borrowed from the European Union, is not directly supported by state resources. The point is to create consulting entities generated by key customers and engaging in research and development on certain issues. The entities in such platforms will be engaged

in planning and forecasting in their areas of specialization, market research, and elaboration of solutions that will shape government policy.

According to the Ministry of Education, it is extremely important to encourage both universities and commercial enterprises to invest in the development of the high-tech industry. The most significant tool given to the agency is the subsidies to industrial firms in order to work with institutions of higher education. Nineteen billion rubles are allocated for that purpose, six in 2010. Any firm, including foreign ones, that managed to have their research and development projects reviewed by experts from the institutions were invited to participate in the contest, which was announced in late June as stipulated in the 218th decree. The subsidy will be available for a period of 1 to 3 years and can amount to up to 100 million rubles per year. Companies are required to come up with matching funds of no less than the amount received from the State.

The contest proved very attractive for business: applications were lodged by 416 organizations from 59 Russian regions. One hundred and fifty-seven schools were selected as the projected sites for research and development in the project. The total amount sought by grant applicants exceeds 70 billion rubles. The average size of a subsidy is approximately 150 million rubles.

Fifty-seven projects won grants; 33 of the projects will be conducted in the regions that are not part of Moscow and St. Petersburg. Among the winners were such large manufacturing companies as the aircraft corporation MiG, truck builder KAMAZ, diamond mining giant ALROSA, and others. In total, 55 companies, in cooperation with 49 Russian universities, will implement the projects.

It is worth noting that the competition is conducted in two phases.

The decision was dictated by the desire to hit the targeted number of applications, since not all organizations had time to prepare the necessary paperwork. The results of the second phase will be announced before October 5.

Government Decrees № 218 and 219 are considered in the academic circle as the second stage of the push toward innovation at Russian universities. The first one was associated with small businesses – with 217-FZ becoming law last year, which allowed universities to establish small businesses. Now is the time for a stronger big business sector, which will be commissioning research and innovation from universities.

DEVELOPMENT OF THE INFRASTRUCTURE FOR INNOVATION

Based on government Decree 219, a selection of various university programs focusing on developing innovation infrastructure was carried out. Some 199 universities took part in the competition. The winner was determined by an analysis of the scientific, educational and innovative potential of the institution over the past 3 years and was presented based on these development programs. The Competition Commission selected the 56 best institutions. Most of them (14) are universities located in Moscow (Moscow State University, Bauman Moscow State Technical University, Moscow Institute of Physics and Technology, Moscow Engineering Physics Institute, Moscow Institute of Electronic Technology, and others). But overall, the geography of the winners is wide - 36 Russian regions in all 8 of the federal districts: Far East - 4; Volga - 7; North-West - 7; North Caucasus - 3; Siberian - 8; Ural - 4; Central - 20; and South - 3.

Each higher educational institution will receive up to 150 million rubles. The funds can be spent on strictly defined areas:

- 1) The development of the innovation infrastructure: business incubators, technology parks, certification centers, technology transfer, shared-use scientific and technical information, and others;
- 2) Evaluation of the results of intellectual activities and their legal protection;
- 3) Implementation and development of targeted training and skills development in small innovative enterprises, including graduate students and young scientists, as well as developing educational-methodological, and scientific-methodological support for small and medium-sized enterprises;
- 4) Training and professional development of university staff in the field of innovative entrepreneurship and technology transfer at foreign universities;
- 5) Consulting services of foreign and Russian experts in the field of technology transfer; creation and development of small innovative companies, including involving faculty members in the normative-methodical and practical support in the creation of such companies.

ATTRACTING TOP SCIENTISTS

Decree 220 is meant to attract the best specialists in the world to universities on a competitive basis to work at the newly established laboratories. To achieve this, the federal government will allocate 12 billion rubles. Given that the size of a single grant is 150 million, up to 80 scientists could win grants.

Some 507 applications have been received from 179 universities in the competition. Of these, most are in the following fields: Physics, Information Technology and Computing Systems, Earth Sciences and Materials Science (40 each), and only five are in the field

of Agricultural Sciences and Engineering.

Competition is intense amongst universities, especially since some of them have filed multiple applications. Lomonosov Moscow State University stands out with 30 applications. St. Petersburg State University and the Moscow Engineering Physics Institute have filed 21 applications, each. The vast majority of universities only submitted 3 to 6 applications.

The Ministry of Education has repeatedly emphasized that leading scientists will be determined not according to their countries of origin; the purpose of the contest is not to attract foreigners. The main goal is to secure real scientific achievements. However, many universities rely on foreign scientists. There are about 130 contenders for the multi-million-ruble grant.

Invited scholars will have to lead a newly created laboratory and work there for at least four months a year. As explained on the eve of the contest by officials of the Ministry of Education, when considering applications, experts will first look not at what research the higher educational institution is able to offer, but in fact, whether the implementation of this proposal will create a world-class laboratory. Under the project's performance guidance, the Ministry of Education wants to create a laboratory that will be able to operate efficiently and successfully after the completion of the project (three years), with a staff able to publish articles in prestigious scientific journals, secure orders from commercial firms, register patents on designs and win in new contests, etc.

Unlike other government initiatives, the idea of attracting leading scientists to Russian universities has

been the one that has ruffled most feathers in the local scientific community. There are many critics and skeptics, and those who generally support the idea have reservations as to how it will be implemented; in particular, whether a competent and honest examination of the applications and, accordingly, the picking of winners, will be possible. How justified these concerns are will become clear when the results of this competition are announced. By then, it will be possible to assess the quality of the winning bids and, therefore, the scientists who will have to create world-class laboratories.

Thus, the government supports university science, not only through the size of the funding for the universities, which are awarded different statuses, but also through related initiatives. Thus, the State seeks to improve the quality of training at Russian universities, develop the research and technological bases of universities, and bolster their innovative dimension.

Skeptics fear that universities will be unable to cope with the tasks and that the money and equipment will fall into the hands of those less able to use them effectively. To such criticism, the leadership of the Ministry of Education has only one answer; resources will be allocated to universities only on the basis of merit and cause. The winners will have to win in a tough competition. No carefree life is promised. And if a university begins losing position throughout the course of the program, it can lose its status. In this case, there is no choice but to believe the word of the officials. Let's see what comes out of this grand-scale project and who, in the end, turns out to be right - the skeptics or optimists. ●

Universities and Business Cooperation Contest Results

Ivan Sterligov, STFR.ru

The results of the second round of the Universities and Business Cooperation Contests have been announced. Deputy Minister of Education and Science Sergey Mazurenko discussed the successes and difficulties related to state support based on Government Resolutions №218 and 21 at a briefing on October 6, 2010.

The government has suggested providing two types of subsidies as a means of state support – one implementing cooperation between universities and business (Resolution 218, with 19 billion rubles), and the other dedicated to creating university-based innovation infrastructure (Resolution 219, with 8 billion rubles). The government has made the decision to allocate this money in a single-installment payment, and the Ministry of Education and Science will receive the funding shortly. The contests have already been conducted, and the latest lists of winners can now be found on the program's website.

In total, 199 applications for participation in the Innovative Infrastructure Contest were submitted; 197 were deemed qualified, of which 56 won. The contest score was therefore 3.5. In terms of geography, applications originated from 35 different regions. Twenty originated from the Central Federal District; 8, from Siberia; and 7, from the Privolzhsky and Northwest Districts. “We are aiming to select applications of the appropriate level,” – the deputy minister said, commenting on the regional distribution of winners.

As we predicted, the winners include 5 federal and 20 national research universities. The winners



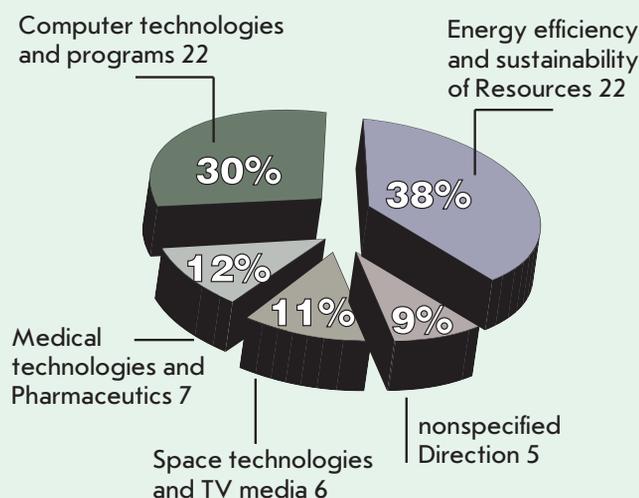
Sergey Mazurenko

will receive up to 50 million rubles a year for 3 years. The funds are meant to be used for the creation and development of Centers of Technology Transfer, as well as for the management (administration) of intellectual property, stimulation of innovations, etc.

Resolution 218 was slightly different. Applicants showed greater interest; however, only 553 of the 806 applications received were deemed eligible for the contest. One hundred and twelve applications, received from 107 companies and 76 universities, were successful. Among them are 25 research universities and 6 federal universities.

Based on the conditions of the contest, the funds allocated for research and development at uni-

Fig. 1. Distribution of winners in the first round of the Universities and Business Cooperation Contest in the priority directions of the Presidential Modernization Board.



Universities leading in cooperation and innovative infrastructure contest

Universities	Cooperation program	Innovative infrastructure
N.E. Bauman Moscow State Technical University	3	1
Moscow Physico-Technical Institute	3	1
National University of Science and Technology "MISIS"	3	1
B. N. Yeltsin Urals Federal University	3	1
Saint Petersburg Electrotechnical University "LETI"	3	1
Kazan State University	3	1
M.V. Lomonosov Moscow State University	3	1

versities will not be channeled directly to the university but will go through its financial partner, which then has to match the sum provided by the state and offer the university an amount of money equal to or higher than the amount offered by the state. In addition, 20% of the matching funds should be used for scientific and technological purposes.

The state's subsidies will not exceed 100 million rubles per year for a period of up to 3 years – a total of 300 million rubles. In addition, the subsidy per year should exceed 20% of the participating company's annual revenue for the last three years. Perhaps, this explains why the actual average of the subsidy for the entire period of the contract amounts to only 70 million roubles, which is less than a quarter of the highest amount possible. According to Sergey Mazurenko, "30 million is enough for a good project to be carried out." Yet, participants who asked for 3-5 million were excluded from the competition. Nevertheless, the amount of off-budget funds exceeded 20 billion rubles because of major contestants (participants). Based on the results of the past two selection rounds, the government will invest 18 billion rubles.

The head of the Department of Strategy and Development

(a branch of the Ministry of Education and Science) Igor Remorenko let it be known that the winners will be carefully monitored. Currently, a computer-based system through which business and universities will be held accountable on co-operation once a month is being developed. Having realized that expenditures will be strictly monitored, two companies have already withdrawn their winning applications.



Igor Remorenko

While commenting on the broad and far-reaching changes in the system of Russian science funding, Sergey Mazurenko gave journalists the following advice:

"We have to inform the public correctly. Very often, conversations arise about the underfunding of a scientific sector. Now, we are discussing the heightened attention that the state is devoting to science. However, the landscape of Russian science includes several areas. When I was a student at the Physico-Technical Institute, Kapitza used to tell us: there is fundamental and applied research, but there is one science. Scientific research is carried out in various organizations. However, we cannot make do without fundamental science either. Now fundamental research is strikingly different from Newtonian experiments, and it requires powerful tools and setups. Even the most developed countries cannot cope alone with the task of creating these setups. We actively participate in such projects as LHC, IER, and the free electron-based laser project. All this confirms serious focus on fundamental science. In addition, one of the tasks is to develop a large setup in Russia; so the PIK in Gatchina is nearly completed. But now we are talking about applied science." ●

International Evaluation Procedure for Supergrants

Ivan Sterligov, STRF.ru

Sergey Ivanets, director of the department of international integration at the Ministry of Education and Science, talks about the specifics of the evaluation procedures for inviting leading scientists (Decree of the Government of the Russian Federation dated April 9, 2010, № 220: "On Measures for Attracting Leading Scientists into Russian Higher Professional Educational Institutions")

Are international experts involved in assessing supergrant applications, as was promised earlier?

Yes, of course. Of the 900 selected experts, more than half are foreigners recommended to us by the American Association of Universities, the Association of European Universities, the Association of Universities and Institutes of Higher Education of Germany, the Association of Universities of Flanders, the International Bureau of the Federal Ministry of Education of the Federal Republic of Germany, the American Charitable Foundation for Informational Support of Education and Science, the United States National Science Foundation, and others. Russian experts were mostly provided by the Russian Foundation for Basic Research.

Each application is sent to two Russian experts and two foreign experts. If the opinions of the experts turn out to be completely opposite and cannot be reconciled, then that application is assessed by additional experts. The experts are paid, as that is the accepted practice, but I cannot say that this is a major expenditure.

Has the evaluation procedure been launched? What is the timeline for this procedure?

We plan on completing the expert evaluation procedure by the end of October. Most of the applications have already been sent out. Some of the experts have asked for more time, because of their personal schedules. But judging by the current progress, most experts need only a few days to complete the task.

Therefore, by the beginning of November we will have our winners, who will be selected by the grant council.

What is the role of formal criteria, including those that characterize publication activity and number of citations?

Formal parameters that characterize scientific expertise were required in the competition documentation. These parameters are very important for the evaluation procedure. In general, the expert evaluates each application for a group of issues. These can be divided into three parts: one characterizes the leading scientist, his standing in the scientific community; the second characterizes the application project, its quality, and the predicted results of the project; the third characterizes the higher education institution and its potential and ability to undertake and complete the project.

Of course, both formal and informal criteria are important for

evaluating all of these issues. However, there are criteria that cannot be disputed, such as citation indexes.

Will all the winners in natural and precise sciences have a high Hirsch index and citation index?

Here I must explain the role of the governmental grant council, which will be making the final decision on the winners. The results of R&D competitions all over the world are decided based on the results of expert evaluation; however, the final decision is made by special committees or the heads of the organization which provides the funding.

It was precisely for these reasons that our council included renowned scientists with the highest bibliometrical indices. They possess wide knowledge, which allows them to assess any application, and they can take various issues into account. However, the results of the expert evaluation are the basis for their decisions.

Were applications concerning public and humanitarian fields also sent to international experts? Is there any difference in the treatment of the various scientific fields?

There is none. There are no quotas and no discrimination concerning the themes of the applications.



Sergey Ivanets

But we can be sure that priority fields of research will yield much stronger applications. Won't this mean that some lines of scientific research will be left with no winners?

The council has selected 21 lines of research. There were no limitations for applications. It is natural that they will be unevenly distributed between the various scientific fields. However, we do not think that any quotas are required. Using them would make inviting the best leading scientists pointless.

On the other hand, it is no good if all of the winners belong to the same scientific field. We would like to see some variety. But resolving this issue is not up to the ministry; it will be decided by the council, which was created exactly for this reason.

Among the applicants, some of which are renowned and highly cited scientists, there are some very elderly people. Is age a factor in evaluating the application?

There are no formal limits on age. We have no reason to incor-

porate them. Of course we would like to see more leading scientists who are young and talented people, since the task of creating a scientific group requires an energetic approach. However, I think many would agree with me that some scientists who are no longer young will give any young researcher a run for their money.

Of course the experts will factor age into their evaluation. Their task is to evaluate the expertise of the applicant, taking activity and age into account. The experts can assess this by analyzing publications in recent years. If a 90-year-old person publishes several articles every year, then there are no reasons to doubt his or her working ability.

There is a certain group of applicants who are heads and deputy heads of major institutes and higher educational institutions. Sometimes they apply together with an educational institution in an adjacent city, let us say from Novosibirsk to Tomsk or from Moscow to

Tver. What can you say about the participation of such people in the competition?

We have a formal requirement: the scientist must apply into a different Federation constituent entity. This requirement must be fulfilled. However there is another requirement: starting in 2011 the scientist must spend at least 4 months in the year working in the higher educational institution. In other words, the work of the leading scientist cannot be limited to "Skype supervision."

If you are a director or deputy director of an organization, then in order to leave your place of employment for several months you must resolve formal issues. We have a number of such applications. I think the council members (if they decide to grant such applications) will require a formal confirmation, which will acknowledge the possibility that the head of the organization will be absent from his post for the allotted time. This will be strictly enforced even before awarding the grants.

How many "stars" will there be among the winners? Are there enough real world-class leading scientists among the applicants?

My answer is this: the average Hirsch index is not that high, it is 17.5. However, more than 70 applicants have an index in excess of 30. In other words, a considerable number of applicants are indeed world-class researchers. The average citation index of the 10 best papers of each applicant is 559. However, more than a 100 people have a citation index in excess of 1,000 and several tens of people have more than 2,000.

Thus, we have a considerable number of world-renowned and acknowledged scientists among our applicants. In other words, we have our choice of specialists. ●

Toll-Like Receptors (TLRs): The Role in Tumor Progression

D.V. Shcheblyakov, D.Y. Logunov*, A.I. Tukhvatulin, M.M. Shmarov, B.S. Naroditsky, A.L. Gintsburg

Gamaleya Research Institute of Epidemiology and Microbiology, Russian Academy of Medical Sciences

*E-mail: ldenisy@yahoo.com

Received 28.08.2010

ABSTRACT Toll-like receptors (TLRs) are major components of the innate immune system that recognize the conserved molecular structures of pathogens (pathogen-associated molecular patterns; PAMPs). TLRs are found in many different cell types, ranging from epithelial to immunocompetent cells. TLR binding triggers the expression of several adapter proteins and downstream kinases, leading to the induction of key pro-inflammatory mediators. This results in the activation of both the innate immune response (elevated expression of antiapoptotic proteins, proinflammatory cytokines, and antibacterial proteins), as well as the adaptive immune response (maturation of the dendritic cells, antigen presentation, etc.). In consequence of their ability to enhance the specific and nonspecific immune reactions of an organism, TLR agonists are widely used in the therapy of infectious diseases and, as adjuvants, in the therapy of malignant neoplasia. However, to date, TLRs have had the opposite effects on tumor progression. On the one hand, TLR ligands can suppress tumor growth. On the other hand, TLR agonists can promote the survival of malignant cells and increase their resistance to chemotherapy. The purpose of this review is to summarize the available data on the effects of TLRs and their agonists on tumor progression, as well as the mechanisms underlying the differences in the effects of TLRs on tumor growth.

KEYWORDS toll-like receptors, agonists of innate immune receptors, tumor progression, innate immune response, inflammation

ABBREVIATIONS TLR – toll-like receptor, LPS – lipopolysaccharide, NF- κ B – nuclear transcription factor κ B, PRR – pattern recognition receptor, PAMP – pathogen-associated molecular pattern, DAMP – damage-associated molecular pattern, IRF – interferon regulatory factor, ss- and dsRNA – single-stranded and double-stranded ribonucleic acid, TNF- α – tumor necrosis factor α , IL – interleukin, IFN – interferon, NK-cells – natural killers, siRNA – small interfering RNA, TGF – transforming growth factor

INTRODUCTION

According to the modern concept, inflammation is one of the main causes behind the appearance and progression of a tumor [1, 2]. The causative mechanism of this phenomenon is not well understood, although some key events at the site of the inflammation, required for the appearance and progression of a tumor, are known.

1) The cells localized on the site of the inflammation sustain a high activity of the transcription factor NF- κ B [3], which is responsible for the expression of anti-inflammatory cytokines, many of them (GRO α , β , γ , IL-8, MIP-3 α) exhibiting a tumor-stimulating action [4, 5]. Moreover, NF- κ B is regarded as a major antiapoptotic factor, as it activates the expression of a series of genes that encode antiapoptotic proteins, such as IAP, Bcl-2, Bcl-X, etc. These proteins elevate the cells' resistance to various stress agents associated with inflammation [6, 7].

2) An inflammation is accompanied by the induction of the oxidative stress that causes the appearance and accumulation of mutations and genome rearrangements in cells [8].

3) Many anti-inflammatory cytokines (GRO α /CXCL1, GRO β /CXCL2, GRO γ /CXCL3, IL-8/CXCL8, MIP-3 α ,

and IL-1) and growth factors (TGF- β 1, PDGF, bFGF, TGF- α , IGF-I, IGF-II) that enhance the migration of stromal (fibroblasts) and epithelial cells, followed by their proliferation, are secreted at late stages of an inflammation [9]. In a chronic inflammation, processes of reparation and alteration often occur simultaneously, which enforces the cells to proliferate under hypoxia and genotoxic stress conditions, thus increasing the risk of mutations.

The most common and well-studied cause of inflammation is microbial invasion; the process by which the pathogen can affect, in one way or another, the host cell's homeostasis.

One of these mechanisms is the interaction of the pathogen's highly conserved molecular domains with pattern-recognizing receptors (PRR – RIG-I-like receptors, Nod-like receptors, C-type lectins, Toll-like receptors (TLRs), etc.) localized on the surface or intracellular membranes of eukaryotic cells [10].

Thanks to their capability to bind with various bacterial ligands, PRRs play a crucial role in the development of an inflammation, by initiating the development of the innate immune response through increasing the expression of some antiapoptotic proteins, anti-inflammatory

cytokines, and antibacterial proteins. They also stimulate the acquired immune response by inducing the maturation of dendrocytes, presentation of the captured antigen, and differentiation of naive T-helpers [11].

Therefore, studies focusing on the role of PRRs in the induction of tumorigenesis and stimulation of tumor progression in the course of bacterial infection are a pressing priority.

In this review, we focused our attention on the role of TLRs in the development of inflammatory processes and attempted to gauge their interrelationship with tumor progression.

The data accumulated to date suggest an association between TLRs and tumor growth. However, these data are contradictory; both the tumor-stimulating [12, 13] and tumor-suppressing [14, 15] effects of TLRs have been reported.

As such, we have sought to summarize the available data and brood over the possible mechanisms that underpin the variations in the observed effects of TLRs on tumor growth.

THE ROLE OF TLRs

TLRs function as members of the PRR family that mediates specific recognition of conserved Pathogen-Associated Molecular Patterns (PAMPs). TLRs activate the innate immune system by binding with PAMP and largely determine the development of the adaptive immune response [16, 17]. The most conserved role of TLRs is the activation of the antimicrobial immune response in both the skin and mucosa of respiratory, gastrointestinal, and urogenital tracts.

TLRs recognize microbial molecules, which results in the development of inflammatory reactions caused by the activation of the NF- κ B regulating expression of anti-inflammatory cytokines (TNF- α , IL-1, IL-6, etc.) and chemokines (MCP-1, MCP-3, GM-CSF, etc.).

TLRs have been implicated in the transcriptional and posttranslational regulation (proteolytic cleavage and

secretion) of antimicrobial factors, such as defensins α and β , phospholipase A2, lysozyme, and so on. [18]. TLRs intensify the phagocytosis of microorganisms and optimize their inactivation by regulating the release of peroxy radicals and nitric oxide [19, 20].

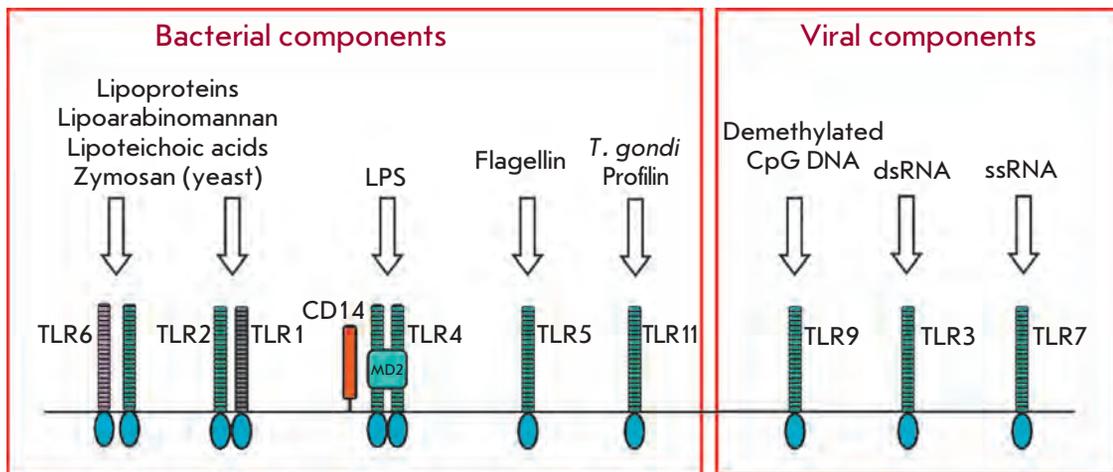
It is known that TLRs localized on the surface of endothelial cells indirectly support the migration of leucocytes into the inflammation focus by stimulating the expression of leucocyte adhesion molecules E-selectin and ICAM-1 [21].

The TLR stimulation directly leads to the elevation of interferon (IFN)- α/β production by both stromal and hematopoietic cells, which is important for the organism to defend itself against certain viral and bacterial infections [22]. Moreover, as was recently discovered, TLRs – via the activation of a series of factors, such as FADD, caspase 8, and protein kinase R (PKR) or the stimulation of IFN- α/β expression – can induce the development of apoptosis, an important mechanism of tissue protection against microbial pathogens [23].

TLRs play a crucial role in the regulation of the adaptive immune response. In particular, the TLR-dependent activation of professional antigen-presenting dendritic cells is determinative in several essential processes providing the development of the adaptive immune response, such as the activation of mature T-cells, processing and presentation of microbial antigens, elevation of the expression of the costimulatory molecules (CD80, CD86) required for the activation of naive CD4+-T-cells, and suppression of regulatory T-cells via IL-6 production [24]. The TLR-dependent activation is also important for B-cell proliferation and maturation during the infection process [25].

Thus, TLRs play an important role in initiating the development of the inflammatory process (activation of innate immune reactions) in response to the introduction of various pathogens (including protozoa, fungi, bacteria, and viruses). Moreover, according to the modern view, pathogen recognition by TLRs is the midpoint

Fig. 1. Toll-like receptors and their ligands.



in the development of the adaptive immune response, which is the second line of defense [11]. TLRs also participate in the functioning of the normal gastrointestinal system and are implicated in the pathogenesis of autoimmune diseases (systemic lupus erythematosus), arthritis, atherosclerosis, and certain other disorders [26, 27]. Recent data indicate that TLRs can either activate antitumor immunity [28, 29] or, on the contrary, stimulate tumor progression [30, 31].

TLR STRUCTURE, EXPRESSION IN DIFFERENT CELL TYPES, AND SPECIFICITY TO VARIOUS MOLECULAR STRUCTURES (PAMP AND DAMP)

Based on their structure, TLRs are members of the family of IL-1 receptors (IL-1R). They are transmembrane proteins that are expressed on the cell's surface and in subcellular compartments (particularly endosomes). TLR localization depends on the type of the recognized ligand. For example, the TLRs 1, 2, 4, 5, and 6, which bind with structural bacterial components, are localized on the cell's surface, while the TLRs 3, 7, 8, and 9, primarily recognizing virus-associated nucleic acids (dsRNA, ssRNA, and DNA), are localized in endosomes, where they interact with the ligands that appear after the deproteination of virions [16].

The TLR's structure includes the N-terminal leucine-rich repeat (LRR) responsible for ligand binding, one transmembrane domain, and the C-terminal cytosolic signaling domain (homologous to the intracellular domain of IL-1R) [32].

TLRs are expressed in most types of cells in the human body, including non-hematopoietic epithelial and endothelial cells. The number of simultaneously expressed TLRs and their combination are specific for each cell type; the largest number is observed in cells of hematopoietic origin, such as macrophages, neutrophils, and dendritic cells (Table 1) [33].

To date, 13 different TLRs have been identified in mammals: 10 in humans and 12 in mice. The TLRs 1 through 9 are conserved in humans and mice; however, some difference exists. The gene encoding TLR10 is only found in humans, whereas the TLR11 gene is present in both species, but possesses functionality in mouse only [34].

The hallmark of TLRs that distinguishes them from the receptors of acquired immunity (T- and B-cell receptors) is their capability to recognize not just unique epitopes, but also the evolutionary-conserved pathogen-associated molecular patterns (PAMP), which are widely distributed in all taxa of microorganisms and viruses, regardless of their pathogenicity.

The specificity of PAMP recognition is well studied for the majority of TLRs. The ligands of TLR 1–9 and 11 are now known (Fig. 1). The biological role and the

Table 1. Activation of the transcription factors NF- κ B and IRF by different TLRs.

TLR type	Activation of NF- κ B	Activation of IRF
TLR2/1/6	+	-
TLR3	+	+ (IRF3)
TLR4	+	+ (IRF3)
TLR5	+	-
TLR7	+	+ (IRF5, 7)
TLR8	+	+ (IRF5, 7)
TLR9	+	+ (IRF5, 7)

specificity of human TLR10 and murine TLR 12 and TLR 13 remain unclear [16].

The most common microbial ligands of TLRs are the following:

- bacterial lipopeptides, lipoteichoic acid (LTA), and peptidoglycans; mycobacterial lipoarabinomannan; and zymosan from the fungal cell wall – which bind with TLR2 that forms heterodimers with TLR1, TLR6, and CD14;
- LPS of Gram-negative bacteria; the TLR4 ligand;
- flagellin, a principal component of bacterial flagella that activates TLR5;
- protozoan profilin-like structures binding with TLR11;
- DNA (demethylated CpG-islets) recognized by TLR9;
- dsRNA, the TLR3 ligand;
- ssRNA, the TLR7 and the TLR8 ligand.

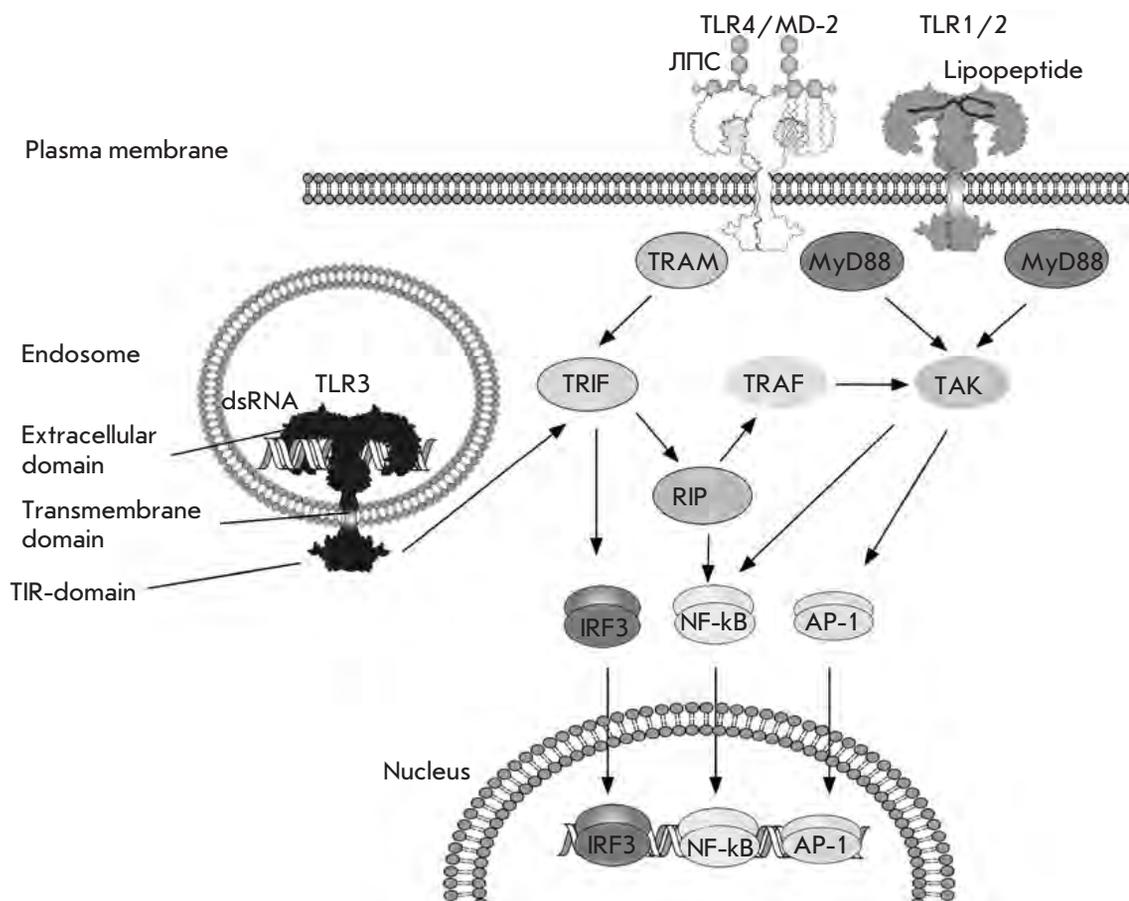
It was shown recently that TLRs can be activated by many endogenous molecules, the so-called alarmins (hyaluronic acid, heat shock proteins, etc.), appearing in tissue decay. These heterogeneous in their nature and structure (PAMP and alarmins) substances recognized by TLRs are presently allied to one family called DAMP (Damage Associated Molecular Patterns) [35].

SIGNALING CASCADE ACTIVATED FOLLOWING THE INTERACTION OF TLRs WITH THEIR OWN LIGANDS

Now, after describing the TLR's structure and its actions, we can focus on the processes set in motion after their binding with their own ligands.

The binding of a ligand with TLR initiates a cascade of signals going back to the plasmatic TIR-domains of TLR. The signal from the TIR-domain – via the adapter molecules MyD88 (myeloid differentiation factor 88), TIRAP (TIR domain-containing adapter proteins), TICAM1 (TRIF), TICAM2 (TIR-containing adapter molecule 2) – is transmitted to the corresponding kinases (TAK, IKK, TBK, MAPK, JNKs, p38, ERK, Akt, etc.) that specifically activate the transcription factors (NF- κ B, AP-1, and IRF) responsible for the expression of various anti-inflammatory and antimicrobial factors. All TLRs, except for TLR3, transmit the signal to kinases using MyD88.

Fig. 2.
TLR-dependent
signaling path-
ways.



The TLR3 transmits the signal via TICAM1, and TLR4 – via both MyD88 and TICAM1 (Fig. 2).

Activation of one or the other factor depends on the type of TLR that is transmitting the signal. In particular, almost all TLRs (TLR2 and its co-receptors TLR1 and TLR6, as well as TLR4–9 and TLR11) bind with their own ligands to activate NF- κ B, one of the main factors regulating the expression of anti-inflammatory cytokines, such as IL-1, -6, -8, etc. Signal transduction via TLR3, 4, 7–9 leads to the activation of another family of anti-inflammatory transcription factors, IRF. Signals transmitted via TLR3 or TLR4 lead to the activation of IRF3, which regulates the expression of IFN- β and is regarded as a crucial component of antiviral immunity. Signal transduction via TLR7–9 results in the

activation of IRF-5 and IRF-7, as well as the expression of IFN- α , which is of critical importance for the antiviral defense. Signaling via TLR2 or TLR5 does not result in the activation of IRF family factors [36].

Thus, the interaction of the particular TLR with its ligand triggers a signaling cascade that leads to the expression of a specific gene pattern (cytokines, antimicrobial factors, etc.).

However, many details of the triggering of TLR-dependent signaling and its downstream effects still remain unclear. To the best of our knowledge, there is no reported data characterizing the overall transcriptome and proteome alterations occurring in response to the activation of distinct TLRs.

Table 2. TLRs in clinical trials.

Cancer	TLR
Non-small cell lung carcinoma, late stage	TLR9
Melanoma, stage IV	TLR7
Melanoma, stage IIIb/c or IV	TLR9
Incompletely re-sectable pancreatic cancer	TLR2/6
Recidivism of non-Hodgkin's lymphoma	TLR9
Recidivism of glioblastoma	TLR9
Chronic lymphocytic leukemia	TLR7

Table 3. Effect of TLRs on tumor growth and development.

Tumor-stimulating activity	TLR	Antitumor activity	TLR
Stimulation of angiogenesis	2, 9	Suppression of angiogenesis	7, 9
Stimulation of proliferation	3, 4	Development of apoptosis	3, 4, 7, 9
Chemoresistance	4	Elevation of chemosensitivity	2, 4, 7
Activation of regulatory T-cells (T _{reg})	4, 5	Inhibition of T _{reg} , antigen presentation	4, 5, 7, 8, 9
		Cytotoxicity	9

TLRS AND TUMORS

Fundamentally different effects of TLRs on tumors have been reported to date. On one hand, TLRs (and their ligands) can suppress tumor growth, on the other hand, they can stimulate tumor progression and influence tumor resistance to chemotherapy. In order to explain these contradictory data, we shall consider both cases in detail.

ANTITUMOR ACTIVITY OF TLRS

Currently, many TLR agonists are in clinical trials as antitumor agents (Table 2). Particularly, both natural (ssRNA) and synthetic (imiquimod) agonists of TLR7 and TLR8 have demonstrated high activity against chronic lymphocytic leukemia and tumors of the skin [37]. The TLR9 ligand CpG can suppress the growth of lymphomas and tumors of the brain, kidney, and skin [28]. The TLR3 ligand poly(IC) possesses a proapoptotic effect on both tumor and surrounding cells (for example, endothelial cells) [38].

It has been reported that the TLR4 agonists LPS from Gram-negative bacteria and OK-432 (picibanil; a lyophilized preparation of a low-virulence strain of *Streptococcus pyogenes*, inactivated by heating with penicillin G) possess high antitumor activity, when administered intratumorally. However, both LPS and OK-432 could not suppress tumor growth upon systemic administration [39]. At present, OK-432 is being tested at the second stage of clinical trials as medication against colorectal and lung cancer. It has also been shown that OM-174, a chemical agonist of TLR2/4, can inhibit the progression of melanoma and increase the survival rate of animals in experiments when introduced together with cyclophosphamide [40]. These experiments have shown that TLR2/4 agonists induce the secretion of TNF- α and expression of inducible NO-synthase. NO is known to induce apoptosis in tumor cells resistant to chemotherapy, thus prolonging the lifespan of mice. Another known antitumor preparation of microbial genesis is BCG (Bacillus Calmette-Guérin), which activates TLR-dependent reactions (TLR2, 4, 9). This preparation has been successfully used in the therapy

of urinary bladder tumors for over 30 years [41].

Of note, various TLR agonists are presently in clinical trials as potential medications against tumors of different geneses (Table 2).

One of the main mechanisms underlying the antitumor activity of TLRs is their capability to activate the development of a tumor-specific immune response. The activation of TLRs

- 1) stimulates (directly or indirectly) the migration of NK-cells, cytotoxic T-cells, and type I T-helpers into the tumor, which causes the lysis of tumor cells via secretion of various effectors (perforin and IFN- γ) [42]; and
- 2) results in the secretion of type I IFNs (IFN- α , β) [43].

Another possible mechanism underlying the antitumor effect of TLRs is the TLR-dependent transition of tumor-stimulating macrophages (M2 type) into the tumor-suppressing type M1. Type M2 macrophages are characterized by the expression of cytokines, such as TGF- β and IL-10 ---components required for tissue repair and remodeling. TGF- β stimulates tumor cell proliferation, while IL-10 directs the development of the immune response to Th2, thus blocking the development of the cellular antitumor immune response. Alternatively, type M1 macrophages express IL-1, -6, -12, TNF- α , and IFN- γ , thereby stimulating the development of the cellular antitumor (Th1) immune response [44].

TUMOR-STIMULATING ACTIVITY OF TLRS

Chronic infections and inflammations are the most important factors known to stimulate the development of a malignant neoplasm. In particular, stomach cancer can be associated with a chronic inflammation induced by a pathogen; namely, *Helicobacter pylori*, and a chronic inflammation of the digestive tract is often associated with colorectal cancer [45]. Moreover, in some cases the use of nonsteroidal anti-inflammatory drugs can decrease the risk of malignant neoplasm development [46].

TLRs are key players in the system of innate immunity in animals, including humans; they participate in the mechanisms of inflammatory response when the cells come into contact with various pathogens. The role of TLRs in the development and progression of vari-

ous tumors is being studied in detail currently. Several mechanisms have been proposed to explain TLR implication in the stimulation of tumor formation and development (Table 3).

NF- κ B is one of the major factors that provide the interrelationship between chronic inflammation and tumor formation [47]. This factor is constitutively activated in more than 90% of human tumors, such as acute and chronic myeloid leukemia, prostate cancer, multiple myeloma, hepatocarcinoma, etc. [48, 49]. In relation to this, the agents that are capable of activating NF- κ B can be directly implicated in tumor development and progression. The interaction of pathogens with TLRs on the cell's surface is known to result in the activation of NF- κ B and expression of NF- κ B –dependent genes, thus determining the participation of TLRs in the stimulation of carcinogenesis. The activation of NF- κ B leads to an increase in the level of IL-1, IL-2, IL-6, IL-10, and TNF- α cytokine production, migration of immune cells to the inflammation's focus due to the increased production of chemokines, “maintenance” of chronic inflammation, increase in the production of anti-apoptotic factors, etc. These indicated properties can promote tumor survival and progression due to the inhibition of apoptosis and cytotoxicity, and induction of angiogenesis [50].

An elevated level of TLRs has been reported in various tumor cells, and the frequency of induced tumor formation is decreased in *TLR* knockout mice [67]. Moreover, the increase in TLR expression on prostate cancer or head and neck tumor cell surfaces can stimulate their proliferation [51].

Huang and associates [31] have demonstrated that *Listeria monocytogenes* has a direct tumor-stimulating effect associated with its ability to activate the TLR2-dependent signaling pathways in ovary cancer cells. Moreover, the TLR2-dependent activation of NF- κ B caused by *L. monocytogenes* results in an enhanced resistance of tumor cells to chemotherapeutics [31]. The interrelationship between TLR2 and tumor progression has been confirmed in another independent study, in which Karin *et al.*, [67] proved this receptor's key role in lung cancer metastasis formation. Metastasis and progression of tumors are essentially retarded in *TLR2* knockout mice, compared with wild-type mice. The key role in lung cancer progression belongs to myeloid cells expressing TNF- α in response to their stimulation by versican (proteoglycan of extracellular matrix, ligand for TLR2, the level of which is elevated in many tumor cell types). In our research, we also studied the role of TLR2 in tumor progression. In particular, infection with mycoplasma (*Mycoplasma arginini*) or the addition of its structural components (LAMP) to the cells expressing TLR2 resulted in the suppression of

apoptosis and acceleration of tumor growth *in vivo* [52, 53]. Thus, TLRs have a tumor-stimulating effect that is mediated by myeloid cells [54].

Similar data was obtained for TLR4, another member of the TLR family. Systemic (intravenous) injection of LPS (the ligand of this receptor) stimulated breast adenocarcinoma cell migration, increased their invasion, and stimulated angiogenesis in tumors as well [30]. Similar results were obtained on another model – colorectal adenocarcinoma; LPS increased the survival of tumor cells, stimulated their proliferation, and, when injected intraperitoneally, enhanced metastasis [55]. Moreover, Huang, *et al.* have demonstrated that tumor cells expressing TLR4 cause a substantially more aggressive development of a disease, reducing the lifespan of mice in comparison with mice of isogenic line with TLR4 inactivated by specific siRNA. The obtained data has made possible the supposition that endogenous ligands (heat shock proteins, β -defensines, and endogenous LPS thrown from the gut) can influence the progression of TLR4-positive tumors, partially akin to the tumor-stimulating effects of TLR2 and its endogenous ligand versican [56].

However, data illustrating the tumor-stimulating effect of TLRs have been obtained not only for TLR2 and TLR4; elevated expression of TLR5 and TLR9 on the cervical epithelial cell's surface can be associated with cervical cancer progression [57] as well. A high level of TLR9 expression was observed in clinical samples of lung cancer and tumor cell lines. In these cells, TLR9 stimulation by specific agonists resulted in increased production of tumor-associated cytokines [58]. The TLR9 level is also elevated on the surface of human prostate cancer cells [59]. The treatment of such cells with the TLR9 ligands CpG-oligodeoxynucleotide (ODN-CpG) or bacterial DNA elevated the invasion of tumor cells. The elevation of tumor cell invasion via TLR9 activation can be regarded as a new mechanism by which chronic infections can stimulate prostate cancer cell growth.

However, not only various infectious agents and their structural components can stimulate carcinogenesis via interaction with the TLRs. DAMP, the nuclear and cytoplasmic proteins of necrotic cells, are known to serve as TLR ligands. DAMP released from damaged cells can be recognized by various TLRs on the surface of immune cells with subsequent activation of TLR-dependent signals resulting in the suppression of the antitumor immune response and, as a consequence, in the stimulation of tumor progression [60].

The molecules that potentially possess a tumor stimulating effect include heat shock proteins (HSP60, 70), ATP and uric acid, the Ca²⁺-modulating protein family (S100), the protein HMGB1 and nucleic acids, whose DNA-binding protein HMGB1 is the most well-studied.

HMGB1 released from damaged cells activates the immune system via interaction with TLRs. In cell cultures, HMGB1 was shown to stimulate melanoma, as well as breast, colon, pancreas, and prostate cancer cell growth. HMGB1 can activate TLR2 and TLR4 on the surface of tumor and immune cells and, as a result, induce tumor progression and metastasis [61].

Elevated expression of DAMP proteins, particularly S100 family members, is shown in melanoma cells; these proteins can stimulate the growth of melanoma cells and lymphocytes, peripheral blood lymphocytes, and act as an autocrine tumor growth factor. The S100A4 protein, being a TLR ligand, stimulates the metastasis of breast cancer cells, and its elevated expression is an indicator of a poor prognosis. In spite of the interrelationship between S100A4 and metastasis, the protein can be expressed in macrophages, lymphocytes, and fibroblasts. Recent studies have shown that the S100A8 and S100A9 proteins produced by primary tumor can activate serum amyloid A (SAA) 3 in lung tissues, thereby creating conditions for the formation of a metastatic niche. SAA3 is a ligand for TLR4 on the surface of lung endothelial cells and macrophages. TLR4 activation facilitates the migration of tumor cells from the primary focus into the lung tissue via the creation of a microenvironment favorable to tumor growth. Thus, the suppression of the S100-TLR4-signaling pathway effectively counteracts the formation of metastasis in the lung [62].

In summing up the described effects, one can come to the conclusion that TLR can, on the one hand, directly or indirectly participate in tumor progression, and, on the other hand, increase tumor cell resistance to proapoptotic agents.

The data presented illustrate a complex mechanism showing the tumor-stimulating effect of TLRs and their ligands, which is to be studied in detail. However, despite the complexity of the issue, one can highlight some key points that determine the tumor stimulating effect of TLRs:

1) TLR interaction with its own ligands induces the activation of the NF- κ B transcription factor and, as a consequence, an increase in the production of various pro-inflammatory cytokines (IL-6, MCP-1, MIF, GRO- α , etc.), as well as a number of anti-apoptotic proteins, which facilitates a direct or indirect tumor-stimulating effect;

2) TLR-dependent activation of myeloid cells and their progenitors is apparently the determining factor in the formation of metastasis. A series of independent works have demonstrated that myeloid cells migrating from the bone marrow (in response to endogenous stimulation) into tissues play a key role in the formation of metastasis niches [30, 54]. Since endogenous (versi-

can, fibronectin, etc.) and exogenous (microbial origin) TLR ligands are known to be capable of stimulating myeloid cells and their progenitors and increasing the metastatic potential of the tumor, it is very likely that an interrelationship exists between the TLR-dependent activation of myeloid cells and their subsequent involvement in metastasis;

3) TLR activation can stimulate angiogenesis via angiogenic factors, such as IL-8, vascular endothelial growth factor (VEGF), and matrix metalloproteinases (MMP), as well as enhance the adhesive and invasive activity of tumor cells, in line with the increase in the permeability of blood vessels.

TOLL-LIKE RECEPTORS IN TUMOR THERAPY

Antitumor therapy based on the delivery of TLR ligands into the foci of tumor growth seems promising, because TLR agonists can induce an antitumor immune response by regulating the function in immune cells localized in the microenvironment of the tumor. The imiquimod containing the TLR7 agonist may represent an example of such therapy. This drug is used against actinic keratosis and basal cell carcinoma. The possibility of using this drug as an adjuvant in the therapy of melanoma is also under study [63, 64]. Yet another TLR7 agonist, 852A, is used in tumor therapy. Presently, the possibility of using 852A against chronic lymphoid leukosis and other solid tumors is under consideration [65]. ODN-CpG, a TLR9 agonist, induces the activation and maturation of dendrite cells and stimulates the development of the T-cellular antitumor response. Clinical trials of TLR9 agonists are presently under way in the context of their safety and efficacy in the therapy of breast, colorectal, and lung cancers, melanoma, glioblastoma, etc. [28]. Macrophage-activating lipopeptide-2 (MALP-2), a TLR2/6 agonist, has shown encouraging results in the therapy of pancreatic cancer; the intratumoral injection of MALP-2 with gemcitabine during laparotomy substantially increased the lifespan of patients with incompletely resectable cancer (from 9 to 17 months) [66]. The described examples demonstrating the effective use of TLR agonists in tumor therapy are evidence of their potential, as well as the reasonableness of further studies directed toward the development of antitumor pharmaceuticals with a similar mechanism of action.

However, as mentioned above, many tumor cells can express TLRs on their surface, and the direct interaction of these cells with TLR ligands can enhance the progression of the tumor and make it less sensitive to chemotherapy. Thus, it is likely that TLR agonists permanently circulating in an organism (pathogenic microorganisms that can overcome the immune barrier; LPS of bacteria comprising normal gut microflora,

which can be thrown into the bloodstream; and own endogenous ligands) can directly or indirectly favor the advancement of tumor progression.

In this regard, a promising avenue for the therapy of malignant tumors is an approach directed toward the suppression of TLR-dependent signaling pathways. The employment of NF- κ B inhibitors can be distinguished as a known approach that is already used in the therapy of malignant tumors.

The constitutive activation of NF- κ B is observed in tumors, such as Hodgkin's lymphoma (Hodgkin's disease), acute lymphoblastic leucosis, multiple myeloma, breast, colon, lung, ovary, and prostate cancers, various lymphomas, liver cancer, and melanoma, amongst others [48, 49].

The following pharmaceutical groups are used for the suppression of NF- κ B activity: nonsteroidal anti-inflammatory agents; substances-inhibiting IKK and COX-2 activities; natural and bioavailable inhibitors of IKK β , such as flavonoids, prostaglandins, BMS-345541, PS1145, SC-514, and SPC839; as well as proteasome inhibitors that inhibit NF- κ B activity by preventing I κ B degradation, such as bortezomib (PS-341), irinotecan, gemcitabine, and other drugs widely used in the therapy of tumors of the colon, small intestine, stomach, pancreas, etc. [67].

Since NF- κ B is a key component of the TLR-dependent signaling pathway, the use of its inhibitors seems promising for the suppression of TLR-dependent stimulation of tumor growth.

In our opinion, other promising targets are TLRs themselves. Since TLR2 and TLR4 (receptors implicated in the stimulation of tumor growth) are expressed on the cell's surface, specific molecules (antibodies, chemical inhibitors) suppressing their functional activity seem usable. Antibodies that inhibit TLR activities are currently available; however, data on their clinical use have not been reported yet.

CONCLUSION

TLRs are members of the PRR family. The actions associated with their activation go beyond reactions of the innate immune response. Participation in dendritic cell activation, regulation of specific T- and B-cellular immune responses, elevation of IFN expression, etc., provide for the TLR's implication both in the development of effective innate and adaptive immune reactions in response to the penetration of various pathogens into the organism, and in supporting tissue homeostasis. The reported data in abundant research is evidence that it is possible to use TLR ligands as adjuvants for the immune therapy of malignant neoplasms. It is known, however, that TLR activation on the surface of tumor cells can result in enhanced progression of tumors of

various origins.

Such a difference in effects primarily depends on the ligand type. As shown in Table 1, TLR can be divided into two groups: inducing and noninducing IFN production. As a rule, the addition of TLR3, 4, 7, 8, and 9 agonists that activate IRE considerably suppresses tumor growth. However, there is currently no data indicating any anti-tumor effect of the agonists of TLR2, which, unlike the listed receptors (TLR3, 4, 7, 8, 9), cannot activate the production of IFN type I. The way they are introduced is yet another feature leading to differences in the effects of TLR agonists on tumors. Intratumoral introduction of TLR3, 4, 7, 8, 9 ligands induces tumor cell death and, in most cases, a decrease in the tumor's size. The most probable explanation for the anti-tumor activity of these TLRs is their capability to (A) induce local expression of IFN types I and II, which are known to induce tumor cells death, and (B) activating cell immunity in response to the interaction with ligand. The death of tumor cells, their phagocytosis, and subsequent presentation of tumor-specific antigens cause additional stimulation of specific anti-tumor immunity. However, a series of works [30, 55, 56] have demonstrated that a systemic introduction of TLR4 ligands is often associated with the stimulation of tumor growth. In our opinion, this difference is related to the fact that intratumoral injection of TLR4 ligand (LPS), compared with its systemic introduction, results in substantially higher IFN accumulation directly in the tumor. Since IFNs are short-distance effector proteins acting at sufficiently high concentrations, their production outside the tumor (at systemic introduction) does not result in tumor cell death and, hence, in the development of anti-tumor immunity. Inflammatory cytokines and chemokines can play a dual role after a local or systemic introduction of LPS: they either favor the development of anti-tumor immunity after intratumoral injection of LPS, or positively influence the growth of the tumor, resistance of its cells, and metastasis potential after the systemic introduction of LPS in the absence of targets for the immune system.

Thus, the available data suggest that TLR agonists have a dual effect on tumor growth. This dual TLR effect is indicative of the more complex functional role of TLRs in tumor biology. Such a role by TLRs is obviously beyond the frame of a simple NF- κ B activation. The influence of various TLR ligands on tumors must be studied taking into account multiple factors, including the TLR expression level, type of tissue from which the tumor originates, the tumor's microenvironment, and many others. The systemic study of TLR's role and functions in tumor cells can contribute substantially to the development of new antitumor pharmaceuticals with a TLR-dependent mechanism of activity. ●

REFERENCES

1. Coussens L.M., Werb Z. // *Nature*. 2002. V. 420(6917). P. 860–867.
2. Mantovani A., Allavena P., Sica A., Balkwill F. // *Nature*. 2008. V. 454(7203). P. 436–444.
3. Okamoto T. // *Endocr. Metab. Immune Disord. Drug. Targets*. 2006. V. 6(4). P. 359–372.
4. Thomas H.G., Han J.H., Balentien E., et al. // *Methods Enzymol*. 1991. V. 198. P. 373–383.
5. Kleeff J., Kusama T., Rossi D.L. // *Int. J. Cancer*. 1999. V. 81. P. 650–657.
6. Deveraux Q.L., Schendel S.L., Reed J.C. // *Cardiol. Clin*. 2001. V. 19(1). P. 57–74.
7. Dong Z., Wang J.Z., Yu F., Venkatachalam M.A. // *Am. J. Pathol*. 2003. V. 163(2). P. 663–671.
8. Bartsch H., Nair J. // *Langenbecks Arch. Surg*. 2006. V. 391(5). P. 499–510.
9. Moser B., Willmann K. // *Ann. Rheum. Dis*. 2004. V. 63 Suppl 2. P. 84–89.
10. Rasmussen S.B., Reinert L.S., Paludan S.R. // *APMIS*. 2009. V. 117(5–6). P. 323–337.
11. Palm N.W., Medzhitov R. // *Immunol. Rev*. 2009. V. 227(1). P. 221–233.
12. Zeromski J., Mozer-Lisewska I., Kaczmarek M. // *Cancer Microenviron*. 2008. V. 1(1). P. 37–42.
13. Jego G., Bataille R., Geffroy-Luseau A., et al. // *Leukemia*. 2006. V. 20(6). P. 1130–1137.
14. Seya T., Akazawa T., Uehori J., et al. // *Anticancer Res*. 2003. V. 23(6a). P. 4369–4376.
15. Grauer O.M., Molling J.W., Bennink E., et al. // *J. Immunol*. 2008. V. 15. P. 6720–6729.
16. Medzhitov R. // *Nat. Rev. Immunol*. 2001. V. 1(2). P. 135–145.
17. Pasare C., Medzhitov R. // *Nature*. 2005. V. 438(7066). P. 364–368.
18. Zasloff M. // *Nature*. 2002. V. 415(6870). P. 389–395.
19. Doyle S.E., O’Connell R.M., Miranda G.A., et al. // *J. Exp. Med*. 2004. V. 5. P. 81–90.
20. Werling D., Hope J.C., Howard C.J., Jungi T.W. // *Immunology*. 2004. V. 111(1). P. 41–52.
21. Nijhuis M.M., Pasterkamp G., Sluis N.I., et al. // *J. Vasc. Res*. 2007. V. 44(3). P. 214–222.
22. Kaisho T., Akira S. // *J. Allergy Clin. Immunol*. 2006. V. 117(5). P. 979–988.
23. Salaun B., Romero P., Lebecque S. // *Eur. J. Immunol*. 2007. V. 37(12). P. 3311–3318.
24. Iwasaki A., Medzhitov R. // *Nat. Immunol*. 2004. V. 5(10). P. 987–995.
25. Pasare C., Medzhitov R. // *Adv. Exp. Med. Biol*. 2005. V. 560. P. 11–18.
26. Li M., Zhou Y., Feng G., Su S.B. // *Curr. Mol. Med*. 2009. V. 9(3). P. 365–374.
27. Curtiss L.K., Tobias P.S. // *J. Lipid. Res*. 2009. V. 50 Suppl. P. 340–345.
28. Krieg A.M. // *J. Clin. Invest*. 2007. V. 117. P. 1184–1194.
29. Chicoine M.R., Zahner M., Won E.K. // *Neurosurgery*. 2007. V. 60. P. 372–381.
30. Harmey J.H., Bucana C.D., Lu W. // *Int. J. Cancer*. 2002. V. 101. P. 415–422.
31. Huang B., Zhao J., Shen S. // *Cancer. Res*. 2007. V. 67. P. 4346–4352.
32. Means T.K., Golenbock D.T., Fenton M.J. // *Life Sci*. 2000. V. 8. P. 241–258.
33. Diebold S.S. // *Handb. Exp. Pharmacol*. 2009. V. 188. P. 3–30.
34. West A.P., Koblansky A.A., Ghosh S. // *Annu Rev. Cell. Dev. Biol*. 2006. V. 22. P. 409–437.
35. Zhang Z., Schluesener H.J. // *Cell Mol. Life Sci*. 2006. V. 63(24). P. 2901–2907.
36. O’Neill L.A., Bowie A.G. // *Nat. Rev. Immunol*. 2007. V. 7(5). P. 353–364.
37. Stockfleth E., Trefzer U., Garcia-Bartels C. // *Br. J. Dermatol*. 2003. V. 149 (Suppl. 66). P. 53–56.
38. Nogueras S., Merino A., Ojeda R. // *Heart. Circ. Physiol*. 2008. V. 294. P. 708–713.
39. Okamoto M., Oshikawa T., Tano T., et al. // *J. Immunother*. 2006. V. 29. P. 78–86.
40. D’Agostini C., Pica F., Febbraro G., et al. // *Int. Immunopharmacol*. 2005. V. 5(7–8). P. 1205–1212.
41. Morales A., Eidinger D., Bruce A.W. // *J. Urol*. 1976. V. 116. P. 180–183.
42. Street S.E., Cretney E., Smyth M.J. // *Blood*. 2001. V. 1; 97(1). P. 192–197.
43. Swann J.B., Hayakawa Y., Zerafa N., et al. // *J. Immunol*. 2007. V. 15;178(12). P. 7540–7549.
44. Martinez F.O., Sica A., Mantovani A., et al. // *Front. Biosci*. 2008. V. 13. P. 453–461.
45. Balkwill F., Coussens L.M. // *Nature*. 2004. V. 431. P. 405–406.
46. Robak P., Smolewski P., Robak T. // *Leuk. Lymphoma*. 2008. V. 49. P. 1452–1462.
47. Pikarsky E., Porat R.M., Stein I., et al. // *Nature*. 2004. V. 431. P. 461–466.
48. Palayoor S.T., Youmell M.Y., Calderwood S.K., et al. // *Oncogene*. 1999. V. 18. P. 7389–7394.
49. Griffin J.D. // *Blood*. 2001. V. 98. P. 2291.
50. Philip M., Rowley D.A., Schreiber H. // *Semin. Cancer Biol*. 2004. V. 14. P. 433–439.
51. Swann J.B., Vesely M.D., Silva A., et al. // *Proc. Natl. Acad. Sci. USA*. 2008. V. 105. P. 652–656.
52. Shcheblyakov D.V., Logunov D.Yu, Zubkova O.V., et al. // *Mol. Genetika, mikrobiol. i virusol. (Mol. Genet. Microbiol. Virol)*. 2008. № 4. P. 163–167.
53. Logunov D., Scheblyakov D., Zubkova O., et al. // *Oncogene*. 2008. V. 27. P. 4521–4531.
54. Kim S., Takahashi H., Lin W.W., et al. // *Nature*. 2009. V. 457. P. 102–106.
55. Luo J.L., Maeda S., Hsu L.C., et al. // *Cancer Cell*. 2004. V. 6. P. 297–305.
56. Huang B., Zhao J., Li H. // *Cancer Res*. 2005. V. 65. P. 5009–5014.
57. Kim W.Y., Lee J.W., Choi J.J., et al. // *Int. J. Gynecol. Cancer*. 2008. V. 18. P. 300–305.
58. Droemann D., Albrecht D., Gerdes J., et al. // *Respir. Res*. 2005. V. 6. P. 1–6.
59. Ilvesaro J.M., Merrell M.A., Swain T.M., et al. // *Prostate*. 2007. V. 67. P. 774–781.
60. Lotze M.T., Zeh H.J., Rubartelli A., et al. // *Immunol. Rev*. 2007. V. 220. P. 60–81.
61. Ellerman J.E., Brown C.K., Vera M., et al. // *Clin. Cancer Res*. 2008. V. 13. P. 2836–2848.
62. Cabezón T., Celis J.E., Skibshoj I., et al. // *Int. J. Cancer*. 2007. V. 121. P. 1433–1444.
63. Adams S., O’Neill D.W., Nonaka D., et al. // *J. Immunol*. 2008. V. 181. P. 776–784.
64. Spaner D.E., Miller R.L., Mena J., et al. // *Leuk. Lymphoma*. 2005. V. 46. P. 935–939.
65. Dudek A.Z., Yunis C., Harrison L.I., et al. // *Clin. Cancer Res*. 2007. V. 13. P. 7119–7125.
66. Schmidt J., Welsch T., Jäger D., et al. // *Br. J. Cancer*. 2007. V. 97. P. 598–604.
67. Karin M., Yamamoto Y., Wang Q.M. // *Nat. Rev. Drug*. 2004. V. 3. P. 17–26.

Molecular Basis of Mammalian Embryonic Stem Cell Pluripotency and Self-Renewal

S. P. Medvedev^{1,2}, A. I. Shevchenko^{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 28.07.2010

ABSTRACT Mammalian embryonic stem cells (ESC) have a number of specific properties that make them a unique object of fundamental and applied studies. In culture, ESC can remain in an infinitely undifferentiated state and differentiate into descendants of all three germ layers – ectoderm, endoderm, and mesoderm – that is, they can potentially produce more than 200 cell types comprising the body of an adult mammal. These properties of ESC are referred to as self-renewal and pluripotency. In this review, the basic signal pathways implicated in the maintenance of ESC pluripotency are considered. The major genes comprising a subsystem of “internal regulators of pluripotency,” their protein products and regulators, are characterized, and interaction with other factors is described as well. The role of epigenetic mechanisms and microRNAs in the system of ESC self-renewal and pluripotency, as well as the relationship between pluripotency and X-chromosome inactivation in female mammals, is discussed.

KEYWORDS: embryonic stem cells, pluripotency, transcription factors, X-chromosome inactivation

ABBREVIATIONS: ESC – embryonic stem cell, mESC – murine ESC, hESC – human ESC, TSC – trophoblast stem cell, iPSC – induced pluripotent stem cell, ICM – inner cell mass, HCNE – highly conserved non-coding element

MAMMALIAN EMBRYOGENESIS AND STEM CELLS. PLURIPOTENT CELLS *in vivo* AND *in vitro*

Development of any mammal begins from a single cell, called the zygote. Its successive division and differentiation of daughter cells eventually results in the formation of an organism composed of more than 200 cell types, each executing a specific function via the formation of tissues and organs in which different cell types interact with each other. These functional, biochemical, and physiological specificities of distinct cell types depend on the characteristic pattern of gene expression. All the diverse forms of RNA transcribed in the cell nucleus are called “transcriptome,” which (unlike genome) is distinctive of the cell type and differs in different cell types within the body. The cell transcriptome is determined by a tight interaction between the genetic and epigenetic systems. The former includes transcription factors, the proteins modulating expression of target genes at the transcription level. The latter includes proteins providing differential gene transcription via the changing and maintenance of the chromatin structure by a chemical modification of the DNA in promoter regions, particularly the 5'-methylation of cytosines in CpG islets, and methylation and acetylation of histones, amongst other modifications. Regulation of gene transcription also involves the microRNAs

(miRNAs) controlling gene expression at the posttranscriptional level [1].

Detached cell types, the so-called stem cells, are present at different stages of mammalian ontogenesis. These cells are characterized by a unique transcriptome and a complex of genetic and epigenetic components allowing self-renewal; i.e., sustaining the stem status for a short or long time (depending on the stem cell type) and maintaining the capability of differentiation into various secondary cell types [2]. Stem cells may be subdivided into several subtypes, depending on how wide the daughter cell type spectrum is [2].

Totipotent cells can differentiate into all the possible cell types that a mature mammalian body consists of. These cells are also implicated in the formation of extraembryonic tissues and organs, particularly placenta. In mammals, the only totipotent cells are the zygote and early cleavage stage blastomers.

Pluripotent cells can differentiate into descendants of all three germ layers: ectoderm, mesoderm, and endoderm. During ontogenesis, pluripotent cells can form a mature mammal body. However, they cannot produce extraembryonic tissues and organs. Pluripotent cells are in the inner cell mass and epiblast of mammal embryos, as well as their descendants, embryonic stem cells growing *in vitro*. Besides, embryonic germ cells that

originate from gamete precursors and epiblast stem cells isolated from postimplantation mouse embryos are also pluripotent [3–5]. In recent years, methods allowing to make mature differentiated cells pluripotent by the introduction of exogenous factors have been developed. These cell types are called induced pluripotent stem cells (iPSCs). To date, iPSC lines of human, mouse, and some other mammals have been isolated [6–8].

Multipotent cells are region-specific stem cells (those of a mature body, which are located among differentiated cells of various tissues) that can differentiate into several cell types. Examples of such cells are hematopoietic stem cells differentiating into blood cells.

Unipotent cells are precursors of only one cell type. For instance, spermatogonial stem cells are precursors of male gametes.

As already mentioned, mammalian embryo blastomeres are totipotent up to an eight-cell stage. Each blastomer of an 8-cell embryo possesses the potential to form all possible cell types, including the cells comprising extraembryonic organs. This property of mammalian blastomeres was confirmed in experiments on the construction of aggregation chimeras. However, a transition to the 16-cell stage (morula) is associated with the first differentiation event. Each of the eight blastomeres undergoes either a symmetrical division to form two polar outer cells of morula or an asymmetrical division to form a polar outer cell and a nonpolar, relatively small, cell localized within the embryo. Thus, the fourth division of zygote produces an embryo composed of two first-specialized cell groups: the outer polar and the inner nonpolar cells [9]. Following the next two-cell divisions, the outer polar cells of morula form the outer cell layer called trophoblast. It is required for the implantation of the embryo into the uterine endometrium and formation of placenta. The trophoblastic cells form a sphere with the inner cavity, the blastocoel. The inner nonpolar cells of the morula are transformed into the embryo inner cell mass (ICM) located at one end of the blastocyst on the inner side of the trophoblastic layer. Preceding implantation into the uterus, ICM is divided into two layers: the epiblast (primitive embryonic ectoderm) and the hypoblast (primitive embryonic endoderm). Throughout the course of its development, the epiblast gives rise to the embryo body (all tissues and organs), whereas the hypoblast forms the yolk sac. Thus, the mammalian preimplantation embryo (blastocyst) is composed of three compartments: the trophoblast, the hypoblast, and the epiblast, two of which (trophoblast and hypoblast) produce extraembryonic tissues and organs and one that forms all the fetal tissues during further development [10]. The structure of preimplantation embryos is very similar in different mammalian species. The blastocysts of primates closely

resemble those of rodents, featuring a greater number of cells and morphology after implantation. However, despite their similar structures, the blastocysts of primates differ from those of rodents in the time of compartmentation (into the trophoblast, hypoblast, and epiblast) and are characterized by a longer preimplantation development (7–10 days in primates compared with four days in mice) [11].

All three compartments of preimplantation embryos (the epiblast, trophoblast, and hypoblast) are sources of stable stem cell lines called embryonic stem cells, trophoblast stem cells, and extraembryonic stem cells, respectively. Cultured stem cells keep most of the features characteristic of their precursors, such as the gene expression pattern, epigenetic features, and many biochemical and physiological parameters. The stem status and correspondence of cell lines to their precursors existing *in vivo* is best demonstrated by injecting the cells into a recipient blastocyst. After ESC injection, both TSC and extraembryonic ectodermal cells become involved in embryogenesis, contributing to the formation of the corresponding embryonic or extraembryonic tissues and organs. All three cell types isolated from preimplantation embryos require specific culture conditions, such as the medium composition and presence of growth factors, in order to maintain their stem status. Sometimes, co-cultivation is necessary with other cell types synthesizing the factors required for maintenance of the undifferentiated state of stem cells [12–18].

Embryonic stem cells that possess self-renewal and pluripotency features are of particular interest. Self-renewal implies the capability of infinite symmetrical division leading to the appearance of two pluripotent cells. As mentioned above, pluripotency is the capability of a stem cell to differentiate into cell derivatives of all three germ layers, but not extraembryonic tissues. Pluripotency is the characteristic that makes ESC a unique model for the investigation of the processes underlying cell differentiation and opens a broad avenue for ESC use in fundamental research into biological events at early stages of mammalian ontogenesis, including chromosome X inactivation, epigenetic genome changes, etc. Moreover, ESCs have practical use. Their unique features can be used in trials of novel drugs and studies of the toxicity of chemical substances. ESCs are a prominent source of biological material for cell replacement therapy in human diseases, although the issue of how safe they are remains unsettled and under study.

At the moment, the focus of studies of ESC self-renewal and pluripotency is on their molecular basis. To date, it is known that a complex system of cell-surface proteins, their molecular signaling pathways, and the

transcription factors initiating or modulating target gene transcription sustain the pluripotency of pre-implantation embryo cells and ESCs. Several signaling cascades/pathways are implicated in maintaining pluripotency. These pathways constitute an external pluripotency regulation subsystem. A considerable body of experimental data suggests that this subsystem is not conservative. In particular, some signaling molecules, such as LIF and BMP4, have opposite effects on the self-renewal of mouse and human ESCs [19–21].

In addition to the subsystem of external regulators of pluripotency, there is a subsystem of internal regulators which includes three transcription factors: OCT4, SOX2, and NANOG. These factors have a broad spectrum of target genes in murine and human ESCs, and their effect, unlike that of signaling pathways, is exclusively directed toward maintaining self-renewal and pluripotency in various mammals [22, 23].

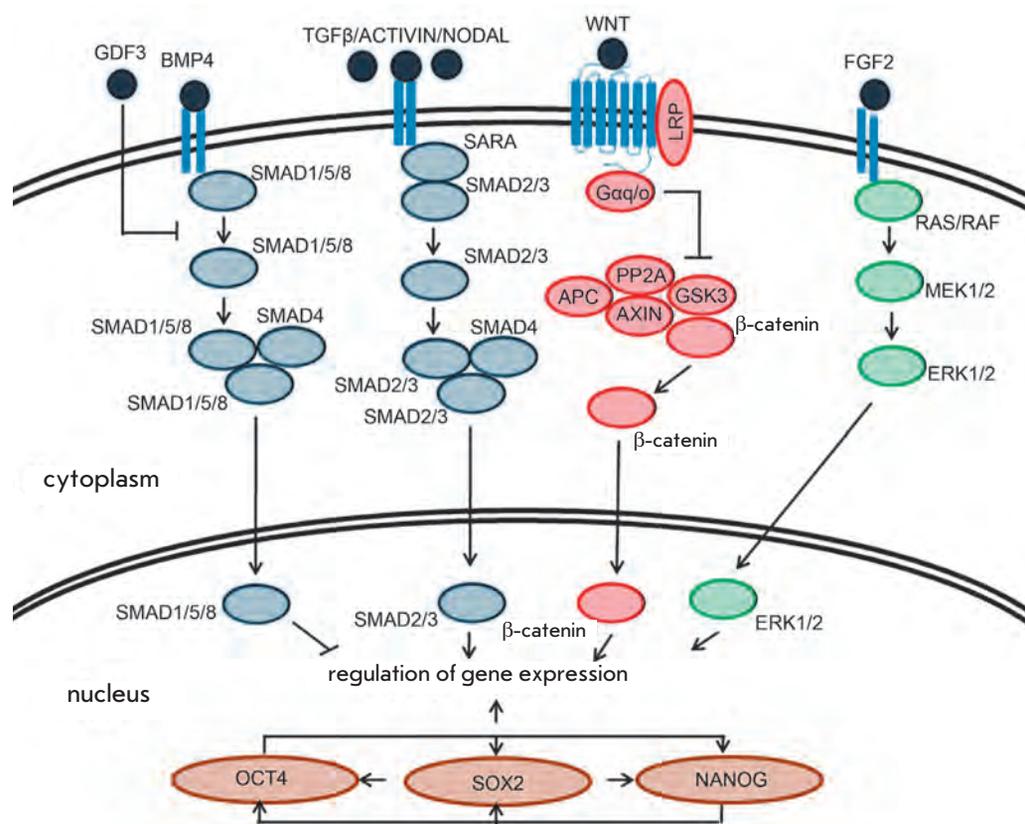
Recent data suggest a tight link between the two subsystems in ensuring a stable functioning of the entire system that sustains self-renewal and pluripotency [24] (Fig. 1). The subsystems have also proved to interact with proteins comprising the system of epigenetic regulation of gene expression (Polycomb and other proteins) and with miRNAs specific for ESCs or various differentiated cell descendants.

SIGNALING CASCADES TRIGGERED BY INTERLEUKIN LIF

Initially, murine ESC lines were obtained using either mitotically inactivated fetal fibroblasts forming a feeder layer or a conditioned medium in which teratocarcinoma cells or fetal fibroblasts were preliminarily grown for a given amount of time [12, 13]. Mouse ESC culturing without a feeder layer or in an unconditioned medium resulted in cell differentiation. The nature of the agent secreted by fetal fibroblasts and required for sustaining the undifferentiated state of ESCs was determined later: the agent was interleukin LIF (Leukemia Inhibitory Factor), a member of the IL-6 family [25]. The recombinant LIF protein is currently widely used to isolate and culture mouse ESCs, in order to avoid the use of both a feeder and conditioned medium [26, 27]. The LIF protein, via interaction with a heterodimeric receptor LIFR β -gp130 localized on the cell membrane, can activate three signaling cascades: JAK (tyrosine kinase Janus)-STAT3 (signal transducer and activator of transcription-3), the PI3K (phosphatidylinositol 3-kinase)-mediated pathway, and the MAPK (mitogen-activated protein kinase) cascade. Only JAK-STAT3 is activated solely by LIF, whereas the other two cascades are triggered by multiple molecular signals [24].

In the JAK-STAT3 cascade, LIF is the initial molecular signal that activates the transcription factor

Fig. 1. System of “external and internal regulators of pluripotency” in human ESC. Signal molecules and receptors on the cell plasmatic membrane are indicated by dark circles and blue rectangles, respectively. Gray, red, and green ovals indicate molecules involved in signal transmission from receptors to the cell’s nucleus and regulation of gene expression, together with the transcription factors OCT4, SOX2, and NANOG in the cell’s nucleus. Arrows represent activation and stoppers – repression as well and activity of SMAD1/5/8 which induces hESC differentiation.



STAT3. Its binding with the LIFR β -gp130 receptor causes the activation of phosphoprotein kinases. Both the Janus tyrosine kinase and immunoreactive phosphotyrosine kinase phosphorylate the tyrosine residues of the cytoplasmic part of the LIFR β -gp130 heterodimer. The transcription factor STAT3 binds to the phosphotyrosine residues of the activated LIFR β -gp130 heterodimer and undergoes subsequent phosphorylation and dimerization. Phosphorylated STAT3 dimers are transported into the cell nucleus, where they interact with the target genes [25, 28].

The important role of the LIF-STAT3 signaling pathway in sustaining the self-renewal of murine ESCs has been convincingly shown [29]. On the other hand, abundant data suggest a LIF-STAT3-independent mechanism sustaining ESC pluripotency in humans and other primates [21, 30].

Up until recently, not much data was available regarding the mechanisms of LIF implication in the maintenance of murine ESC pluripotency. Recent data provided by Niwa and colleagues [24] suggests the integration of LIF-triggered molecular cascades into the subsystem of internal regulators of pluripotency, whose components are the genes *Oct4*, *Sox2*, and *Nanog* (Fig. 1). This integration takes place via two pathways: JAK-STAT3 and PI3K-mediated. STAT3 activates the expression of KLF4 (Krüppel-like factor 4), which, in turn, is a positive regulator of the *Sox2* gene. In parallel, the PI3K signaling cascade activates a gene encoding another transcription factor, TBX3, which then activates the *Nanog* gene. Moreover, inhibitors of MAPK that mediate the third, LIF-activated signaling cascade cause the accumulation of the TBX protein in the cell's nucleus and the activation of the *Tbx3* and *Nanog* genes (Fig. 2). Moreover, LIF can indirectly (via MAPK) activate TBX3 transport (Fig. 2), but this pathway seems to be inhibited by some third factor.

TGF β /ACTIVIN/NODAL, BMP, AND BFGF (FGF2) SIGNALING PATHWAYS

The TGF β (transforming growth factor- β) superfamily includes a large number of signaling molecules (about 40 putative protein ligands of TGF β are found in humans). This superfamily may be subdivided into two large groups of activators:

1) TGF β /ACTIVIN/NODAL activating the transcription of SMAD2/3 (similar to mothers against decapentaplegic homologue) transcription factor genes via receptors of ALK4 (activin-like kinase 4), ALK5, and ALK7;

2) BMP (bone morphogenic protein)/GDF (growth differentiation factor) activating SMAD1/5 via the receptors of ALK1, ALK2, ALK3, and ALK6. Also, two proteins of the SMAD family are inhibitors of the same

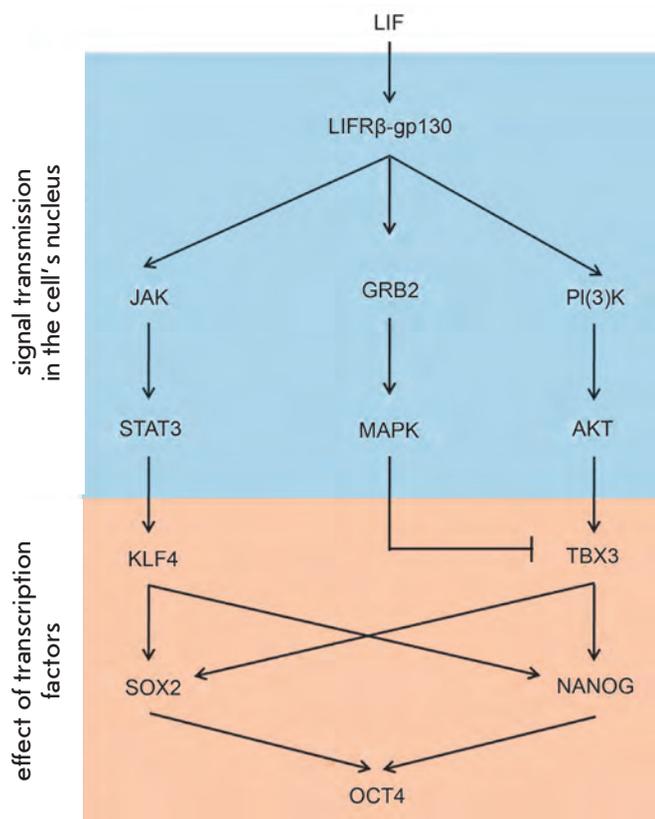


Fig. 2. Interplay between transcription factors and LIF-mediated signaling pathways in mouse embryonic stem cells [24].

family members: SMAD6 selectively inhibiting the SMAD1/5 and SMAD7 suppressing protein activators in both branches of the TGF β superfamily [31].

The TGF β /ACTIVIN/NODAL signaling pathway plays an essential role in sustaining human ESC (hESC) pluripotency. The ACTIVIN and NODAL proteins act via the same receptors to suppress hESC differentiation [32–34]. Human ESCs express both the ACVR1B and ACVR1B receptors and the TDGF1 (CRYPTO) coreceptor of NODAL, as well as NODAL itself [35]. In hESC cultures, NODAL and ACTIVIN can act in concert with other protein factors, such as bFGF (FGF2) and WNT, in the maintenance of hESC self-renewal [32, 34, 36, 37]. ACTIVIN can activate the production of bFGF, which is necessary for sustaining the self-renewal of hESC cells cultured in the absence of a serum [36]. Both ACTIVIN and NODAL can be replaced with TGF β in hESC cultures.

The protein factor NODAL plays many different roles during the early embryogenesis of mice. It is necessary for the normal development of embryo epiblast and the maintenance of human OCT4 and NANOG gene expression. A high level of activity of NODAL/

ACTIVIN/TGF β directed toward the activation of SMAD2/3B transcription factors (Fig. 1) is observed in undifferentiated hESCs. The transcription factors SMAD2 and SMAD3 are implicated in *NANOG* transcription activation. Undifferentiated hESCs exhibit a high level of phosphorylated (active) forms of SMAD2 and SMAD3. Inhibition of SMAD2/3 phosphorylation by specific inhibitors results in a lowered *OCT4* and *NANOG* gene expression level [32]. During differentiation, the activity of SMAD2 and SMAD3 decreases with an increase in the level of SMAD1/5/8 activating BMP (Fig. 1). SMAD1/5/8 proteins suppress *NANOG* gene transcription. Moreover, SMAD2 and SMAD3, as well as the signaling pathway triggered by FGF2 (bFGF), inhibit the expression of BMP4, thus blocking the spontaneous differentiation of hESCs [37].

The GDF protein, a member of the TGF β superfamily, not only inhibits BMP, but also potentiates the effect of NODAL (Fig. 1). The *GDF* gene is expressed in pluripotent cells of humans and mice, through implication in the suppression of the BMP-mediated differentiation. Eventually, GDF influences ESC self-renewal by establishing a balance between SMAD2/3 and SMAD1/5/8 [38, 39] (Fig. 1).

The protein factors of the BMP group have the opposite effect on the self-renewal of murine ESCs (mESCs). BMP4 suppresses neural differentiation and, taken in combination with LIF, sustains the symmetrical self-renewal of mESCs in the absence of both the feeder layer and serum. Murine BMP4 activates the expression of the transcription factor SMAD4 that, in turn, activates the transcription of genes belonging to the *Id* (inhibitor of differentiation) family. It is likely that the expression of the *Id1*, *Id2*, and *Id3* genes in ESCs suppresses the effects of the early expression of genes encoding proneuronal transcription factors (bHLH), particularly *Mash* members, thus preventing differentiation [40]. However, BMP4 in the absence of LIF can have an opposite activating effect on SMAD family factors (SMAD1/5/8), which inhibit the expression of *Id* family genes [40].

Thus, the signaling pathways and cascades triggered by BMP and LIF-JAK-STAT3 have opposite effects on hESC and mESC self-renewal. This is not a solitary example of the species-specific effect of signal molecules. For instance, bFGF is critical in sustaining the undifferentiated state of hESC, but it causes the differentiation of mESCs [41, 42]. Human ESCs express both the receptor and bFGF itself [43, 44]. The effect of bFGF is mediated by tyrosine kinase receptors (ERK1 and ERK2), and inhibition of this signaling pathway results in hESC differentiation (Fig. 1). It is worth noting, however, that the mechanism of bFGF action on hESCs has yet to be understood in detail.

WNT SIGNALING PATHWAY

Recent studies show that the signaling pathway, of which WNT is a key element, is involved in the short-term maintenance of the pluripotent state of both mESCs and hESCs [45]. The inhibition of glycogen synthase kinase 3 (GSK-3) results in the activation of the WNT signaling pathway and leads to the accumulation of β -catenin in the cells' nuclei and the activation of a series of target genes. However, the expression of the *Oct4*, *Rex1*, and *Nanog* genes, which is specific to pluripotent cells, is retained even in the absence of LIF [46] (Fig. 1). Nevertheless, it was shown later that retaining hESC self-renewal for longer requires TGF β and bFGF, and that the action of WNT is restricted to intensifying proliferation [47, 48].

TRANSCRIPTION FACTORS OCT4, NANOG, AND SOX2

The transcription factors OCT4, NANOG, and SOX2 play a central role in the subsystem of internal regulators of pluripotency. Unlike a system of species-specific signaling pathways, the functions of these transcription factors essentially intersect in sustaining the pluripotency of hESCs and mESCs. Both the genomic organization of the *Oct4*, *Sox2*, and *Nanog* genes and the domain structure of the OCT4, NANOG, and SOX2 proteins are conserved, and the spectra of the target genes of these factors are similar in mESCs and hESCs. Both in mice and humans, OCT4, NANOG, and SOX2 act collectively to form an integrated system regulating gene transcription, including the autoregulation of their own genes.

TRANSCRIPTION FACTOR OCT4

The OCT4 protein is a POU (Pit, Oct, Unc) class V family of transcription factors. The POU-domain is a bipartite domain composed of two subunits separated by a nonconserved region whose length varies. The N-terminal subunit (75 aa) is known as the POU-specific (POUs) domain, while the C-terminal subunit (60 aa) is a homeobox domain [49–51]. OCT4 expression is characteristic of early embryogenesis in humans, mice, and other mammals [52–56]. In mice, its expression is traced above the two-cell stage and is restricted to the pluripotent cells of ICM and epiblast. In postimplantation embryos, OCT4 is only expressed in germ cells. In human blastocyst, OCT4 is found not only in ICM, but also in the trophoblast [57]. The role of OCT4 in embryogenesis was demonstrated in experiments with *Oct4* knockout mice (the gene *Oct4* is also known as *Pou5f1*, *Oct3*, *Oct3/4*, *OTF3*, and *OTF4*). Homozygous (*Oct4*^{-/-}) mouse embryos died at the stage of implantation due to their inability to form ICM (composed of pluripotent cells), while the development of trophoblast was normal [53]. Later, Niwa and colleagues demonstrated the effect of

the *Oct4* expression level on mESC self-renewal [58]. A variation in the *Oct4* transcription level by 50% results in the differentiation of mESCs into either trophoblastic or endodermal derivatives [58]. Moreover, OCT4 activity is necessary for the normal life of murine germ cells [59]. Inhibition of *Oct4* gene expression in murine and human ESCs by means of RNA-interference also results in their differentiation into different derivatives expressing trophoblastic and endodermal markers, such as *Cdx2*, *Gata6*, and *AFP* [60, 61]. Studies of the *Oct4* genomic structure, the structure of the regulatory region, and chromosomal localization in humans and other mammals have shown a highly conserved molecular organization of this gene. Besides, it was shown that in a series of mammal species the *Oct4* gene maps onto a similar syntenic group [62–65]. In mice and other animals, *Oct4* is composed of five exons and localized in the main histocompatibility complex (MHC) region [62, 63, 65, 66]. The human *OCT4* gene is also localized within the MHC region; however, three variants of its alternative splicing, *OCT4A*, *OCT4B*, and *OCT4B1*, were demonstrated [62, 67]. The proteins encoded by these isoforms are involved to different levels in the maintenance of pluripotency. OCT4B, which is largely localized in the cytoplasm, cannot sustain the pluripotency of ESCs [68]. In humans, OCT4A is expressed at the stage of compact morula and blastocyst, whereas OCT4B is expressed in all embryonic cells above the four-cell stage. Moreover, OCT4B is not expressed in hESCs [69]. Unlike the *Oct4* of other animals, human *OCT4* has four exons. The second exon is subdivided into four subexons (2a, 2b, 2c, and 2d) that differently combine with each other in three different variants of transcript splicing. The variant *OCT4A* is composed of exons 1, 2b, 2d, 3, and 4; *OCT4B* – 2a, 2b, 2d, 3, and 4; *OCT4B1* – 2a, 2b, 2c, 3, and 4. Thus, the human *OCT4* gene transcripts are identical in 3'-area (exons 3 and 4), and the difference only concerns exons 1 and 2. The *OCT4A* transcript encodes a protein with a length of 360 aa. The size of OCT4B is 265 aa; however, in some parts of the population this protein type is not translated because of the polymorphism of its start codon (AU/GG) [62, 70]. Another two isoforms, OCT4B-190 and OCT4B-164 composed of 190 and 164 aminoacid residues, respectively, can be translated from the recently found alternative ribosome-binding site localized in the area of the *OCT4B* transcript subexons 2a-b. The protein product of *OCT4B1* was not found, maybe because of the *UGA* stop codon in the subexon 2c that is absent in *OCT4B* [67]. It was shown earlier that the protein encoded by the *OCT4A* isoform is localized in the cell nucleus and involved in the regulation of gene transcription, while the OCT4B protein is localized in the cytoplasm and cannot maintain pluripotency [57,

68]. Later, it was found that both *OCT4B* products, OCT4B-190 and OCT4B-164, are dispersed in both the nucleus and cytoplasm [70]. In hESCs, the level of OCT4B-190 considerably increases in response to stress conditions and can inhibit apoptosis. No expression of OCT4B-265 and OCT4B-164 was found in hESCs [70].

TRANSCRIPTION FACTOR NANOG

The transcription factor NANOG is a homeodomain protein that shares a very high degree of homology and structural resemblance with NK family proteins [71]. The expression of the *NANOG* gene is characteristic of preimplantation embryo pluripotent cells (ICM and epiblast), as well as murine and human ESCs [72–74]. *Nanog* overexpression can lead to the retention of mESC pluripotency, even in the absence of interleukin LIF [72]. The mutation analysis of *Nanog* has shown how necessary this gene expression is for normal embryonic epiblast development and maintenance of mESC self-renewal. ESCs with the genotype *Nanog*^{-/-} differentiate into extraembryonic derivatives [73]. Suppression of *NANOG* expression in hESCs results in differentiation accompanied by elevated expression of endodermal (*GATA4*, *GATA6*, *LAMININ B1*, and *AFP*) and trophectodermal (*CDX2*, *GATA2*, *hCG- α* , and *hCG- β*) markers.

The murine *NANOG* molecule in mice has several functional elements involved in transactivation. These include the homeodomain, which occupies a central position in the protein, and three C-terminal elements: CD1 (C-terminal domain 1), CD2, and the tryptophan repeat (W-repeat) localized between CD1 and CD2. Homeodomain, CD2, and the W-repeat chiefly contribute to the transactivation of murine *NANOG* [75, 76]. Recent studies have shown that the murine *NANOG* W-repeat provides transactivation; it is involved in *NANOG* protein dimerization that is necessary for LIF-independent maintenance of mESC pluripotency [77, 78]. In all likelihood, tryptophan residues are the most essential. Their replacement by alanines results in a substantial decrease in the murine *NANOG* transactivation capability, whereas the replacement of any other aminoacid residue in the W-repeat monomer has no significant effect on this activity [76]. Another important transactivation element of *NANOG* is CD2 [79]. This element is required for the *NANOG*-mediated self-renewal of mESCs. CD2 domain activity depends on aromatic residues (phenylalanine and tyrosine). Replacement of these aminoacids leads to a decrease in the CD2 and entire *NANOG* activity and, as a result, leads to ESC differentiation [79]. Nevertheless, a comparison of different mammalian CD2 sequences has shown that the majority of aromatic aminoacid residues are not conserved to a high level [80].

TRANSCRIPTION FACTOR SOX2

Transcription factor SOX2 (SRY-related HMG box) contains the DNA-binding HMG (high mobility group)-domain. Expression of SOX2, as well as OCT4, is characteristic of the ICM, epiblast, and germ cells of mice embryos [81].

Homozygous SOX2 mutant embryos die at the stage of implantation because of the epiblast hypoplasia. It is impossible to obtain stable ESC lines from mutant embryos, but both TSC and extraembryonic endodermal lines are easily prepared [81]. In addition, a normal expression of *Sox2* gene is necessary for sustaining the self-renewal of murine and human ESCs [82, 83]. Both suppression and overexpression of SOX2 cause trophodermal differentiation of hESCs [83].

GENE TARGETS FOR OCT4, NANOG, AND SOX2 TRANSCRIPTION FACTORS

OCT4, NANOG, and SOX2 are the proteins whose expression is necessary for sustaining murine and human ESC pluripotency under standard culturing conditions. It is evident that their function as transcription factors determines general ESC features. At present, much data exist concerning the spectrum of target genes for these proteins.

The transcription factor OCT4 can serve as the activator and repressor of many target genes, such as *Fgf4*, *Opn*, *Utf1*, and the genes encoding human chorionic gonadotropin α - and β -subunits, most of them participating, in one way or another, in early embryo development [84–88]. The regulatory forms of OCT4 are mono-, homo-, and heterodimer, depending on the target gene [89]. For instance, the OCT4 homodimer regulates the transcription of *Opn*. The enhancer element of *Opn* contains a palindromic sequence called PORE (palindromic-oct-regulatory-element), which contains the octameric ATGCAAAT motif (octameric site for the docking of the OCT monomer) and the ATTTG sequence separated from the octameric motif by two nucleotides. The SOX2 docking site is localized near PORE; however, it exerts a suppressing effect on *Opn* transcription [84].

The OCT4 and SOX2 factors can act in conjunction. They positively regulate *Fgf4* and *Utf1* genes, and their docking sites are localized in the enhancers localized in the 3'-nontranslated regions of these genes [85, 86, 90]. Aside from PORE, another element exists which interacts with OCT monomers and dimers; it was named MORE (more of PORE, ATGCATATGCAT) [91, 92]. MORE and PORE elements are substantially different in the relative localization of the POU-specific domain and POU-homeodomain. Despite an initially artificial synthesis of MORE, similar elements were found in the regulatory regions of natural genes [91–93]. OCT1 and

OCT4 transcription factors can influence the target gene's transcription via MORE elements in response to genotoxic and oxidative stresses [93].

The development of high-performance methods for the distribution analysis of transcription factors at the whole genome level has substantially broadened our knowledge of the spectrum of OCT4, NANOG, and SOX2 target genes. Two research groups have localized OCT4, NANOG, and SOX2 in the human and mouse ESC genomes using ChIP-on-Chip and ChIP-PET (chromatin immunoprecipitation – paired-end ditag) methods [22, 23].

The distribution of binding sites for the OCT4, NANOG, and SOX2 transcription factors in human H9 ESCs was analyzed by Boyer and colleagues [22]. The genes encoding transcription factors and the components of signaling pathways which, as determined earlier, are implicated in early embryogenesis, cell differentiation, organogenesis, and maintenance of ESC's self-renewal and pluripotency in culture are a major part among the genes associated with these transcription factors. Particularly, they are OCT4, SOX2, NANOG, LEFTY2/EBAF, CDX2, HAND1, DPPA4, GJA1/CONNEXIN43, FOXO1A, CRIPTO/TDGF1, and ZIC3 [94–102]. Then, the distribution of binding sites for the NANOG and SOX2 factors was determined in the same way. The OCT4, SOX2, and NANOG transcription factors jointly regulate 353 genes in hESCs; they can act as transcription activators or repressors [22]. The transcription factor genes, such as OCT4, NANOG, and SOX2, as well as genes encoding components of the signaling pathways implicated in the sustenance of self-renewal, such as TGF β (TDGF1, LEFTY2/EBAF) and WNT (DKK1, FRAT2), make up a large portion of the genes positively regulated by OCT4, NANOG, and SOX2 [22]. Also, transcription factor genes, such as *Rcor2*, *Esrrb*, and *Phc1*, were revealed in mESCs among the genes positively regulated by the OCT4 and NANOG factors [23]. In addition, *Pou5f1* (encoding the OCT4 transcription factor) and *Sox2* are on the list of the genes positively regulated by NANOG in mESCs. Early experimental data also suggest the involvement of the OCT4 and SOX2 factors in the regulation of the *Pou5f1*, *Sox2*, and *Nanog* genes [97, 99, 100, 103]. It appears that autoregulation is the general property of the system maintaining pluripotency in mice and humans. The abundance of genes encoding transcription factors (such as REST, SKIL, HESX1, and STAT3) among the positively regulated genes suggests that OCT4, NANOG, and SOX2 are only the tip of an iceberg, hiding a far more complex system of transcription regulation in ESCs. Many genes encoding the transcription factors (for instance, ESX1I, HOXB1, MEIS1, PAX6, LHX5, LBX1, MYF5, ONECUT1) involved in differentiation

during embryo development are found among the negatively regulated genes [22]. It is very likely that the OCT4, NANOG, and SOX2 factors participate in the repression of these genes both *in vivo* and *in vitro*.

Data published in two studies seem to indicate that the binding sites for the OCT4, NANOG, and SOX2 transcription factors are associated with the genes encoding miRNAs [22, 23]. In hESCs, the docking sites for OCT4, NANOG, and SOX2 are found in the promoters of 14 miRNA genes; they are all present in the *mir-137* and *mir-301* promoters. In mESCs, the binding sites for NANOG are localized within a distance of 6 kbp from four miRNA genes: *mir-296*, *mir-302*, *mir-124a*, and *mir-9-2*. Moreover, no other genes are found in close vicinity to the NANOG site in the case of the *mir-296*, *mir-124a*, and *mir-9-2* genes. The docking site for NANOG is found within a 30-kbp region for the *mir-135* gene, while OCT4 binds in close vicinity to the NANOG site near the *mir-296* and *mir-302* genes.

A comparison of data on the revelation of target genes for the OCT4 and NANOG transcription factors in mouse and human ESCs demonstrates that only a minor portion of these genes are simultaneously found in the genomes of both species. So, only 9.1% OCT4-associated genes and 13% NANOG-associated genes overlap in these two species. This may be evidence of an essential difference in the composition of the gene sets controlled by OCT4 and NANOG. However, a group of 32 genes is regulated by OCT4 and NANOG in both mouse and human, 18 of them encoding transcription factors, including OCT4, SOX2, and NANOG, suggesting the importance of these genes in sustaining pluripotency [23]. The difference in the compositions of the target genes found in these two studies may have a technical origin, because two different methods were employed in the works (ChIP-on-Chip and ChIP-PET, respectively) to obtain data and process them.

Thus, the OCT4, NANOG, and SOX2 transcription factors are in the center of a broad regulatory net including transcription factors (whose genes are positively or negatively regulated), components of signaling pathways, and miRNAs. The stable functioning of this net is evidently necessary for sustaining self-renewal and pluripotency in both human and murine ESCs.

PROTEIN-PROTEIN INTERACTIONS IN THE PLURIPOTENCY MAINTENANCE SYSTEM IN EMBRYONIC STEM CELLS

Recent studies provide much evidence pointing to the fact that the transcription factors involved in the maintenance of ESC self-renewal not only exert joint control on target gene transcription, but they are also in physical contact with each other [104–108]. For instance, NANOG can suppress BMP-initiated mesodermal differentiation in mESCs via binding with SMAD1 and

physically interact with the SALL4 transcription factor to execute joint positive regulation of the *Nanog* and *Sall4* gene enhancers [104, 106]. However, these were only particular examples of protein-protein interactions of NANOG directed toward the maintenance of ESC self-renewal. A wider analysis of NANOG interactions on mESCs has demonstrated a broad spectrum of interactions. The set of interacting proteins includes not only transcription factors, but also the proteins implicated in chromatin structure regulation [105]. Wang and colleagues [105] applied a method based on the expression of the NANOG protein containing FLAG-epitope and a short amino acid sequence serving as the substrate for BirA, the *Escherichia coli* biotin ligase. A NANOG containing two additional epitopes was expressed (at the level of about 20% of the endogenous level) in mESCs that also expressed BirA, and the cells retained self-renewal and pluripotency. NANOG complexes were isolated from the nuclei extract either by using streptavidin-agarose or by immunoprecipitation with antibodies against FLAG, followed by purification on streptavidin-agarose. The proteins interacting with NANOG were identified using mass-spectrometry (whole-lane liquid chromatography–tandem mass spectrometry, LC–MS/MS).

Several regularities can be detected from the data obtained in this experiment. In the first, the group of proteins interacting with NANOG is much enriched in the factors required for the normal viability and differentiation of ICM cells (pluripotent embryo compartment) in mouse blastocyst. Besides, according to data in the literature, more than 80% of NANOG-interacting proteins, such as OCT4, DAX1, NAC1, ZFP281, and SALL4, are required for the acquisition and maintenance of general ESC properties [105].

Secondly, most of the revealed proteins possess a similar expression pattern, which is typical of pluripotent cells, and their expression is suppressed under ESC differentiation, likely suggesting their involvement in similar processes or in the same regulatory system [105].

Thirdly, in human and murine ESCs a large portion of genes encoding the proteins interacting with NANOG represents putative targets for OCT4 and NANOG [22, 23, 63]. The system sustaining ESC self-renewal and pluripotency in mice and humans undergoes autoregulation with positive and negative feedbacks, a feature that is necessary for both the stability of the undifferentiated state and the realization of a strict differentiation program in a distinct direction [97, 99, 100, 103]. It is likely that the maintenance of pluripotency or direction of cell differentiation depends on the stoichiometric ratio between the molecules of different factors. This hypothesis is supported by experimental data on

ESC multidirectional differentiation depending on the transcription levels of the *Oct4*, *Nanog*, and *Sox2* genes [58, 83, 109].

Fourthly, NANOG can interact via its partners with multiple proteins involved in the epigenetic regulation of gene expression. NANOG – via interaction with OCT4, DAX1, NAC1, ZFP281, or SALL1/4 – can form complexes with components of the NuRD complex (P66B and HDAC2) possessing histone acetylase activity, Polycomb proteins (YY1, RNF2, and RYBP), and components of the SWI/SNF (BAF155) chromatin remodeling complex [105]. The interaction of NAC1 and SALL1 with histone deacetylases has been experimentally proved [110, 111].

Complexes formed by NANOG and other proteins are apparently functional, rather than occasional short-live intermediates observed in a single experiment. For instance, the complex formed by NAC1, ZFP281, and NANOG interacts with the *Gata6* promoter (the marker of endodermal differentiation) to inhibit it.

Interaction between OCT4 and other protein molecules was also reported in a series of works [85, 112–114]. Particularly, these are interactions between OCT4 and transcription factors with the formation of heterodimers that are necessary for regulating target gene transcription. In addition, proteins interacting with OCT4 and inhibiting this factor were found using the yeast two-hybrid system and co-immune precipitation [115]. SUMO-ligase PIASy, as well as PIAS1, and PIAS3 proteins belonging to the same family interact with OCT4. Despite PIASy being a known SUMO-ligase and the OCT4 protein containing sumoylation sites, the inhibitory effect of the ligase was experimentally proved to be independent of this activity. PIASy exerts the inhibitory effect on the transactivating capability of OCT4, acting as a monomer, homo-, or heterodimer, whereas PIAS1 and PIAS3 do not exhibit this action. Also, the PIASy, PIAS1, and PIAS3 proteins induce the relocation of OCT4, pushing it to the nucleus's periphery [115].

Two research groups recently carried out a study on the patterns of protein-protein interactions of the OCT4 factor, as well as its known partners: SALL4, TCF2L1, DAX1, and ESRRB in mESCs [107, 108]. The authors employed an approach based on the expression of a chimeric protein (OCT4 in this case) containing known epitopes, which are necessary for the isolation of its complexes with other proteins from cellular or nuclear extracts. Expression of *Oct4* under its natural promoter was used in one of these works [107]. Like in the work by Wang and colleagues [105], components of the NuRD, SWI/SNF, and PRC1 complexes involved in the regulation of the chromatin structure were found among the proteins interacting with

OCT4 [107, 108]. Also, some enzymes participating in the epigenetic regulation of gene expression, such as MYST2 (histone acetyltransferase H4), DNMT3A (*de novo* DNA-methyltransferase), and some other proteins, were found in OCT4 complexes. Some proteins involved in the posttranslational modification of OCT4 itself, particularly OTG (enzyme attaching O-bound N-acetyl glucosamine), modifying OCT4 in hESCs were also found [107, 108, 116]. The transcription factors necessary for the maintenance of ESC self-renewal, such as KLF4, SOX2, SALL4, and ZFP281, are found among the partners of OCT4. Analysis of the functions of OCT4 partners using information from databases has shown that they are mostly involved in early development and cell differentiation and that their knockout induces death during the early development of the embryo [107]. All genes determined as OCT4 partners in mESCs have human homologues. Moreover, the amino acid sequences of the proteins encoded by these genes show a very high, more than 94%, homology between mice and humans (the average level is 77%). These genes are closely associated with the development of inherited diseases in humans (mostly with developmental disorders) and cancer [107].

Thus, gene regulation directed toward the maintenance of self-renewal and pluripotency appears more complex than was earlier thought. Not just the suppression or activation of target genes by distinct protein molecules of the transcription factors is involved in the regulation. There are molecular complexes implicated in the regulation which contain not only transcription factors, but also proteins remodeling the chromatin structure. The quantitative ratio of the molecules in each separate cell can directly or indirectly influence its self-renewal and direction of differentiation.

EPIGENETIC REGULATION OF CELL PLURIPOTENCY

ESCs possess a virtually unlimited potential for self-renewal and differentiation into a very broad spectrum of cell types. Global changes in morphology, physiology, division rate, and other parameters occur during cell differentiation. These changes are caused and accompanied by a global change in the gene expression pattern. Currently, gene expression is known as a process that is strictly regulated at the epigenetic level. Epigenetic regulation includes covalent modification of histones (nucleosome-forming proteins) and DNA methylation in gene promoter regions. Numerous types of histone modifications, such as acetylation, methylation, phosphorylation, ubiquitylation, etc., are known, consisting in the attachment of chemical groups and regulatory peptides (ubiquitin, SUMO) to the amino acid residues of histones. Histone modifications alter the physical properties of nucleosomes, thus making chro-

matin more or less accessible to the factors providing gene transcription. The modifications associated with active chromatin and actively transcribed genes and, alternatively, the modifications associated with inactive chromatin and often associated with transcription suppression have been distinguished. In particular, the acetylated forms of the histones H3 and H4 and histone H3 trimethylated at position K4 (H3K4me3) (K is lysine, according to the single-letter nomenclature) are “active” modifications. Alternatively, histone H3 di- and trimethylation at position K9 (H3K9me2 and H3K9me3) and histone H3 di- and trimethylation at position K27 (H3K27me2 and H3K27me3) are “inactive” chromatin modifications. Moreover, inactive chromatin is characterized by histone deacetylation.

In mammals, DNA methylation touches cytosines within the CpG-dinucleotide islets. These islets are often observed in promotor regions, and their hypermethylation is commonly associated with transcription suppression.

Studies on the localization of modified histones in the genomes of embryonic stem cells demonstrate that the distribution of active and inactive modifications is rather unusual and specific to pluripotent cells. The presence of bivalent domains simultaneously containing the labels of active and inactive chromatin (H3K4me3 and H3K27me3) was discovered in the mESC genome [117]. A distribution analysis of histone H3 active and inactive modifications has been carried out for 56 regions enriched with highly conserved noncoding element (HCNE) sequences. HCNE-enriched regions generally not contain many genes: however, they do encompass a relatively large number of genes encoding the transcriptional factors implicated in the regulation of development. For instance, all four clusters of the HOX gene family are localized in these regions. Within HCNE, Bernstein and colleagues [117] have demonstrated the presence of 343 H3K4me3-enriched regions averaging 3.4 kb; 63% of them are co-localized with the transcription initiation points of known genes. In addition to this, 192 regions enriched with H3K27me3 were found within HCNE sequences with a maximum length of 18 kb in the gene clusters of the HOX family. A comparison of data on the H3K4me3 and H3K27me3 distributions has revealed domains containing both modifications. Nine similar domains were found in HOX clusters; 95, in other HCNE-enriched regions; whereas only five bivalent domains were identified in control loci [117].

Most bivalent domains in HCNE-enriched regions are associated with the transcription initiation points of the genes encoding transcription factors. These genes are members of the SOX, FOX, PAX, IRX, and POU families, whose distinct members play important roles in cell differentiation during development. However, 26

bivalent domains were found beyond the transcription initiation points; they are also associated with the genes implicated in development. For instance, bivalent domains are found in the 3'-regions of the *Npas3*, *Meis2*, *Pax2*, and *Wnt8b* genes. Bivalent domains are present in the *Fgf8* and *Prok1* genes that do not encode transcription factors but participate in the development of the nervous system [117].

An extremely small number of bivalent domains were found in the differentiated cells studied by the authors, such as embryonic fibroblasts, primary lung fibroblasts, C2C12 myoblasts and Neuro2a neuroblastoma cells. At the same time, extended regions enriched separately with H3K4me3 or H3K27me3 were found. In differentiated cells, the overwhelming majority of the domains among the bivalent ones revealed in mESCs contained extended regions represented by only one modification, depending on cell type [117].

The genes, whose transcription start points are associated with bivalent domains, are characterized by a low transcription level, despite the presence of H3K4me3, the active chromatin label, thus suggesting a “predomination” of H3K27me3 over H3K4me3. The authors studied the distribution of the modifications in the bivalent domains associated with gene transcription start points during a directed ESC differentiation into neural cell progenitors. The genes with a high transcription level after the differentiation lost the H3K27me3 modification, while the genes transcribed at a low level enhanced enrichment in H3K4me3, in line with the preservation of H3K27me3, and nontranscribed genes preserved H3K27me3 and lost H3K4me3 [117].

Thus, a specific system of epigenetic gene regulation exists in the ESCs involved in cell differentiation. Apparently, it enables pluripotent cells to keep distinct genes in the “low start” state and rapidly start their transcription depending on differentiation programs and type of the formed differentiated cells.

The molecular mechanisms underlying the epigenetic regulation of gene expression in ESCs are being studied in detail. Polycomb group proteins are one of the main actors in the epigenetic regulatory system in embryo development in organisms and ESCs [118]. Numerous biochemical and genetic studies show that Polycomb proteins form two independent complexes: PRC1 (Polycomb repressive complex 1) and PRC2. The main components of these complexes are highly evolutionarily conserved and necessary for a normal embryonic development of various organisms, from drosophila to mouse and humans [118–122].

Boyer and colleagues have analyzed the localization of the main components of PRC1 (PHC1, RNF2) and PRC2 (SUZ12, EED) complexes in the regions flank-

ing the transcription start points (from -8 to +1 kb) of about 16,000 genes in mESCs [123]. The authors have found that PRC1 and PRC2 proteins are mostly localized within 1 kb of the transcription start point of 512 genes, thus suggesting their involvement in the regulation of these genes. Earlier, PRC2 was shown to be responsible for the establishment of H3K27me₃, the label of inactive chromatin. The study has demonstrated that H3K27me₃ enrichment takes place in all 512 genes whose 5'-regions can carry PRC1 and PRC2 proteins. The genes encoding homeodomain-containing transcription factors represent a significant portion of them. Previously known targets of Polycomb proteins – factors of the HOX family and other transcription factors belonging to DLX, IRX, LHX, POU, PAX and SIX families – belong to this group. A common feature of the factors belonging to these families is their implication in cell differentiation and regulation in embryo development and organogenesis. Genes whose products are implicated in the regulation of development despite the absence of a homeodomain were also revealed. These are proteins of FOX, SOX, GATA, and TBX families. In cells bearing homozygous EED mutation (one of the PRC2 components), a high expression level of genes belonging to various families is observed, although these genes are silenced in normal ESCs [123].

Thus, multiple genes implicated in differentiation and embryo development in mESCs are targets for PRC. Hence, PRC1 and PRC2 can play a substantial role in the maintenance of ESC self-renewal and pluripotency.

Similar results were obtained in a study on the localization of the SUZ12 protein (another PRC2 component) in hESCs [124]. The authors also succeeded in proving that PRC2 components are localized in the vicinity of the transcription start points of the genes implicated in many processes associated with cell differentiation during embryo development. Besides, in hESCs the SUZ12 protein is localized in the promoters of genes encoding the protein components of the signaling pathways involved in gastrulation, differentiation during ontogenesis, as well as in the self-renewal and differentiation of ESCs in culture. Among these genes, there are components of signaling pathways triggered by TGFβ, BMP, WNT, and FGF. SUZ12 is localized in distinct promotor regions, as well as in extended sites encompassing several genes belonging to one signaling pathway, such as WNT (WNT1, WNT2, WNT6) and TGFβ (BMP2, GDF6). Earlier, these signaling pathways were proven to participate in the regulation of self-renewal and pluripotency in murine and human ESCs [46]. Hence, PRC2 can directly influence ESC properties by regulating the components of these signaling pathways. However, the later data suggest maintenance of pluripotency in mESCs bearing the mutant EED [125]. ESCs bearing the homozygous EED mutation express

OCT4 and NANOG and form chimeras. Although this work was a failure in its attempt to obtain adult chimeric mice, a large extent of chimerization was observed in 12.5 day-old embryos. An attempt to obtain a culture of embryonic fibroblast descendants of the mutant ESCs failed, likely due to the low viability of differentiated mutant cells. Chimeric embryos (with a high percentage of chimerization) also had several defects, such as hypertrophy of allantois and underdeveloped neuroectoderm and embryonic mesoderm, which possibly caused embryo death at the stage of about 10.5 days after fertilization [125].

A high frequency of SUZ12 co-localization with the already-mentioned HCNE (bivalent chromatin domains are localized in HCNE-enriched regions) is an interesting regularity described by Lee and colleagues [124]. Approximately 8% of the 14,000 HCNEs discovered earlier [126] are enriched in SUZ12, and the more evolutionary conserved the element, the higher the level of enrichment [124].

Besides, one-third of the genes regulating development and binding to PRC2 are also under the control of at least one of the three transcription factors, OCT4, SOX2, or NANOG, which are the key factors in the system of murine and human ESC pluripotency maintenance [22, 124]. These genes (such as *ESX1L*, *ONECUT1*, *HAND1*, and *HOXB1*) are involved in the development of extraembryonic tissues, as well as the ecto-, meso-, and endoderm. The transcription factors comprising the system of pluripotency maintenance can directly regulate the genes encoding PRC2 components. So, the transcription factors OCT4 and STAT3 (the activity of the latter depends on the presence of interleukin LIF) are positive regulators of the *Eed* gene in mESCs [127]. At the same time, OCT4 is a negative regulator of the *Hdac4* gene encoding histone deacetylase in mESCs [128]. These factors are examples of collaboration between the system of epigenetic regulation of gene expression and the system of ESC pluripotency maintenance.

The genes implicated in the maintenance of ESC pluripotency and self-renewal are subjected to epigenetic alterations during differentiation in embryogenesis and during induced or spontaneous differentiation *in vitro*. The *Oct4* gene slightly differs from other genes (*Dppa3/Stella/PGC7*, *Nanog*, and *Sox2*) with expression typical of pluripotent cells in its mechanism of epigenetic silencing during differentiation [129]. The labels of inactive chromatin, namely H4K9 methylation, and the presence of the HP1 heterochromatin protein are typical of the *Oct4* promotor region in differentiated cells of P19 embryonic carcinoma and post-implantation murine embryos. The same type of epigenetic silencing is observed in the *Rex1* gene, whose transcription is positively regulated by OCT4 (this mechanism was not observed

in the *Nanog* and *Sox2* genes). The presence of H3K9me and HP1 favors the *de novo* recruiting of DNMT3A and DNMT3B, two DNA-methyltransferases methylating the promoter region of the *Oct4* gene, which is necessary for complete and stable transcription suppression. DNA methylation is the secondary process in relation to H3K9 methylation; a fact proven when studying the differentiation of mESCs bearing homozygous mutations of the gene encoding G9a histone methyltransferase and the genes *Dnmt3a* and *Dnmt3b* encoding DNA-methyltransferases. It was also shown recently that mutations interfering with H3K9 and DNA methylation inhibit ESC differentiation [129].

The necessary presence of euchromatin proteins in the sustenance of mESC pluripotency was established recently [130]. Repression of *Chd1* gene transcription suppresses both ESC cell culture growth and *Oct4* gene promoter activity. This gene encodes the euchromatin protein, which is co-localized with the H3K4me3 label of active chromatin and RNA-polymerase II in mESCs. Suppression of *Chd1* transcription results in elevated expression of neural markers in ESCs and impairment of cell differentiation into primitive endoderm and, as a consequence, into cardiac mesoderm descendants. At that stage, the ectodermal type of differentiation occurs normally. Besides, *Chd1* suppression drastically reduces the efficacy of iPSC production from embryonic fibroblasts [130]. Note that the transcription factors, such as OCT4, SOX2, NANOG, SMAD1, ZFX, and E2F1, required for the maintenance of cell pluripotency are localized within the *Chd1* gene in ESCs [131].

The link between pluripotency and epigenetic mechanisms can be tracked in experiments on murine and human iPSC production. Several studies have demonstrated that specific inhibitors of the enzymes implicated in the epigenetic modification of histones and DNA substantially increase the efficacy of iPSC production and can even substitute some “factors of pluripotency.” Currently, similar properties are known for the inhibitors of histone deacetylases (valproic acid (VPA), trichostatin A (TSA), and suberoylanilide hydroxamic acid (SAHA)) [132, 133], G9a histone methyltransferase inhibitor BIX-01294 [134, 135], and DNA methyltransferase inhibitors 5-azacytidine and RG108 [132–136]. Besides, the use of RNA interference against *Dnmt* increases the efficacy of reprogramming due to the enhancement of partially reprogrammed cell transitions to a completely reprogrammed state [136]. Numerous studies devoted to the production of iPSCs of human and other animals have demonstrated that reprogramming somatic cells to the pluripotent state is accompanied by DNA demethylation in the promoter regions of *Oct4* and *Nanog*, whereas they are hypermethylated in murine somatic cells and TSCs [137, 138].

miRNAs AND PLURIPOTENCY

Numerous studies indicate that miRNAs play an important role in the regulation of the expression of multiple genes during embryo development. In particular, a mutation in the *Dicer* gene encoding the RNase involved in small noncoding RNA processing cause early death of murine embryos [139]. A similar effect is observed in mutants bearing the *Dgcr8* gene mutation (its protein product comprises the Microprocessor complex, which is also implicated in miRNA biogenesis). ESCs with mutations in the *Dicer* and *Dgcr8* genes are characterized by the disturbance of both the cell cycle and ability to differentiate [140–142].

To date, more than 500 various miRNAs with their expression typical in various tissues and types of differentiated cells are known. The expression of some miRNA families is restricted to undifferentiated ESCs. In murine ESCs, these families are *mir-290* (*mir-290*, *mir-291a*, *mir-291b*, *mir-292*, *mir-293*, *mir-294*, and *mir-295*) and *mir-302* cluster (*mir-302a*, *mir-302b*, *mir-302c*, *mir-302d*, and *mir-367*). In human ESCs, expression of *mir-302* family miRNAs occurs, as well as *mir-371* (*mir-371*, *mir-372*, and *mir-373*), whose representatives are homologous to those of murine *mir-290* [143].

To date, the data in numerous experiments support the notion of a tight interaction between the system of miRNA-mediated regulation of gene expression and the system of transcription factors of pluripotency, including OCT4, NANOG, and SOX2. Transcription factors can regulate the transcription of distinct genes and whole miRNA clusters, whereas miRNAs are capable of regulating *Oct4*, *Nanog*, and *Sox2* expression at the posttranscriptional level [22, 23, 143]. For instance, *mir-134*, *mir-296*, and *mir-470*, whose expression is elevated under induced differentiation of mESCs, can regulate *Oct4*, *Nanog*, and *Sox2* in various combinations, causing a decrease in the levels of the corresponding proteins [144]. An increase in the *mir-134* transcription level can induce mESC differentiation via the ectodermal pathway. The mRNAs of the *Nanog* and *LRH1* genes, whose protein products positively regulate the *Oct4* gene, are the targets of *mir-134* [145]. It was shown later that *mir-200c*, *mir-203*, and *mir-183* miRNAs can co-repress *Sox2* and *Klf4*, which is also required for the maintenance of mESC pluripotency [146]. Induction of this miRNA decreases the capability of self-renewal and leads to the induction of differentiation markers. In hESCs, the miRNA named *mir-145* is also found. Its activation induces differentiation. The OCT4 transcription factor represses the transcription of *mir-145* in nondifferentiated hESCs [147].

Localization of transcription factors in the human and murine ESC genomes has demonstrated that the

genes encoding miRNAs are among the targets of OCT4, SOX2, and NANOG [22, 23].

Highly precise distribution mapping of the transcription factors OCT4, SOX2, NANOG, and TCF3 in mESCs has shown that these transcription factors are co-localized within 55 miRNA loci, including three polycistronic clusters comprising 20% of all annotated mammalian miRNAs [143]. These miRNAs include both those that are actively transcribed in ESCs, as well as silencing forms. Hence, the transcription factors of pluripotency can act as activators and repressors of miRNA transcription in ESCs. Besides, polycomb proteins executing di- and trimethylation of histone H3 at position K27 (the label of inactive chromatin) are found in the promoter regions of the miRNA genes repressed by OCT4, SOX2, NANOG, and TCF3 (Fig. 3). Transcription rearrangement occurs under cell differentiation, resulting in the formation of specific patterns of miRNA expression in each of the differentiated cell types [143].

Apparently, miRNA can be implicated in cell reprogramming as well. It was shown that ESC-specific *mir-291-3p*, *mir-294*, and *mir-295* miRNAs can increase the efficacy of murine iPSC production without the use of *c-Myc* [148]. Besides, human iPSCs have been successfully produced using ectopic expression of the *OCT4*, *SOX2*, *KLF4*, and *LIN28* genes. The protein product of the *LIN28* gene inhibits the production of let-7 family miRNA, which participates in cell differentiation [149].

TRANSCRIPTION FACTORS OCT4, Sox2, AND NANOG AND X-CHROMOSOME INACTIVATION

The system that sustains cell pluripotency self-renewal with OCT4, SOX2, and NANOG playing the central role is associated with the fundamental genetic processes that occur in early embryogenesis in mammals, particularly, X-chromosome inactivation.

In females of higher mammals, one of two genetically equivalent X-chromosomes undergoes inactivation, which is heterochromatinization and transcription silencing of most of its genes. The inactivation process consists of several studies and is managed by a complex genetic locus, the inactivation center, localized in the X-chromosome [150]. Imprinted inactivation is observed in all murine embryo blastomers at very early developmental stages; i.e., exclusively the male-derived X-chromosome undergoing inactivation. Reactivation of the inactivated X-chromosome occurs after separation from ICM, a pluripotent compartment of the blastocyst. The imprinted inactivation is retained in extraembryonic tissues, whereas random inactivation of the X-chromosome becomes established in epiblast cells under differentiation. *Xist* and *Tsix*, which are transcribed antisenses to *Xist* from a complementary DNA strand, are two genes of the inactivation center that play a crucial

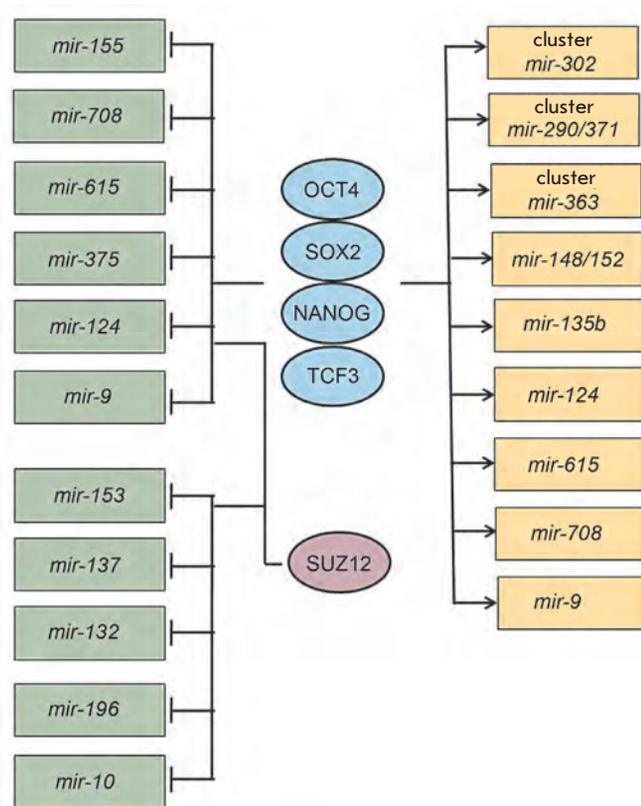


Fig. 3. miRNA genes associated with transcription factors OCT4, SOX2, NANOG, and TCF3 (indicated by arrows) and co-regulated by these factors and PRC2 (SUZ12) in murine and human embryonic stem cells [143].

role in X-inactivation [151–153] (Fig. 4). Both genes encode nontranslated nuclear RNAs. The RNA of the *Xist* gene spreads along the X-chromosome, triggering the inactivation. *Tsix* transcription has a suppressing effect on the transcription of the *Xist* gene [152].

Both X-chromosomes are active in both ESCs and ICM cells, in which the level of *Xist* RNA is extremely low [154]. Fluorescence *in situ* hybridization (FISH) allows to visualize the *Xist* transcript as a shining dot on each X-chromosome. Random inactivation of one parent X-chromosome occurs during ESC differentiation, like in embryonic epiblast cells after the implantation into the uterus. Therefore, an obvious correlation between pluripotency and X-chromosome inactivation exists in female mammal cells. The molecular origin of this correlation was not elucidated until recently.

A link was recently found between OCT4, SOX2, and NANOG. It appears to play a key role in the maintenance of ESC pluripotency and X-chromosome inactivation [155]. OCT4, SOX2, and NANOG proteins collectively bind to a DNA site in the first intron of the *Xist* gene, followed by the suppression of its transcription

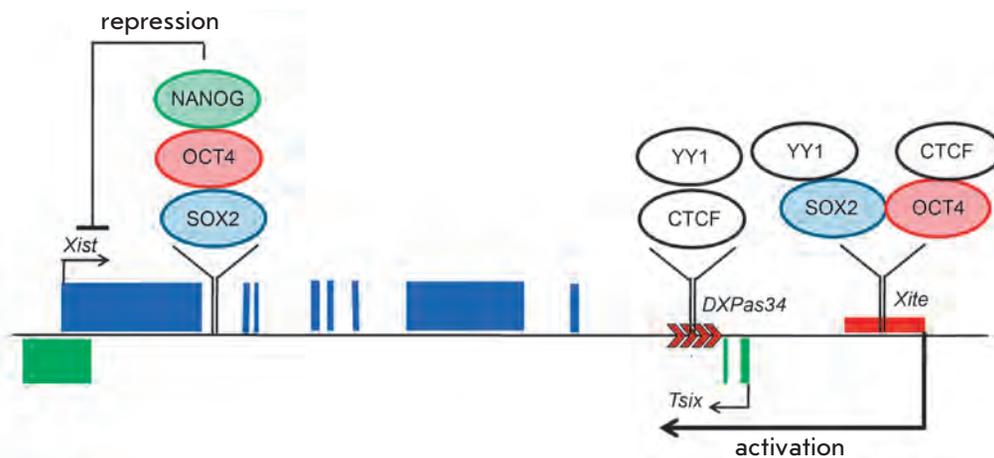


Fig. 4. Scheme of the mouse *Xist/Tsix* locus with the binding sites of OCT4, SOX2, and NANOG, and YY1 and CTCF proteins interacting with these transcription factors. *Xist* and *Tsix* exons are indicated by blue and green rectangles, respectively. The red rectangle represents enhancer *Xite*. Arrows indicate the directions of gene transcription.

in undifferentiated mESCs (Fig. 4). Reversible, limited activation of the *Xist* gene is observed in ESCs bearing a homozygous mutation of the *Nanog* gene, while normal binding of OCT4 and SOX2 proteins occurs with the first intron of *Xist*. Repression of all three factors—OCT4, SOX2, and NANOG—is accompanied by rapid accumulation of *Xist* RNA in ESC nuclei. Thus, the factors of pluripotency can directly repress the *Xist* gene via a *Tsix*-independent mechanism.

However, OCT4 and SOX2 have been shown to participate in *Xist* gene regulation via activation of its repressors, *Tsix* and *Xite* [156] (Fig. 4). A bioinformatic analysis has revealed one site for SOX2 binding and two sites for OCT4 binding in the inactivation center of the murine X-chromosome. One site of OCT4 binding is mapped to the vicinity of the CTCF and YY1 (E site) binding sites, at a distance of 1 kb from the *DXPas34* regulatory element, which also contains several sites for CTCF and YY1 binding (D site). A composite site for the binding of the OCT4 and SOX2 transcription factors was found within the region of 1.2 kb, which is known as the *Xite* enhancer (Fig. 4). The fact of OCT4 binding with the sequence of the E site, as well as the OCT4 and SOX2 factors binding to the *Xite* nucleotide sequence, was confirmed *in vitro* by gel-retardation and *in vivo* by chromatin immunoprecipitation methods. The *Xite* enhancer region containing the native OCT4 and SOX2 sites can substantially enhance the activity of the major promoter of the *Tsix* gene within the luciferase reporter gene constructs temporarily transfected into mESCs. Mutations in the OCT4 and SOX2 sites substantially decrease the capability of the *Xite* enhancer of *Tsix* promoter activation. Suppression of *Oct4* gene expression by RNA-interference causes a substantial decrease in the levels of *Tsix* and *Xite* RNA in female-derived mESCs. Suppression of *Sox2* expression also slightly increased the levels of *Tsix* and *Xite* RNA. Apparently, SOX2 plays a substantially less significant role in *Tsix* transactivation, compared to OCT4.

Apart from protein-DNA interactions, the transcription factors OCT4 and SOX2 are characterized by protein-protein interactions, which are directly associated with the X-chromosome inactivation process: OCT4 interacts with the CTCF protein, while SOX2 interacts with the YY1 protein (Fig. 4). It was shown earlier that CTCF and its cofactor YY1 are involved in X-chromosome pairing at the stages of counting and choosing of the future inactive chromosomes. Suppression of *Oct4* expression in mESCs disturbs the X-chromosome pairing in the same way as was observed under *Ctcf* repression, where the fall in the level of *Sox2* expression has no significant effect on this process. Aberrant biallelic *Xist* expression was observed during the formation of embryoid bodies from ESCs, with suppressed *Oct4* expression; this might be a result of the impairment of the X-chromosome counting process. Biallelic expression of *Xist* was not observed under suppression of *Ctcf* and *Sox2* expression [156].

Taking into account the data in the two studies mentioned above, one could assume that OCT4 regulates the *Xist* gene in two ways: direct repression – together with SOX2 and NANOG – of *Xist* transcription and activation of the *Tsix* gene. During cell differentiation, the OCT4 factor, in cooperation with CTCF, mediates normal X-chromosome pairing in the *Tsix/Xite* region, providing counting and choice of the future active and inactive X-chromosomes. The decrease in the *Oct4* gene transcription level results in the loss of OCT4 binding with one of the X-chromosomes, repression of *Tsix*, and activation of *Xist*, and it is the chromosome that becomes inactive. Yet, the residual amount of OCT4 in the coming active X-chromosome supports *Tsix* expression and *Xist* repression. This model very elegantly represents the association between pluripotency and the status of the X-chromosome in the cells of female mammals. However, a number of unanswered questions still remain. For instance, how is the difference in the binding force or in the amounts of OCT4, SOX2,

and NANOG in the coming active and inactive X-chromosomes achieved during differentiation. Logically, one should expect that, under random X-chromosome inactivation, this is a random process as well. But the question of what happens with imprinted inactivation when only a male-derived X-chromosome becomes inactive arises. Active and inactive X-chromosomes are genetically equivalent; meaning that the difference in protein binding cannot be ascribed to the difference in nucleotide sequences. There is an evident need for a search for additional protein molecules or epigenetic factors capable of modulating this process.

Another interesting question is associated with the status of the X-chromosome in human ESCs. The first produced hESC line (H9) possessed two active X-chromosomes, as was observed in mESCs [15, 157], and was devoid of *XIST* gene transcription. Random inactivation of one of the X-chromosome and an increase in *XIST* expression were observed during H9 differentiation. However, it was found later that several H9 subclones had the ability to express *XIST*, whose RNA covers the inactive chromosome even in nondifferentiated cells [158]. Moreover, other hESC lines were found to express *XIST* and to bear the inactive X-chromosome [157–159]. The data published by the International Consortium of Stem Cell Networks demonstrates that about half of the analyzed ESC lines produced in different laboratories express the *XIST* gene, simultaneously with pluripotency markers, such as OCT4, SOX2, and NANOG [159]. The analysis of eleven human ESC lines carried out by Silva and associates [160] has enabled to separate ESC into three groups: 1) cells with two active X-chromosomes, one of which is inactivated during the

differentiation process; 2) cells with one inactive chromosome in both the undifferentiated and differentiated states; and 3) cells which do not express *XIST* in both the undifferentiated and differentiated states [160]. Further experiments have shown that, in spite of the lack of *XIST* transcription, the cells belonging to the third group have an inactive X-chromosome [160]. All of the examples given above suggest that, in humans, *XIST* transcription and X-chromosome inactivation are not directly associated with pluripotency and the activity of OCT4, SOX2, and NANOG.

CONCLUSION

Embryonic stem cells are a unique object for fundamental and applied studies. Their uniqueness is rooted in two of their properties – self-renewal and pluripotency. Modern methods in cell biology and molecular genetic analysis have allowed to look anew at the molecular basis of and factors controlling self-renewal and pluripotency. The basic properties of embryonic stem cells are determined by a complex multicomponent system including transcription factors, signaling cascades, as well as a system of epigenetic regulation and miRNAs. Certainly, the obtained information will help to better understand the nature of many processes occurring in the embryogenesis of animals, including humans. Additionally, new knowledge will allow a more effective use of ESC in applied research, as well as to understand the causes behind many inherited human diseases and facilitate the development of therapies. ●

This work was supported by the Program of the RAS Presidium Molecular and Cellular Biology.

REFERENCES

1. Surani M.A., Hayashi K., Hajkova P. // *Cell*. 2007. V. 128. P. 747–762.
2. Jaenisch R., Young R. // *Cell*. 2008. V. 132. P. 567–582.
3. Durcova-Hills G., Ainscough J., McLaren A. // *Differentiation*. 2001. V. 68. P. 220–226.
4. McLaren A., Durcova-Hills G. // *Reprod. Fertil. Dev.* 2001. V. 13. P. 661–664.
5. Tesar P.J., Chenoweth J.G., Brook F.A., et al. // *Nature*. 2007. V. 448. P. 196–199.
6. Takahashi K., Yamanaka S. // *Cell*. 2006. V. 126. P. 663–676.
7. Takahashi K., Tanabe K., Ohnuki M., et al. // *Cell*. 2007. V. 131. P. 861–872.
8. Wernig M., Meissner A., Foreman R., et al. // *Nature*. 2007. V. 448. P. 318–324.
9. Johnson M.H., Ziomek C.A. // *Cell*. 1981. V. 24. P. 71–80.
10. Zernicka-Goetz M. // *Nat. Rev. Mol. Cell Biol.* 2005. V. 6. P. 919–928.
11. Rossant J. // *Cell*. 2009. V. 138. P. 1047–1050.
12. Evans M.J., Kaufman M.H. // *Nature*. 1981. V. 292. P. 154–156.
13. Martin G.R. // *Proc. Natl. Acad. Sci. USA*. 1981. V. 78. P. 7634–7638.
14. Thomson A., Wojtacha D., Hewitt Z., et al. // *Cloning Stem Cells*. 2008. V. 10. P. 89–106.
15. Thomson J.A., Itskovitz-Eldor J., Shapiro S.S., et al. // *Science*. 1998. V. 282. P. 1145–1147.
16. Tanaka S., Kunath T., Hadjantonakis A.K., et al. // *Science*. 1998. V. 282. P. 2072–2075.
17. Kunath T., Arnaud D., Uy G.D., et al. // *Development*. 2005. V. 132. P. 1649–1661.
18. Shevchenko A.I., Demina V.V., Mazurok N.A., et al. // *Rus. J. Genetics*. 2008. V. 44. P. 1280–1289.
19. Xu R.H., Chen X., Li D.S., et al. // *Nat. Biotechnol.* 2002. V. 20. P. 1261–1264.
20. Xu R.H., Peck R.M., Li D.S., et al. // *Nat. Methods*. 2005. V. 2. P. 185–190.
21. Humphrey R.K., Beattie G.M., Lopez A.D., et al. // *Stem Cells*. 2004. V. 22. P. 522–530.
22. Boyer L.A., Lee T.I., Cole M.F., et al. // *Cell*. 2005. V. 122. P. 947–956.
23. Loh Y.H., Wu Q., Chew J.L., et al. // *Nat. Genet.* 2006. V. 38. P. 431–440.
24. Niwa H., Ogawa K., Shimosato D., Adachi K. // *Nature*. 2009. V. 460. P. 118–122.
25. Burdon T., Chambers I., Stracey C., et al. // *Cells Tissues Organs*. 1999. V. 165. P. 131–143.

REVIEWS

26. Smith A.G., Heath J.K., Donaldson D.D., et al. // *Nature*. 1988. V. 336. P. 688–690.
27. Williams R.L., Hilton D.J., Pease S., et al. // *Nature*. 1988. V. 336. P. 684–687.
28. Niwa H., Burdon T., Chambers I., Smith A. // *Genes Dev*. 1998. V. 12. P. 2048–2060.
29. Matsuda T., Nakamura T., Nakao K., et al. // *EMBO J*. 1999. V. 18. P. 4261–4269.
30. Sumi T., Fujimoto Y., Nakatsuji N., Suemori H. // *Stem Cells*. 2004. V. 22. P. 861–872.
31. Shi Y., Massague J. // *Cell*. 2003. V. 113. P. 685–700.
32. James D., Levine A.J., Besser D., Hemmati-Brivanlou A. // *Development*. 2005. V. 132. P. 1273–1282.
33. Vallier L., Reynolds D., Pedersen R.A. // *Dev. Biol*. 2004. V. 275. P. 403–421.
34. Vallier L., Alexander M., Pedersen R.A. // *J. Cell Sci*. 2005. V. 118. P. 4495–4509.
35. Assou S., Le Carrour T., Tondeur S., et al. // *Stem Cells*. 2007. V. 25. P. 961–973.
36. Xiao L., Yuan X., Sharkis S.J. // *Stem Cells*. 2006. V. 24. P. 1476–1486.
37. Greber B., Lehrach H., Adjaye J. // *Stem Cells*. 2007. V. 25. P. 455–464.
38. Chen C., Ware S.M., Sato A., et al. // *Development*. 2006. V. 133. P. 319–329.
39. Levine A.J., Brivanlou A.H. // *Cell Cycle*. 2006. V. 5. P. 1069–1073.
40. Ying Q.L., Nichols J., Chambers I., Smith A. // *Cell*. 2003. V. 115. P. 281–292.
41. Amit M., Carpenter M.K., Inokuma M.S., et al. // *Dev. Biol*. 2000. V. 227. P. 271–278.
42. Hamazaki T., Kehoe S.M., Nakano T., Terada N. // *Mol. Cell. Biol*. 2006. V. 26. P. 7539–7549.
43. Dvorak P., Dvorakova D., Koskova S., et al. // *Stem Cells*. 2005. V. 23. P. 1200–1211.
44. Dvorak P., Hampl A. // *Folia Histochem. Cytobiol*. 2005. V. 43. P. 203–208.
45. Aubert J., Dunstan H., Chambers I., Smith A. // *Nat. Biotechnol*. 2002. V. 20. P. 1240–1245.
46. Sato N., Meijer L., Skaltsounis L., et al. // *Nat. Med*. 2004. V. 10. P. 55–63.
47. Dravid G., Ye Z., Hammond H., et al. // *Stem Cells*. 2005. V. 23. P. 1489–1501.
48. Lu J., Hou R., Booth C.J., et al. // *Proc. Natl. Acad. Sci. USA*. 2006. V. 103. P. 5688–5693.
49. Okamoto K., Okazawa H., Okuda A., et al. // *Cell*. 1990. V. 60. P. 461–472.
50. Rosner M.H., Vigano M.A., Ozato K., et al. // *Nature*. 1990. V. 345. P. 686–692.
51. Scholer H.R., Balling R., Hatzopoulos A.K., et al. // *EMBO J*. 1989. V. 8. P. 2551–2557.
52. Palmieri S.L., Peter W., Hess H., Scholer H.R. // *Dev. Biol*. 1994. V. 166. P. 259–267.
53. Nichols J., Zevnik B., Anastassiadis K., et al. // *Cell*. 1998. V. 95. P. 379–391.
54. Kirchhof N., Carnwath J.W., Lemme E., et al. // *Biol. Reprod*. 2000. V. 63. P. 1698–1705.
55. Hansis C., Grifo J.A., Krey L.C. // *Mol. Hum. Reprod*. 2000. V. 6. P. 999–1004.
56. Mitalipov S.M., Kuo H.C., Hennebold J.D., Wolf D.P. // *Biol. Reprod*. 2003. V. 69. P. 1785–1792.
57. Cauffman G., van de Velde H., Liebaers I., van Steirteghem A. // *Mol. Hum. Reprod*. 2005. V. 11. P. 173–181.
58. Niwa H., Miyazaki J., Smith A.G. // *Nat. Genet*. 2000. V. 24. P. 372–376.
59. Kehler J., Tolkunova E., Koschorz B., et al. // *EMBO Rep*. 2004. V. 5. P. 1078–1083.
60. Velkey J.M., O'Shea K.S. // *Genesis*. 2003. V. 37. P. 18–24.
61. Hay D.C., Sutherland L., Clark J., Burdon T. // *Stem Cells*. 2004. V. 22. P. 225–235.
62. Takeda J., Seino S., Bell G.I. // *Nucl. Acids Res*. 1992. V. 20. P. 4613–4620.
63. van Eijk M.J., van Rooijen M.A., Modina S., et al. // *Biol. Reprod*. 1999. V. 60. P. 1093–1103.
64. Nordhoff V., Hubner K., Bauer A., et al. // *Mamm. Genome*. 2001. V. 12. P. 309–317.
65. Medvedev S.P., Shevchenko A.I., Elisaphenko E.A., et al. // *BMC Genomics*. 2008. V. 9. P. 162.
66. Yeom Y.I., Ha H.S., Balling R., et al. // *Mech. Dev*. 1991. V. 35. P. 171–179.
67. Atlasi Y., Mowla S.J., Ziaee S.A., et al. // *Stem Cells*. 2008. V. 26. P. 3068–3074.
68. Lee J., Kim H.K., Rho J.Y., et al. // *J. Biol. Chem*. 2006. V. 281. P. 33554–33565.
69. Cauffman G., Liebaers I., van Steirteghem A., van de Velde H. // *Stem Cells*. 2006. V. 24. P. 2685–2691.
70. Wang X., Zhao Y., Xiao Z., et al. // *Stem Cells*. 2009. V. 27. P. 1265–1275.
71. Wang S.H., Tsai M.S., Chiang M.F., Li H. // *Gene Expr. Patterns*. 2003. V. 3. P. 99–103.
72. Chambers I., Colby D., Robertson M., et al. // *Cell*. 2003. V. 113. P. 643–655.
73. Mitsui K., Tokuzawa Y., Itoh H., et al. // *Cell*. 2003. V. 113. P. 631–642.
74. Hyslop L., Stojkovic M., Armstrong L., et al. // *Stem Cells*. 2005. V. 23. P. 1035–1043.
75. Pan G.J., Pei D.Q. // *Cell Res*. 2003. V. 13. P. 499–502.
76. Pan G., Pei D. // *J. Biol. Chem*. 2005. V. 280. P. 1401–1407.
77. Mullin N.P., Yates A., Rowe A.J., et al. // *Biochem. J*. 2008. V. 411. P. 227–231.
78. Wang J., Levasseur D.N., Orkin S.H. // *Proc. Natl. Acad. Sci. USA*. 2008. V. 105. P. 6326–6331.
79. Wang Z., Ma T., Chi X., Pei D. // *J. Biol. Chem*. 2008. V. 283. P. 4480–4489.
80. Medvedev S.P., Elisaphenko E.A., Shevchenko A.I., et al. // *Dokl. Biochem. Biophys*. 2009. V. 425. P. 102–105.
81. Avilion A.A., Nicolis S.K., Pevny L.H., et al. // *Genes Dev*. 2003. V. 17. P. 126–140.
82. Masui S., Nakatake Y., Toyooka Y., et al. // *Nat. Cell Biol*. 2007. V. 9. P. 625–635.
83. Adachi K., Suemori H., Yasuda S.Y., et al. // *Genes Cells*. 2010. V. 15. P. 455–470.
84. Botquin V., Hess H., Fuhrmann G., et al. // *Genes Dev*. 1998. V. 12. P. 2073–2090.
85. Yuan H., Corbi N., Basilico C., Dailey L. // *Genes Dev*. 1995. V. 9. P. 2635–2645.
86. Nishimoto M., Fukushima A., Okuda A., Muramatsu M. // *Mol. Cell. Biol*. 1999. V. 19. P. 5453–5465.
87. Liu L., Roberts R.M. // *J. Biol. Chem*. 1996. V. 271. P. 16683–16689.
88. Liu L., Leaman D., Villalta M., Roberts R.M. // *Mol. Endocrinol*. 1997. V. 11. P. 1651–1658.
89. Pesce M., Scholer H.R. // *Mol. Reprod. Dev*. 2000. V. 55. P. 452–457.
90. Ambrosetti D.C., Basilico C., Dailey L. // *Mol. Cell. Biol*. 1997. V. 17. P. 6321–6329.
91. Tomilin A., Remenyi A., Lins K., et al. // *Cell*. 2000. V. 103. P. 853–864.
92. Remenyi A., Tomilin A., Pohl E., et al. // *Mol. Cell*. 2001. V. 8. P. 569–580.

REVIEWS

93. Kang J., Gemberling M., Nakamura M., et al. // *Genes Dev.* 2009. V. 23. P. 208–222.
94. Abeyta M.J., Clark A.T., Rodriguez R.T., et al. // *Hum. Mol. Genet.* 2004. V. 13. P. 601–608.
95. Brandenberger R., Khrebtukova I., Thies R.S., et al. // *BMC Dev. Biol.* 2004. V. 4. P. 10.
96. Catena R., Tiveron C., Ronchi A., et al. // *J. Biol. Chem.* 2004. V. 279. P. 41846–41857.
97. Kuroda T., Tada M., Kubota H., et al. // *Mol. Cell. Biol.* 2005. V. 25. P. 2475–2485.
98. Niwa H. // *Cell Struct. Funct.* 2001. V. 26. P. 137–148.
99. Okumura-Nakanishi S., Saito M., Niwa H., Ishikawa F. // *J. Biol. Chem.* 2005. V. 280. P. 5307–5317.
100. Rodda D.J., Chew J.L., Lim L.H., et al. // *J. Biol. Chem.* 2005. V. 280. P. 24731–24737.
101. Sato N., Sanjuan I.M., Heke M., et al. // *Dev. Biol.* 2003. V. 260. P. 404–413.
102. Wei C.L., Miura T., Robson P., et al. // *Stem Cells.* 2005. V. 23. P. 166–185.
103. Chew J.L., Loh Y.H., Zhang W., et al. // *Mol. Cell. Biol.* 2005. V. 25. P. 6031–6046.
104. Suzuki A., Raya A., Kawakami Y., et al. // *Proc. Natl. Acad. Sci. USA.* 2006. V. 103. P. 10294–10299.
105. Wang J., Rao S., Chu J., et al. // *Nature.* 2006. V. 444. P. 364–368.
106. Wu Q., Chen X., Zhang J., et al. // *J. Biol. Chem.* 2006. V. 281. P. 24090–24094.
107. Pardo M., Lang B., Yu L., et al. // *Cell Stem Cell.* 2010. V. 6. P. 382–395.
108. van den Berg D.L., Snoek T., Mullin N.P., et al. // *Cell Stem Cell.* 2010. V. 6. P. 369–381.
109. Hatano S.Y., Tada M., Kimura H., et al. // *Mech. Dev.* 2005. V. 122. P. 67–79.
110. Lauberth S.M., Rauchman M. // *J. Biol. Chem.* 2006. V. 281. P. 23922–23931.
111. Korutla L., Wang P.J., Mackler S.A. // *J. Neurochem.* 2005. V. 94. P. 786–793.
112. Tomioka M., Nishimoto M., Miyagi S., et al. // *Nucleic Acids Res.* 2002. V. 30. P. 3202–3213.
113. Brehm A., Ohbo K., Zwerschke W., et al. // *Mol. Cell. Biol.* 1999. V. 19. P. 2635–2643.
114. Guo Y., Costa R., Ramsey H., et al. // *Proc. Natl. Acad. Sci. USA.* 2002. V. 99. P. 3663–3667.
115. Tolkunova E., Malashicheva A., Parfenov V.N., et al. // *J. Mol. Biol.* 2007. V. 374. P. 1200–1212.
116. Webster D.M., Teo C.F., Sun Y., et al. // *BMC Dev. Biol.* 2009. V. 9. P. 28.
117. Bernstein B.E., Mikkelsen T.S., Xie X., et al. // *Cell.* 2006. V. 125. P. 315–326.
118. Ringrose L., Paro R. // *Annu. Rev. Genet.* 2004. V. 38. P. 413–443.
119. O'Carroll D., Erhardt S., Pagani M., et al. // *Mol. Cell. Biol.* 2001. V. 21. P. 4330–4336.
120. Pasini D., Bracken A.P., Jensen M.R., et al. // *EMBO J.* 2004. V. 23. P. 4061–4071.
121. Shumacher A., Faust C., Magnuson T. // *Nature.* 1996. V. 383. P. 250–253.
122. Voncken J.W., Roelen B.A., Roefs M., et al. // *Proc. Natl. Acad. Sci. USA.* 2003. V. 100. P. 2468–2473.
123. Boyer L.A., Plath K., Zeitlinger J., et al. // *Nature.* 2006. V. 441. P. 349–353.
124. Lee T.I., Jenner R.G., Boyer L.A., et al. // *Cell.* 2006. V. 125. P. 301–313.
125. Chamberlain S.J., Yee D., Magnuson T. // *Stem Cells.* 2008. V. 26. P. 1496–1505.
126. Woolfe A., Goodson M., Goode D.K., et al. // *PLoS Biol.* 2005. V. 3. P. e7.
127. Ura H., Usuda M., Kinoshita K., et al. // *J. Biol. Chem.* 2008. V. 283. P. 9713–9723.
128. Addis R.C., Prasad M., Yochem R.L., et al. // *J. Cell. Biochem.* 2010. [Epub ahead of print]
129. Feldman N., Gerson A., Fang J., et al. // *Nat. Cell Biol.* 2006. V. 8. P. 188–194.
130. Gaspar-Maia A., Alajem A., Polesso F., et al. // *Nature.* 2009. V. 460. P. 863–868.
131. Chen X., Xu H., Yuan P., et al. // *Cell.* 2008. V. 133. P. 1106–1117.
132. Huangfu D., Maehr R., Guo W., et al. // *Nat. Biotechnol.* 2008. V. 26. P. 795–797.
133. Huangfu D., Osafune K., Maehr R., et al. // *Nat. Biotechnol.* 2008. V. 26. P. 1269–1275.
134. Shi Y., Despons C., Do J.T., et al. // *Cell Stem Cell.* 2008. V. 3. P. 568–574.
135. Shi Y., Do J.T., Despons C., et al. // *Cell Stem Cell.* 2008. V. 2. P. 525–528.
136. Mikkelsen T.S., Hanna J., Zhang X., et al. // *Nature.* 2008. V. 454. P. 49–55.
137. Hattori N., Nishino K., Ko Y.G., et al. // *J. Biol. Chem.* 2004. V. 279. P. 17063–17069.
138. Hattori N., Imao Y., Nishino K., et al. // *Genes Cells.* 2007. V. 12. P. 387–396.
139. Bernstein E., Kim S.Y., Carmell M.A., et al. // *Nat. Genet.* 2003. V. 35. P. 215–217.
140. Kanellopoulou C., Muljo S.A., Kung A.L., et al. // *Genes Dev.* 2005. V. 19. P. 489–501.
141. Murchison E.P., Partridge J.F., Tam O.H., et al. // *Proc. Natl. Acad. Sci. USA.* 2005. V. 102. P. 12135–12140.
142. Wang Y., Medvid R., Melton C., et al. // *Nat. Genet.* 2007. V. 39. P. 380–385.
143. Marson A., Levine S.S., Cole M.F., et al. // *Cell.* 2008. V. 134. P. 521–533.
144. Tay Y., Zhang J., Thomson A.M., et al. // *Nature.* 2008. V. 455. P. 1124–1128.
145. Tay Y.M., Tam W.L., Ang Y.S., et al. // *Stem Cells.* 2008. V. 26. P. 17–29.
146. Wellner U., Schubert J., Burk U.C., et al. // *Nat. Cell Biol.* 2009. V. 11. P. 1487–1495.
147. Xu N., Papagiannakopoulos T., Pan G., et al. // *Cell.* 2009. V. 137. P. 647–658.
148. Judson R.L., Babiarz J.E., Venere M., Blueloch R. // *Nat. Biotechnol.* 2009. V. 27. P. 459–461.
149. Yu J., Vodyanik M.A., Smuga-Otto K., et al. // *Science.* 2007. V. 318. P. 1917–1920.
150. Chureau C., Prissette M., Bourdet A., et al. // *Genome Res.* 2002. V. 12. P. 894–908.
151. Brockdorff N., Ashworth A., Kay G.F., et al. // *Cell.* 1992. V. 71. P. 515–526.
152. Lee J.T., Davidow L.S., Warshawsky D. // *Nat. Genet.* 1999. V. 21. P. 400–404.
153. Lee J.T., Lu N. // *Cell.* 1999. V. 99. P. 47–57.
154. Wutz A., Jaenisch R. // *Mol. Cell.* 2000. V. 5. P. 695–705.
155. Navarro P., Chambers I., Karwacki-Neisius V., et al. // *Science.* 2008. V. 321. P. 1693–1695.
156. Donohoe M.E., Silva S.S., Pinter S.F., et al. // *Nature.* 2009. V. 460. P. 128–132.
157. Dhara S.K., Benvenisty N. // *Nucleic Acids Res.* 2004. V. 32. P. 3995–4002.
158. Hoffman L.M., Hall L., Batten J.L., et al. // *Stem Cells.* 2005. V. 23. P. 1468–1478.
159. Adewumi O., Aflatoonian B., Ahrlund-Richter L., et al. // *Nat. Biotechnol.* 2007. V. 25. P. 803–816.
160. Silva S.S., Rowntree R.K., Mekhoubad S., Lee J.T. // *Proc. Natl. Acad. Sci. USA.* 2008. V. 105. P. 4820–4825.

Protein Engineering of Penicillin Acylase

V. I. Tishkov^{1,2,3*}, S. S. Savin^{2,3}, A. S. Yasnaya^{2,4}

¹Department of Chemical Enzymology, Faculty of Chemistry, Lomonosov Moscow State University

²Bach Institute of Biochemistry, Russian Academy of Sciences

³Innovations and High Technologies MSU Ltd.

⁴National Research Center "Kurchatovskii Institute"

*E-mail: vitishkov@gmail.com

Received 05.09.2010

ABSTRACT Penicillin acylases (PA) are widely used for the production of semi-synthetic β -lactam antibiotics and chiral compounds. In this review, the latest achievements in the production of recombinant enzymes are discussed, as well as the results of PA type G protein engineering.

KEYWORDS penicillin acylase, *E.coli*, expression, structure, protein engineering

ABBREVIATIONS PA – penicillin acylase, *EcPA* and *AfPA* – penicillin acylases from *Escherichia coli* and *Alcaligenes faecalis*, respectively, 6-APA – 6-aminopenicillanic acid, D-PGA – D-phenylglycine amide, FAA – phenylacetic acid.

INTRODUCTION.

Penicillin acylase (PA, EC 3.5.1.11) was discovered 60 years ago as a catalyst of the hydrolysis of the amide bond in penicillin antibiotics [1]. This enzyme belongs to the class of hydrolases, a subclass of aminohydrolases, and represents a group of so-called N-terminal nucleophilic hydrolases. PA has been found in bacteria, yeast, and fungi. The physiological role of the enzyme remains poorly understood. It seems possible that its main function is in utilizing heterocyclic compounds as a source of carbon.

PA has been extensively studied for more than 50 years. In practice, this enzyme is commonly used to produce 6-aminopenicillanic acid, which is the main synthon in the synthesis of penicillin antibiotics. PA is also used for the synthesis of various semi-synthetic β -lactam antibiotics. Broad substrate specificity and high regio-, chemo- and stereoselectivity of the enzyme are used for the production of chiral compounds (which are more and more in demand in modern pharmaceuticals), as well as for the protection of hydroxy and amino groups in peptide and fine organic synthesis.

Currently, the most commonly used PA is that from *Escherichia coli* (*EcPA*). This enzyme has been better studied and characterized in comparison with the other PAs; however, the efficiency of the acyl transfer into β -lactam cores, catalysed by *EcPA*, is not high enough to make the enzyme competitive as compared with the out-of-date methods of antibiotic synthesis. At the same time, intensive spreading of pathogens resistant to currently available antibiotics has inevitably prompted the search for new antimicrobial compounds with artificial side chains. The catalytic efficiency of

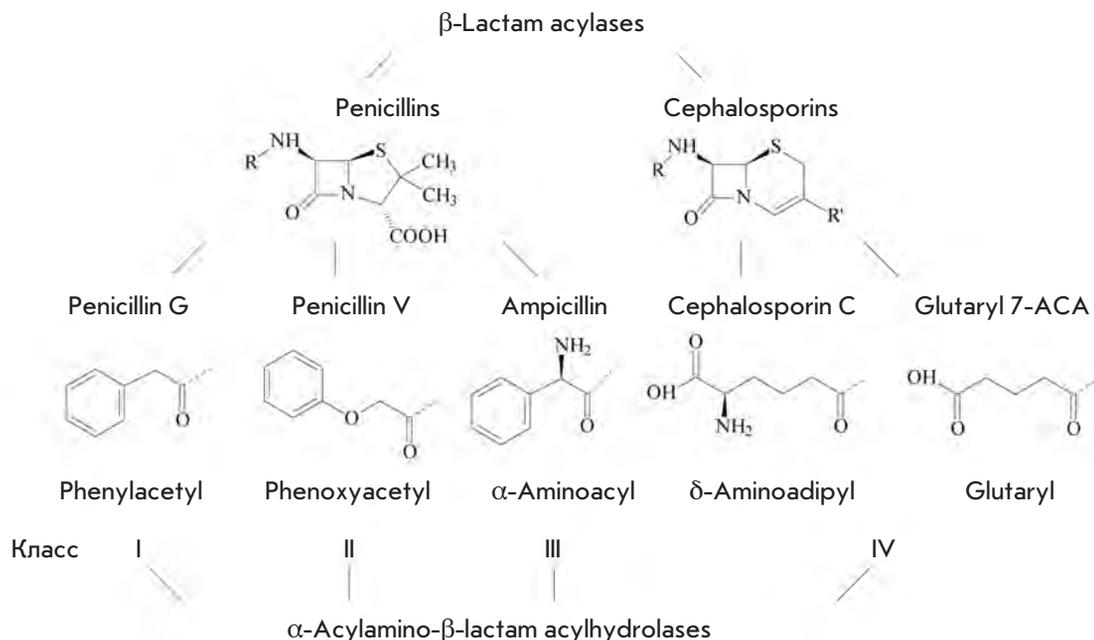
wild type PA would be too low for the synthesis of these substances: therefore, preparation of new mutant enzymes with improved synthetic parameters with artificial substrates is of great importance for practical and fundamental science. Another, equally important problem is improving a PA's stereoselectivity in reactions of amino-alcohols acylation and hydrolysis of N-acyl derivatives, since a PA's stereoselectivity in reactions of amino-alcohols production is incomparably lower than for amino acids.

New enzymes with improved properties can be obtained by various means: for example, random mutagenesis (directed evolution, gene shuffling, etc.) and site-directed mutagenesis based on an analysis of 3D protein structures (rational design). The second approach has become more widespread; however, it requires knowledge of the enzyme's 3D structure, which can be obtained in an experiment (X-ray diffraction, NMR) or by computer modeling. In the present review we will discuss the recent achievements in the production of recombinants PAs with improved properties.

CLASSIFICATION OF PENICILLIN ACYLASES

Penicillin acylases represent a group of β -lactam acylases and can be classified according to the type of the hydrolysed substrate. Therefore, enzymes can be grouped as those that hydrolyse penicillin G, penicillin V, or ampicillin. In 1963 it was suggested to divide penicillin acylases into classes I and II [2]. Class I enzymes basically hydrolyse penicillin V (phenoximethylpenicillin), while class II enzymes use penicillin G (benzylpenicillin) as a substrate. Later, the class III, including the enzymes which hydrolyse ampicillin, was added [3-6].

Fig. 1. Modern classification of β -lactam acylases according to the type of substrate [7]. 7-ACA – 7-aminocephalosporanic acid



An up-to-date classification of acylases of β -lactams according to the preferred substrate core structure and side chains [7] is shown in Fig. 1. According to this classification, class II PAs can be additionally divided into two subgroups: class IIa – enzymes that basically hydrolyse aromatic amides, and class IIb – enzymes that hydrolyse aliphatic amides [7,8].

SOURCES AND LOCALIZATION OF PENICILLIN ACYLASES

The activity of penicillin acylase was described for the first time by Japanese scientists in 1950 [1]. They found that the mycelium of both *Penicillium chrysogenum* and *Aspergillus oryzae* was able to convert benzylpenicillin into phenylacetic acid and another compound, which the authors called “penicin” and which was later identified as 6-aminopenicillanic acid (6-APA) [1,9,10]. More detailed studies of this enzyme were carried out in the 1960s. At the time, a large number of papers on the enzyme properties, and its potential application, were published. Today, the amount of papers devoted to penicillin acylases increases with every year.

It should be noted that penicillin acylase activity was also detected in bacteria, yeast, and fungi [11-14]. At the present time, PAs from more than 40 different microorganisms have been described. Many genes of penicillin acylases were found in annotated genomes of microorganisms.

Depending on the species of the microorganism, the enzyme can dwell either outside or inside the cell. Localization in periplasma is characteristic for active forms of G-class penicillin acylases (class II). Extracellular expression is also typical for some strains produc-

ing penicillin acylases V (class I) and penicillin acylases G (class II). The physiological role of PAs remains unclear despite a 60-year-long history of studying them. It is highly probable that PAs are needed for the utilization of aromatic amides as carbon sources [15]. Basic properties of some well-studied penicillin acylases are presented in Table 1.

CHARACTERISTIC FEATURES OF PENICILLIN ACYLASE G EXPRESSION

PA-G gene encodes a precursor polypeptide which consists of 4 structural elements: a signal peptide, α - and β -subunits, and an inter-subunit spacer. The mature PA-G molecule is a heterodimer with a molecular weight of 86 kDa. It consists of two subunits, α - and β -, with molecular masses of 23 and 63 kDa, respectively [23,24]. In addition, the molecule contains a bound Ca^{2+} ion, which, according to data, is important for enzyme processing [18].

Posttranslational modification of PA-G is a multi-stage process, which has been well studied for the enzyme from *E.coli*. The first step includes transport of the inactive precursor from the cytoplasm to the periplasmic compartment, a process driven by the signal peptide, which is then removed after the transport is completed. Afterwards, the inter-subunit spacer undergoes two-step proteolysis, which results in the formation of an active heterodimer [23,25,26].

In *E.coli* cells, transport of most of the proteins across the membrane is mediated by Sec-translocase. Sec-system recognizes signal peptides that do not have similar sequences but are quite close in their physico-

Table 1. Main characteristics of penicillin acylases from different sources

Class	Source	Preferable substrate*	Induction of biosynthesis	Cell localization	References
Class I PA-V	<i>Streptomyces lavendulae</i>	Pen-V, Pen-K	no	outside	[3]
	<i>Dermatophytes</i>	Pen-V	yes	inside	[3]
	<i>Penicillium chrysogenum</i>	Pen-V, Pen-K	yes	inside	[3]
	<i>Bacillus sphaericus</i>	Pen-V	no data	outside	[16]
	<i>Fusarium sp.</i>	Pen-V	no data	outside	[17]
Class II PA-G	<i>E. coli</i>	Pen-G, Pen-X	yes	inside	[18]
	<i>Bacillus megaterium</i>	Pen-G, Cephalexin, Cephaloglycine, Cephaloridine	no data	outside	[19]
	<i>Alcaligenes faecalis</i>	Pen-G	no data	inside	[20]
	<i>Providencia rettgeri</i>	Pen-G	no data	inside	[21]
Class III Amp-PA	<i>Pseudomonas melanogenum</i>	Amp, Cephaloglycine, Cephadrine, Cephalexin	no data	inside	[22]

* Amp – ampicillin, Pen-G – penicillin G, Pen-V - penicillin V, Pen-K – penicillin K, Pen-X – penicillin X

chemical properties. SecB-chaperone recognizes a short sequence containing aromatic and positively charged amino acids, which subsequently is processed by Sec-translocase [27]. SecB is a basic component that participates in the translocation of many secretory and periplasmic proteins. In work [28], it was shown that SecB is essential for *EcPA-G* maturation.

Another pathway of protein secretion is the Tat-system (twin arginine translocation). This mechanism is considered to mediate the transport of folded proteins with bound cofactors across the membrane. Evidently, the Tat-system and Sec-system work independently. In Tat-mediated transport, it is necessary for the signal peptide sequence to include the SRRXFLK motif, which contains two arginine residues [29]. The sequence of the signal peptide in the *EcPA* precursor contains 2 arginines separated by an asparagine residue (Fig. 2). Some literature data indicate that such a signal peptide can provide Tat-mediated enzyme processing. More recent data directly show that the *E.coli* signal peptide accounts for the enzyme translocation by the Tat-system [30]. PA from *Alcaligenes faecalis* is transported via the Sec-system [31].

Experiments on PA protein precursors lacking the signal peptide show that the inter-subunit spacer is initially cleaved at the N-terminus of the β -subunit, which, as a result, yields a free β -subunit and a polypeptide consisting of the α -subunit and the spacer [23]. Spacer cleavage occurs similarly in two stages in the course of PA maturation in an intracellular medium [23]. The authors of [23] demonstrated that spacer cleavage from the N-terminus of the β -subunit could

take place only in the periplasmic compartment, while spacer removal from the C-terminus of the α -subunit can also take place in cytoplasm. At the same time, work [26] revealed that, at the very beginning, the α -subunit acquires a folded structure before beginning to function as a template for the correct folding of the β -subunit. Later, it was shown that spacer cleavage is an autocatalytic process [25]. Amino acid residues Lys299 and Thr263 of the inactive precursor were shown to play the primary role in the correct folding of the enzyme. Amino acid substitutions in these positions lead to a decreased yield of the active enzyme and accumulation of the inactive precursor [25, 32].

In summing up the available data, it is important to note that all stages of PA-G processing play an important role in the production of an active and soluble protein. The introduction of amino acid substitutions can affect any steps in protein processing; therefore, the production of active forms of novel mutant enzymes requires a careful selection of appropriate conditions for cultivation.

EXPRESSION OF RECOMBINANT PAs IN *E. coli*

Depending on the purposes being pursued, cloning and expression of recombinant enzymes can be performed in various systems; for example, bacteria [33-36], yeast [37-40], plants [41-43], insect cells [44-47], or cell-free systems [48-51]. Since PA-G is found only in bacteria, *E.coli*-based expression systems are considered to be the most appropriate for recombinant enzyme production. Below, we will discuss various approaches of PA-G expression in *E.coli* cells.

Table 2. Cultivation conditions and their influence on yields of recombinant PA's

Enzyme source	Producing strain		Promoter	Antibiotics resistance	Inductor	T, °C	Medium	Additives	Enzyme yield	Reference
	Organism	Strain								
<i>E. coli</i>	<i>E. coli</i>	JM109	trc		IPTG/ arabinose	30	LB		112 U/L/A ₆₀₀ , 193 U/L/A ₆₀₀	[66]
	<i>E. coli</i>	JM109 BL21 HB101	T7	Kn	arabinose	30	LB	glycerol	330 U/L /A ₆₀₀ , 820 U/L, 440 U/L /A ₆₀₀ , 690 U/L	[63]
	<i>E. coli</i>	GM48 HB101	T7 Trc	Kn	IPTG	30	LB		84 U/L /A ₆₀₀	[32]
	<i>E. coli</i>	MC1000	lacZ	Kn	IPTG	28	M9	PAA	14–650 U/g	[67]
	<i>E. coli</i>	DH5α	tac lacI	CmR	IPTG					[68]
	<i>E. coli</i>	x6212	trc		IPTG				1000 U/L, 700 U/g	[69]
	<i>E. coli</i>	JM101, JM103, JM105	lac	Kn	IPTG galactose		M9	glucose, galactose, glucose+ galactose	20–800 U/L	[70]
	<i>E. coli</i>							SecB- chaperone	40–126 U/L/A ₆₀₀	[28]
	<i>E. coli</i>			Cm	IPTG	28	M9	Ca ²⁺	1 g/L 1700 U/g	[31]
	<i>E. coli</i>		trc T7 araB	Kn Cm	IPTG, arabinose			degP- chaperone	PA: 20-800 U/L PA: 10450 U/L	[65]
	<i>E. coli</i>	BL21 (DE3)	T7							[60, 71]
	<i>A. faecalis</i>	<i>E. coli</i>			Cm	IPTG		M9	Ca ²⁺	2.3 g/L, 14000 U/g (AfPA)
<i>E. coli</i>		HB101							1000 U, 23 mg/L	[20]
<i>E. coli</i>		JM109	rhaBAD	Amp	rhamnose		M9 LB	glucose, proline	4500 U/L, 0.96 U/mg	[72]
<i>E. coli</i>		JM109 (DE3)	T7	Kn	IPTG		LB		(200–270)×10 ³ U/L	[21]
<i>K. citrophila</i>	<i>E. coli</i>	BL21 (DE3)	T7	Kn	IPTG		YE TH M9 MR		13460 U/L 1190 U/L /A ₆₀₀ 28056 U/L 1576 U/L /A ₆₀₀	[60]

Selection of promoter and expression vector

A substantial amount of information on homo- and heterological PA expression in *E. coli* can be found in the literature. In terms of choosing an expression vector, both high- and low-copy number plasmids have been used. The following promoters have been applied for control of protein expression: *lac*, *tac*, *trc*, *T7* and *araB*. PA-G genes were cloned from *E. coli* [52, 53], *Actino-*

myces viscosus [54], *Providencia rettgeri* [55], *Kluyvera citrophila* [56], *Bacillus megaterium* [35, 57], and *A. faecalis* [20]. The influence of the promoter type on the PA yield is shown in Table 2. There is no strong correlation between the type of promoter and the yield of the active enzyme. As a rule, stronger promoters (e.g. promoter of a T7 phage RNA-polymerase) provided higher levels of expression; however, in many cases it

leads to abnormal protein folding and accumulation of an insoluble enzyme as the so-called inclusion bodies. The yield of the soluble protein can be increased by using less strong promoters. Generally, expression of each individual enzyme requires selecting an optimal vector.

Factors that affect the level of PA expression are transcription and translation [53]. The limiting stage depends on the selected vector and the host strain. Nevertheless, most researchers choose strong promoters, such as *tac* and *T7*, and isopropyl- β -D-thiogalactoside (IPTG) is used as inducer. The yield of the soluble native protein is usually increased through the optimization of cultivation parameters.

Influence of the temperature

Work [58] revealed that the yield of active PA strongly depends on the temperature of cultivation. The experiments on protein expression were carried out at three different temperatures: 22, 28, and 37°C. For all combinations of plasmids and host strains, the highest yield of the native protein was observed at 22°C. This effect is probably explained by the fact that, at a low temperature, the enzyme undergoes successfully all stages of posttranslational modification and acquires the native conformation, which leads to an increase in the yield of the soluble active enzyme. An increase in temperature causes an increase in the rate of protein synthesis; therefore, the insoluble enzyme is accumulated in the cytoplasm.

Influence of the medium composition and various additives

Numerous studies have shown that the composition of the cultivation medium strongly affects the level of expression and the yield of the active enzyme. Work [59] showed that the highest yield of active PA-G is achieved when the minimal medium M9 is used with glucose as a carbon source. This approach allows to increase the portion of the periplasmic enzyme nearly 100-fold, relative to the intracellular enzyme. High enzyme yields were also achieved using the richest medium, YE, of all those studied (YE, TH, MR, M9), as described in [60].

One of the key parameters of PA expression is the presence of calcium ions in the cultivation medium. Ca^{2+} is important both for cellular growth [61] and for the formation of active soluble PA, since calcium ions are present in molecules of active PA [24,60]. According to some authors, Ca^{2+} plays an important role in the translocation of the protein precursor across the membrane, as well as in the correct folding and maturation of the enzyme in the periplasm [31].

Another approach that allows to increase the level of

soluble PPA in the cell is adding low-molecular carbon sources into the medium. It has been shown that additional carbon sources allow to increase the biomass, as well as diminish the accumulation of the insoluble protein in the cell [62]. The most common low-molecular carbon-containing compounds are lactose, glucose, arabinose, and glycerol. Works [62,63] showed that the addition of glycerol reduces the level of the insoluble protein in the periplasm. Some authors consider glycerol as not only a carbon source, but also a “chemical chaperone” that facilitates protein folding and maturation.

Regulation of the posttranslational modification

Work [59] focuses on selecting the optimal strain for expression and optimization of medium compositions, as well as modifying the PA translocation system by means of alteration of the signal peptides structure. As it was previously discussed, the signal peptide of *EcPA* is responsible for the translocation of the precursor by the Tat-system [31]. However, as is known from the literature, Sec-mediated protein transfer occurs more rapidly [64]. The alteration of the translocating system allowed a substantial increase in the level of the soluble protein in the periplasmic fraction.

The use of chaperons for correct PA folding

An additional, actively developing approach that has emerged over the past decade is the use of chaperones for the optimization of protein expression. In case of PA, most researchers used degP chaperones or the chaperones of the Sec-system, if they participate in the intracellular transport of the enzyme [65]. This approach helps to achieve a many-fold increase in the yield of the active enzyme.

In summary, it should be noted that PA-G expression depends on numerous factors. To create a strain for PA overexpression, one should be well aware of the details of limiting stages of expression, which are specific for each organism and expression system. The yield of the active enzyme can be improved if proteinases-lacking strains or strains with an altered translocation system are used.

PA-G STRUCTURE

Primary structure of PA-G

At present, databases (GeneBank, EMBL) contain full nucleotide and amino acid sequences of PA-Gs from 40 organisms. Figure 2 shows the alignment of amino acid sequences of PA-Gs from different sources. In order to show comparison, the most distant enzymes are presented. The primary structure of *E.coli* PA is given as a full-length amino acid sequence, whereas for the

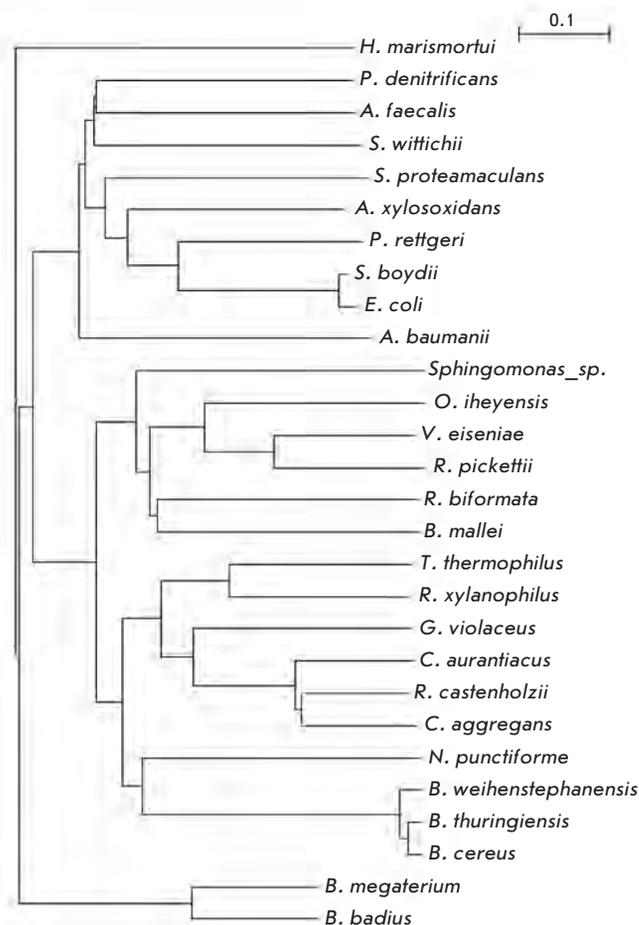


Fig. 3. Phylogenetic tree of penicillin acylases G.

other enzymes only the differing residues are shown. Well-conserved residues are marked by dark gray, and similar residues are given in light gray.

As is seen in the figure, the signal peptide region contains no conserved residues or consensus sequences. They all differ not only in the amino acid sequence but also in length. Intra-subunit spacers differ as well and do not have conserved amino acids in their sequences.

The sequences shown in Fig. 2 contain two important conserved amino acid motifs: (D/E)RXXQ(M/L) (E/D) in the α -subunit and NXXYADXXGNI(G/A)Y in the β -subunit. Apart from these two regions, there are some well-conserved residues, e.g., first and second residues in the β -subunit are serine, which is important for catalysis and neighboring asparagine. The total amount of highly conserved residues in PA- sequences does not exceed 7%.

Figure 3 shows the phylogenetic tree of PA-Gs. As is seen, most of these enzymes occur in proteobacteria.

Quaternary structure of PA-G.

The first 3D structure of the PA-G molecule was determined in 1995 for a PA from *E. coli* with a resolution of 1.9 Å (PDB 1PNK) [24] (Fig. 4). The active enzyme is a heterodimer with a molecular mass of 86 kDa. It consists of a smaller α -subunit (24 kDa) and a bigger β -subunit (62 kDa). At the N-terminus of the β -subunit, there is the Ser catalytic residue (marked red in Fig. 4). The average dimensions of the heterodimer are 70 x 50 x 55 Å. The polypeptide chains of the two subunits form a pyramid-like structure with a deep bowl-like cavity in the middle, with the active site located on its bottom.

Like all Ntn-hydrolases, PA-G contains the characteristic motif $\alpha\beta\beta\alpha$ [73]. PA has one $\alpha\beta\beta\alpha$ - domain which consists of α - and β -subunits [24], and the domain is one of the biggest $\alpha\beta\beta\alpha$ ones among all known Ntn-hydrolases [73].

In terms of the structure, the protein globule can be divided into three regions - A, B, and C. Region C includes eight elements of a secondary structure conserved in all representatives of Ntn-hydrolases. These elements form three antiparallel layers of four $\alpha\beta\beta\alpha$ -motifs, each. Regions A and B consist of similar elements of a secondary structure whose spatial arrangement is quite the same in all representatives of Ntn-hydrolases. However, these regions might occupy different positions in polypeptide chains of various Ntn-hydrolases. For example, in the PA-G molecule, regions A, B, and C are arranged in the following order: C-B-A.



Fig. 4. Structure of the active heterodimer of PA-G from *E. coli* (PDB 1PNK) [24]. α - and β -subunits are shown in yellow and dark blue, respectively. Catalytic β Ser1 residue and Ca^{2+} ion are shown in red and green, respectively.



Fig. 5. Structure of the active heterodimer of PA-G from *A. faecalis* (PDB 3K3W). α - and β -subunits are shown in yellow and dark blue, respectively. Catalytic β Ser1 residue and Ca^{2+} ion are shown in red and green, respectively. Disulfide bond in β -subunit between residues Cys492 and Cys525 in wild-type enzyme is shown in magenta. Insert in right part of figure shows fixation of N-terminus of α -subunit and C-terminus of β -subunit due to creation of new disulfide bond (shown in orange) after double mutation α Q3C/ β P751C.

At the end of 2009, the 3D structure of recombinant PA from *A. faecalis* (*AfPA*) expressed in *E. coli* cells was deposited into the protein data bank (PDB), although this enzyme was isolated and described early in the 1990s. In May 2010, the second structure of *AfPA* obtained from orthorhombic crystals appeared in the PDB (PDB 3ML0). *AfPA* is of particular interest due to high thermal stability compared to one for all known penicillin acylases G. An analysis of the *AfPA* structure revealed that the elevated thermal stability is associated with the presence of a disulfide bond between Cys492 and Cys525 in the β -subunit (marked pink in Fig. 5), which are absent in *EcPA*. The structures of *AfPA* and *EcPA* share a high spatial homology. A computer superposition of the structures of these two enzymes shows high spatial homology in the organization of their polypeptide chains, especially in the region of the active site. Some minor deviations are observed only in the outer regions of the enzyme molecule. It should be noted that both *AfPA* structures have a resolution only of 3.3 and 3.5 Å for PDB3K3W. ENT and PDB3ML0. ENT, respectively. This resolution is considered low in modern X-ray crystallography and does not allow to determine exactly the spatial arrangement and interactions of residues in the active site, which is important for understanding catalytic mechanism.

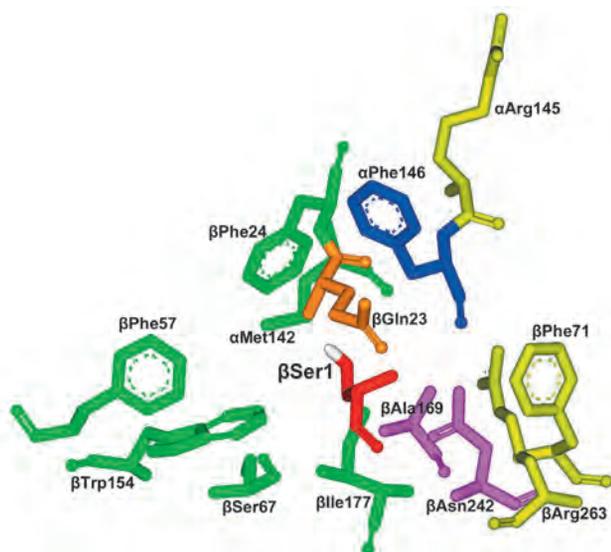


Fig. 6. Main amino acid residues in the active site of PA from *E. coli* (structure PDB 1A14). Catalytic β Ser1 residue and residue β Gln23 are shown in red and orange, respectively. Residues from oxanion hole are presented in magenta. Residues from substrate-binding domain are shown in green and yellow for subdomains S1 and S2, respectively. Residue α Phe146 belonging to both subdomains is in blue.

STRUCTURE OF *EcPA*'S ACTIVE SITE

Structure of the catalytic domain

The early studies of PA revealed the key role played by a Ser residue in catalysis. Covalent modification of this residue by phenylmethylsulfonyl fluoride (PMSF) was shown to lead to the full loss of the enzyme activity [74]. X-ray diffraction data on PA covalently modified with PMSF (PDB 1PNM) [18] allowed to determine that the catalytically important residue is the N-terminal serine of the β -chain. It is worth noting that in the vicinity of β Ser1 there are no imidazole or carboxyl groups that participate in the catalytic processes of acylation and deacylation by means of relay proton transfer and are usually found in active sites of serine proteases. Therefore, PAs were classified as so-called hydrolases with the N-terminal nucleophile, or Ntn-class [75].

Based on the structural data obtained on *EcPA*, a model of the catalytic mechanism of Ntn-hydrolases was proposed in 1995. N-terminal serine β Ser1 functions as a nucleophile (marked red in Fig. 6). Residue β Gln23 (shown in red in Fig. 6), which is positioned close to β Ser1, improves the nucleophilic properties of the catalytic serine. During the reaction, the covalent intermediate is formed through a transition state, which is stabilized by a so-called oxianionic hole. The hole is formed by residues β Asn242 and β Ala169 (marked pink in Fig. 6).

A comparison of the amino acid sequences of PA-Gs from different sources (Fig. 3) reveals that residue β Asn242 is highly conserved in all mentioned enzymes, while residue β Ala69 occurs only in the half sequences presented and Asp residue is present in this position in other enzymes.

Structure of the substrate-binding domain

The substrate-binding domain of the active site consists of two subdomains – S1 and S2. Subdomain S1 is responsible for interacting with the acyl part of the substrate. It exhibits high binding specificity and consists basically of hydrophobic residues. The “closed” structure of this domain, as well as its high hydrophobicity, makes the enzyme very selective for benzene rings with small radicals at C_{α} -atoms or aromatic rings. The main residues which form the “closed” structure of the S1 subdomain are α Met142, α Phe146, β Phe24, β Phe57, β Trp154, β Ile177, and β Ser67 (marked green in Fig. 6) and also β Pro22, β Gln23, β Val56, β Thr68, β Phe71, β Leu253, and β Phe256 (not shown in Fig. 6). As is seen, only two residues are located in the α -subunit, while most of the residues that are essential for binding the acyl part of substrate are located in the β -subunit. The sterically limited hydrophobic cavity in the “open” state has a thermodynamically unfavorable contact with the solvent. Hydrophobic interactions between the substrate and the enzyme result in a double energy gain due to, first, the energy of transfer of a hydrophobic substrate into the enzyme hydrophobic cavity from the solvent and, second, due to the exclusion of water molecules from the active site.

It is suggested that π - π interactions exist between the benzene rings of the phenylacetic acid (PAC) and β Phe24, which are located side by side. Residue α Phe146 is located at the opposite side of the hydrophobic pocket. It interacts with PAC and shields the active site from the solvent. Another phenylalanine residue, β Phe57, is located at the bottom of the hydrophobic cavity. The shortest distance between this residue and the inhibitor is 4.7 Å, which is insufficient for a direct interaction. However, this residue may be essential for the proper overall structure maintenance of the substrate-binding site, since it is located close to β Pro22 and β Phe24 (the distances are 3.5 and 3.9 Å, respectively), which directly participate in substrate binding.

The second subdomain of the substrate-binding site (S2) is responsible for the binding of the nucleophilic part. This subdomain represents the bottom of the bowl-like cavity in the middle of the molecule and accounts for the enzyme’s quite broad specificity for the nucleophilic part. Besides, this domain is enantioselective, which allows to use the enzyme for kinetic separa-

tion of the optical isomers of amines. Domain S2 consists of the following residues: α Arg145, α Phe146, β Phe71, and β Arg263 (marked yellow in Fig. 6) [76–80].

According to the analysis of the structure of the *Ec*PA inactive mutant β Asn241Ala complex with penicillin [79], at the first stage the antibiotic binds in the “open” conformation, which allows the bulky amide part of the substrate to enter the enzyme’s active site. In the “open” conformation the β Phe71 phenol ring becomes parallel to the substrate’s β -lactam ring. Moreover, substrate sorption results in Van-der-Waals interactions between the methyl group of the penicillin molecule and $C^{\delta 1}$ -atoms of α Phe146, which also strengthens the bond.

A complex of the native enzyme with a “slow” substrate penicillin G sulfoxide (PDB 1PNM) gives information about the enzyme’s structure in pre-catalytic act stage [78]. In this case, the positions of residues α Arg145 and α Phe146 correspond to the “closed” conformation, and the amide group of β Asn241 forms a hydrogen bond with O-atom of the substrate’s carbonyl group.

Results of site-directed mutagenesis of residues α Arg145 and β Arg263 show that both of them are essential for catalysis and β Arg263 is also essential for autocatalytic maturation [77]. Structural and kinetic data indicate that β Arg263 participates in the stabilization of the transition state, since its positive charge increases the polarity of the oxianionic hole. It should be noted that both β Arg263 and α Arg145 are highly conserved and present in PA-Gs of nearly all known sources (see Fig. 2).

Another special feature of PA catalytic mechanism is that the active site can exist in “open” and “closed” conformations, as was shown in [80]. The authors resolved the structure of complexes of *Ec*PA with various ligands. The “open” conformation is important for substrate binding, while effective catalysis requires protection of the active site from solvent molecules, which is achieved in the “closed conformation.”

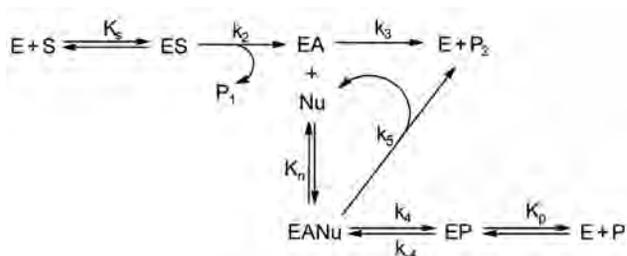
PA-G PROTEIN ENGINEERING

In the past, the main approach to increasing the efficiency of industrial processes had been to optimize conditions in the biocatalytic process. Such optimization was based on kinetic parameters and the physicochemical properties of the reaction medium’s components. This approach was successful and led to some very good results. Today, the approach has practically exhausted its potential: actual design of the enzyme is becoming increasingly popular.

X-ray diffraction studies on PA, its mutants, and complexes with different substrates have provided a prospective for understanding the structure-function relationship, as well as for rational design, of the active

site. The purpose of this design is to improve the enzyme's properties, which could increase the efficiency of PA-mediated synthesis of antibiotics and separation of enantiomers.

Because of thermodynamic constraints, the synthesis of antibiotics is based not on direct condensation but on nucleophilic substitution. The minimal scheme of the process is given below:



Scheme 1. A minimal scheme for half-synthetic β -lactam antibiotics preparation by nucleophilic substitution. E, S, and ES are the enzyme, substrate (donor of acyl moiety), and the enzyme-substrate complex, respectively. EA, Nu, and EANu are the acyl-enzyme, nucleophile, and the double complex of acyl and nucleophile. K_s , K_n , and K_p - are binding constants of substrate free enzyme E, the nucleophile with acyl-enzyme EA, and the product with free enzyme E, respectively.

Papers [81, 82] contain a detailed analysis of this scheme and its various derivatives. The basic result following from this analysis is that the description of the process's efficiency cannot be accomplished by using such common parameters as catalytic constants, Michaelis constants, or constants of inhibition by the reaction products. The parameters α , β , and γ are used instead. Parameter $\alpha = (k_{-4}/K_p)/(k_2/K_s)$ describes the specificity of the enzyme for the product and initial acyl donor, parameter $\beta = k_4/(k_3 \cdot K_n)$ reflects the maximum nucleophilic activity of acceptor, and $\gamma = k_5/k_4$ describes the ability of the triple complex EANu to undergo either synthetic or hydrolytic transformation (i.e., the ratio of the rates of synthesis and hydrolysis, S/H). Therefore, we shall use the above-mentioned parameters in estimations of the efficiency of PA engineering and in studies of PA mutant forms, as applied to antibiotic synthesis.

Production of chimeric PAs by means of DNA shuffling

At present protein engineering exploits a broad range of approaches based on the introduction of both directed and random amino acid substitutions, or even fragments of a polypeptide chain. One of these approaches, namely DNA shuffling, is now increasingly gaining in popularity and is successfully applied in the production of industri-

Table 3. Synthetic efficiency of parent and hybrid PA (15 mM D-PGA, 25 mM 6-APC, pH 7.0)

Enzyme	α	β , mM ⁻¹	γ	Pmax, mM	v_{Ps} %
PA <i>E. coli</i>	7.8	78	0.14	2.2	100
PA <i>K. cryocrescens</i>	12.7	98	0.19	2.1	110
PA <i>P. rettgeri</i>	5.8	32	0.30	1.9	59
PA 6G8	11.1	130	0.12	2.6	149
PA 73C4	11.6	120	0.12	2.4	95
PA 6B11	6.4	115	0.13	2.5	115

ally significant enzymes with enhanced properties. This method is based on a recombination of genes coding for homologous enzymes from various organisms, which results in the production of chimeric proteins that consist of fragments of original enzymes.

DNA shuffling has been applied in engineering PAs from *E. coli*, *K. cryocrescens*, and *P. rettgeri* [83]. The extent of homology of these enzymes is equal to 77% (*E. coli* - *K. cryocrescens*), 61% (*E. coli* - *P. rettgeri*), and 61% (*K. cryocrescens* - *P. rettgeri*). Screening of 81 transformants revealed three chimeric enzymes that exhibited higher catalytic activity (v_{Ps}) as well as improved synthetic properties (see Table 3).

It is noteworthy that in enzyme 6B11 all parameters were selectively improved. It was shown in an experiment that a decrease in α was caused by a two-fold increase in K_p for ampicillin (the product), while in 73C4 an increase in α was caused by a two-fold increase in K_s for D-PGA.

DNA sequencing showed that chimeric mutants 6B11 and 73C4 were obtained by inclusion of some parts of the PA gene from *K. cryocrescens* into that of *E. coli*. A certain number of random mutations that do not occur in the parent genomes and are located far from the active site were also described. However, these mutations can have a substantial effect on the enzyme's properties. For example, the substitution of β Gly385, which is positioned 13.5 Å from the active site, by serine residue (β Gly385Ser) in the hybrid 6G8 can affect the availability of the active site. To study the influence of such substitution on the catalytic properties of enzyme, the authors introduced the corresponding mutation into *EcPA*. The results were amazingly unpredictable: the authors reported a 22% increase in the maximum yield and a 80% increase in the S/H ratio compared to the wild type enzyme under the same conditions. The parameter α did not change. Similarly, random mutation of α Asp148Gly works in the same way, i.e. it results in a 20% increase in the maximum yield and an 80% increase in the S/H ratio. In the case of this amino acid change the observed effect can be

Table 4. Catalytic efficiency of mutant *EcPA* in reactions of antibiotic synthesis [84]

Enzyme	Acyl donor	Ampicillin			Cephalexin		
		α	β , mM ⁻¹	1/ γ	α	β , mM ⁻¹	1/ γ
PAS2	ΦGA	13.2	0.5	6	7.3	0.5	59
	ΦGM	16.4	0.5	6	9.1	0.5	59
αR160K αF161L βF24M	ΦGA	218	6.3	364	58.8	18.4	69
	ΦGM	8.8	6.3	364	2.4	18.4	69
βF24A	ΦGA	209	10.2	286	88.8	21.5	154
	ΦGM	11.6	10.2	286	4.9	21.5	154
αR160P αF161A βF24A	ΦGA	200	4.9	167	57.2	21.1	58
	ΦGM	7.6	4.9	167	2.2	21.1	58

explained by the fact that the αAsp148 is located at the base of the catalytic α-loop, which is responsible for the conformational transitions in the active site. As it was shown for residues αArg145 and αPhe146, any changes in this loop can lead to the alteration of both the substrate specificity and catalytic activity.

Thus, DNA shuffling can be employed not only as a tool in the production of enhanced versions of PA, but also as a promising approach to the design of site-directed mutations that are impossible to predict based on knowledge of the enzyme's 3D structure.

Combinatory saturating mutagenesis

In work [84], saturating mutagenesis was combined with the introduction of random mutations and applied to residues αR160, αF161, and βF24, which were juxtaposed to the enzyme's active site (single, double, as well as triple mutants were obtained). Different amino acid changes were introduced in these positions with degenerate primers, and a library of 700 various clones was obtained. HPLC-based screening was used to test mutants for their ampicillin synthetic activity. As a result, three mutants with improved catalytic properties were selected (Table 4). Moreover, one of the mutants (βF24A) also exhibited good catalytic properties in the reaction of cephalexin synthesis.

It should be noted that the method of random mutagenesis has not found broad application, because of the absence of a simple mutant screening procedure. At the same time, directed mutagenesis has turned out to be the most popular method.

RATIONALE DESIGN

Acyl-binding domain

Processes of preparation of β-lactam cores by hydrolysis of natural antibiotics will be more efficient in the

case of reduced affinity of the enzyme to phenylacetic acid, which is a byproduct of the reaction and a competitive inhibitor of PA. A requirement for effective synthesis of half-synthetic unnatural antibiotics is a high specificity of the enzyme for the FAA derivatives carrying substitutions at C_α atom (D-phenylglycine, phenylglycolic acid, C_α-methyl-FAA). In case of natural PAs, the values of the Michaelis constants for FAA derivatives carrying substitutions at C_α are about 10-100 mM, which is 100-1,000 times higher than for FAA [85]. Therefore, production of mutant enzymes with low values of the Michaelis constant for FAA derivatives carrying substitutions at the C_α atom and high inhibition constants for FAA is of great interest.

Paper [86] describes point mutations βPhe24Ala and αPhe146Tyr and their combination βPhe24Ala/αPhe146Tyr meeting these requirements. For example, in case of mutant *EcPA* βPhe24Ala, FAA inhibition was 18 times lower in comparison with the wild-type enzyme ($K_p = 1.1$ and 0.06 mM, respectively), but affinity to C_α-substituted FAA derivatives was 5 times higher. Constructing the mutants, the authors suggested that substitution of the hydrophobic residue that forms the base of the acyl-binding domain could dramatically change the enzyme's specificity for both C_α-substituted FAA derivatives and the natural substrate.

X-ray analysis of complexes of the mutants with FAA derivatives provided structural explanations of the observed phenomena. Removal of the aromatic ring in the mutant βF24A leads to structural perturbations in the active site. As a result, in this mutant FAA binds in the "open" conformation, as opposed to the wild-type enzyme in which FAA binds in the "closed" conformation. This, in its turn, results in the disappearance of hydrogen bonds between the carbonyl O-atom in FAA and residues βAla69 and βAsn241 that form the oxian-

ionic hole. This explains the multi-fold increase in the value of the inhibition constant.

Unlike FAA itself, its derivative C_α-methyl FAA binds in the “closed” conformation in the mutant *EcPA* βF24A. Simultaneously, C_α-CH₃ takes the place of the removed phenyl group, which leads to a 10-fold increase in binding.

Notably, the mutation of βPhe24 strongly affects the enzyme stereospecificity for (*R*)-isomers of C_α-substituted FAA derivatives. Therefore, βPhe24 can be a potential target for alteration of the enzyme enantioselectivity, which can be used for obtaining optically pure C_α-substituted FAA derivatives.

Substitution of phenylalanine by tyrosine in the αF146Y mutant does not cause conformational rearrangements in the active site, which explains the FAA binding comparable to that in wild type *EcPA*. However, the van der Waals interactions between a C_α-group of FAA derivatives and OH-group of αTyr146 make binding more efficient.

In work [87], the natural specificity of *EcPA* was altered from penicillini to cephalosporin C, which gave the opportunity to create a *EcPA*-based biocatalyst for preparation of 7-aminocephalosporinic acid directly from cephalosporine C. An alignment of the amino acid sequences of *EcPA* and cephalosporinacylase (CA) from *P. diminuta* was performed in order to determine the binding site for the aromatic hydrophobic acyl part of penicillin in PA and that for the linear hydrophilic (glutaryl) acyl part of cephalosporin C in CA. As a result, seven residues from the hydrophobic pocket of *EcPA* were substituted by the corresponding residues from CA. The obtained mutant had a specificity for cephalosporin C 8-fold higher than wt-*EcPA*. Unfortunately, these results are of little practical use aside from scientific interest, since the activity of the prepared mutant enzyme was much lower than that of the native CA.

Site-directed mutagenesis of the nucleophile-binding site

Increasing the acyl part-binding properties is insufficient for improving the enzyme synthetic efficiency. For example, for mutant *EcPA* βPhe24Ala it was shown that increased D-phenylglycinamide binding leads to an increase in product (ampicillin) binding, which finally elevates parameter α. Interestingly, substitutions of residues βPhe24 and αPhe146, which are responsible for binding the acyl moiety, also account for a two-fold increase in the nucleophile properties of 6-APA.

The fact that alterations of the protein structure in a particular region can cause structural rearrangements in adjacent regions does not contradict the basic principles of protein molecule organization. Mutations in the acyl-binding part (S1) could actually cause structural changes

Table 5. Kinetic parameters for ampicillin synthesis for wild-type enzyme and mutant *EcPA* with substitution of αArg145 (15mM D-PGA, 25mM 6-APC, pH 7.0)

Enzyme	α	β, mM ⁻¹	γ	P _{max} , mM
<i>EcPA</i> wild-type	7.7	80	0.14	2.2
<i>EcPA</i> αArg145Gly	29	420	0.04	3.6
<i>EcPA</i> αArg145Ser	14	350	0.05	3.3
<i>EcPA</i> αArg145Leu	15	280	0.06	2.8

in the nucleophile-binding site (S2) so long as residues βPhe24 and αPhe146 are located in the border region of S1 and S2 subdomains and these residues produce the hydrophobic neck of acyl-binding domain (Fig. 6).

On the other hand, it is impossible to determine whether the increase in nucleophile binding (K_n) or decrease in water reactivity (k₃, k₅) plays the key role. βPhe24Ala substitution could affect the spatial position of the neighboring βGlu23, which coordinates the deacylating water molecule and, therefore, decreases the rate of hydrolysis. However, using available data, it was shown that improving the enzyme nucleophilic specificity causes an increase in β: so the effects of point mutations in the nucleophile-binding site are rather predictable. In work [89], the authors achieved a substantial increase in parameter β by mutations of αArg145, which interacts with the ampicillin carboxylic group in the closed conformation and can affect 6-APA binding, according to X-ray analysis. In order to study all the possible effects of a substitution of αArg145, this residue was subjected to saturating mutagenesis. Preliminary results of ampicillin synthesis in diluted solutions showed that all 19 mutants have an increased ability to transfer an acyl group to 6-APA. Three of the most promising mutants of *EcPA* (αArg145Leu, αArg145Gly and αArg145Ser) were selected and subjected to a detailed kinetic study (table 5).

The obtained results show that the removal of a positive charge in position α145 leads to a substantial increase in the nucleophilic properties of 6-APA. A nearly synchronized increase in α leads to the conclusion that elevated nucleophile binding (K_n, see the scheme above), expressed as an increase in β, inevitably results in higher affinity to the reaction product (K_p, see the scheme above), and it is impossible to improve both parameters simultaneously. However, the negative effects on α can be mitigated to some extent if ampicillin is removed from the reaction medium.

According to X-ray diffraction data, the residue αPhe146 interacts with the methyl group of penicillin

upon its binding in the open conformation and could affect the binding of the β -lactam ring of 6-APA. In order to clarify the role of this residue, saturation mutation analysis was also carried out. Preliminary studies of enzymatic ampicillin synthesis showed that all 19 mutants have an increased ability to transfer the acyl group to 6-APA, but the catalytic activity of all mutants decreases at least 10-fold, which did not allow the authors to study the synthetic abilities of the mutants in detail.

Work [90] represents a good example of employing rational design for improving *EcPA* catalytic properties. The authors performed a structural analysis and molecular modeling of ligand binding in the active site of the enzyme. The results showed that mutation of β Phe71 alters the substrate's orientation and increases the interaction of substrates with the oxianionic hole. Therefore, it would be reasonable to expect the mutation of β Phe71 to increase the PA catalytic activity and alter both the stereoselectivity and specificity of the enzyme for the leaving group. Four mutants with amino acid changes β Phe71Lys, β Phe71Glu, β Phe71Leu, and β Phe71Arg were prepared. Three of them showed improved catalytic properties, and two mutants had higher stereoselectivity in terms of separation of enantiomers; however, the effect of an introduction of a particular amino acid was quite opposite. For example, the β Phe71Leu mutant had the best catalytic properties in the antibiotics synthesis reactions, while its enantioselectivity was lower than that of the wild type enzyme. The most pronounced increase in enantioselectivity was observed for the mutant *EcPA* β Phe71Lysones for the mutant *EcPA* β Phe71Leu. Based on the obtained data, one made the conclusion that a simultaneous improvement of all basic catalytic properties of the enzyme seems to be impossible. Evidently, particular tasks could be solved only by individual fine-tuning of the enzyme based on the substrate's structure.

Obtaining a permuted enzyme

Another approach to alter the enzyme's properties is preparation of so-called permuted enzymes. As seen in Figs. 4 and 5, in PAs from *E. coli* and *A. faecalis*, the N-terminus of the α -subunit is placed in juxtaposition with the C-terminus of the β -subunit, which allows to obtain a permuted one-chain enzyme by linking both termini by a flexible spacer. Engineering such an enzyme would allow to obtain the recombinant protein directly in the cytoplasm without transportation of the precursor to the periplasmic compartment and cleavage of the inter-subunit spacer, since these stages were shown to be the main reasons of the low yield of the enzyme produced in available systems of PA-G expression. To this day, there is only one work that describes preparation of the permuted enzyme [91]. A random

sequence of 4 amino acid residues was chosen as a linker between the subunits. A library of mutant clones carrying the gene of the permuted enzyme was created. Subsequent screening of the clones revealed 20 clones that produced the active enzyme; however, all the mutants were much less efficient in terms of specific activity as compared to the wild type enzyme.

A/PA stabilization by formation of disulfide bonds

As it was described earlier, PA from *A. faecalis* is one of the most thermostable penicillin acylases known to this day [20], and this remarkable property is attributed to the presence of a disulfide bond in the β -subunit of A/PA (Fig. 5). Additional disulfide bonds were introduced into the enzyme's globule in order to further stabilize the molecule [92]. Since 2006, when the experiments were performed, the structure of the enzyme has not been determined. The authors constructed a computer model using homology modeling based on the structure of *EcPA*. The search for possible sites of disulfide location was performed with the MODIP program. The computer analysis resulted in 32 possible variants of disulfide bonds in different positions. An additional analysis yielded two final variants - pairs of substitutions α Q3C/ β P561C and α T52C/ α Y64C. The first pair of substitutions should lead to the formation of the covalent bond between the N-terminus of the α -subunit and the C-terminus of the β -subunit (see inset in Fig. 5), which is basically unstructured (Fig. 5). In the case of the second pair of amino acid changes, the disulfide bond should stabilize a loop that connects two α -helices. The analysis of the properties of the obtained mutants showed that introduction of the point mutations α Glu3Cys and β Pro561Cys did not affect the rate of thermal inactivation, while in the case of the double substitution, the temperature at which the enzyme loses 50% of its activity after 20 minutes of incubation increased from 50°C to 53°C. The second pair of substitutions did not affect the enzyme's stability, which suggests that the stability of the two α -helices is not critical for the stability of the whole protein globule. This suggestion corroborates recent X-ray diffraction data, which show that the mentioned α -helices form additional contacts with the structural elements of the β -subunit.

CONCLUSIONS

Despite the large volume of work on the engineering of penicillin acylases, the problem of altering the catalytic properties and stability of PAs is far from being solved. The data presented show that the catalysis is affected by residues located at 15-20 Å around the catalytic site of the enzyme. Therefore, a detailed investigation of the roles of amino acid residues located

in the active site in catalysis and substrate specificity, as well as gaining insight into how they affect catalytic parameters, appears to be important for both biotechnology and fundamental science. It's one's hope that isolation and investigation of penicillin acylases from

other sources could reveal the basic principles of the "structure-function" relationship in this enzyme. ●

The work was supported by the Russian Foundation for Basic Research (grant № 10-04-01334-a)

REFERENCES

1. Sakaguchi K., Murao S. // J. Agr. Chem. S 1. oc. Japan. 1950. V. 23. P. 1–3.
2. Claridge C.A., Luttinger J.R., Lein J. // Proc. Soc. Exp. Biol. Med. 1963. V. 113. P. 1008–1012.
3. Hamilton-Miller J.M. // Bacteriol. Rev. 1966. V. 30. P. 761–771.
4. Valle F., Balbas P., Merino E., et al. // Trends Biochem. Sci. 1991. V. 16. P. 36–40.
5. Arroyo M., de la Mata I., Acebal C., et al. // Appl. Microbiol. Biotechnol. 2003. V. 60. P. 507–514.
6. Rajendhran J., Gunasekaran P. // J. Biosci. Bioeng. 2004. V. 97. P. 1–13.
7. Polderman-Tijmes J.J. Biochemical characterization of alpha-amino acid ester hydrolases. Ph.D. Thesis. Groningen: University of Groningen, 2004.
8. Deshpande B.S., Ambedkar S.S., Sudhakaran V.K., et al. // World J. Microbiol. Biotechnol. 1994. V. 10. P. 129–138.
9. Murao S. // J. Agr. Chem. Soc. Japan. 1955. V. 29. P. 404–407.
10. Batchelor F.R., Doyle F.P., Nayler J.H.C., et al. // Nature. 1959. V. 183. P. 257–259.
11. Vanderhaeghe H. // Methods Enzymol. 1975. V. 43. P. 721–728.
12. Shewale J.G., Deshpande B.S., Sudhakaran V.K., et al. // Process Biochem. Int. 1990. V. 25. P. 97–103.
13. Shewale J.G., Sudhakaran V.K. // Enzyme Microb. Technol. 1997. V. 20. P. 402–410.
14. Bruggink A., Roos E.C., De Vroom E. // Org. Process. Res. Dev. 1998. V. 2. P. 128–133.
15. Švyadas V.K., Klesov A.A., Nys P.S. // Antibiotiki (Antibiotics). 1976. V. 21. P. 698–704
16. Pundle A., SivaRaman H. // Curr. Microbiol. 1997. V. 34. P. 144–148.
17. Sudhakaran V.K., Shewale J.G. // Hindustan Antibiot. Bull. 1995. V. 37. P. 9–15.
18. Hewitt L., Kasche V., Lummer K., et al. // J. Mol. Biol. 2000. V. 302. P. 887–898.
19. Savidge T.A., Cole M. // Methods Enzymol. 1975. V. 43. P. 705–721.
20. Verhaert R.M., Riemens A.M., van der Laan J.M. // Appl. Environ. Microbiol. 1997. V. 63. P. 3412–3418.
21. Cheng T., Chen M., Zheng H. // Protein. Expr. Purif. 2006. V. 46. P. 107–113.
22. Kim D.J., Byun S.M. // Biochim. Biophys. Acta. 1990. V. 1040. P. 12–18.
23. Schumacher G., Sizmann D., Haug H 23. , et al. // Nucleic. Acids Res. 1986. V. 14. P. 5713–5727.
24. Duggleby H.J., Tolley S.P., Hill C.P., et al. // Nature. 1995. V. 373. P. 264–268.
25. Kasche V., Lummer K., Nurk A., et al. // Biochim. Biophys. Acta. 1999. V. 1433. P. 76–86.
26. Lindsay C.D., Pain R.H. // Eur. J. Biochem. 1990. V. 192. № 1. P. 133–141.
27. Economou A. // FEBS Lett. 2000. V. 476. P. 18–21.
28. Chou C.P., Tseng J.H., Kuo B.Y., et al. // Biotechnol. Prog. 1999. V. 15. P. 439–445.
29. Berks B.C. // Mol. Microbiol. 1996. V. 22. P. 393–404.
30. Ignatova Z., Hornle C., Nurk A., et al. // Biochem. Biophys. Res. Commun. 2002. V. 291. P. 146–149.
31. Kasche V., Ignatova Z., Markl H., et al. // Biotechnol. Prog. 2005. V. 21. P. 432–438.
32. Chou C.P., Yu C.C., Tseng J.H., et al. // Biotechnol. Bioeng. 1999. V. 63. P. 263–272.
33. Cai G., Zhu S., Yang S., et al. // Appl. Environ. Microbiol. 2004. V. 70. P. 2764–2770.
34. Zhang L.F., Li Z.W., Zhang Q.J. // Chin. J. Biotechnol. 1991. V. 7. P. 63–72.
35. Martin L., Prieto M.A., Cortes E., et al. // FEMS Microbiol. Lett. 1995. V. 125. P. 287–292.
36. Polderman-Tijmes J.J., Jekel P.A., de Vries E.J., et al. // Appl. Environ. Microbiol. 2002. V. 68. P. 211–218.
37. Kang H.K., Park J.Y., Ahn J.S., et al. // J. Microbiol. Biotechnol. 2009. V. 19. P. 172–177.
38. Lu Y., Chi X., Yang Q., et al. // Extremophiles. 2009. V. 13. P. 875–884.
39. Lu Y., Zhao H., Zhang C., et al. // Biotechnol. Lett. 2009. V. 31. P. 1525–1530.
40. Madhavan A., Tamalampudi S., Ushida K., et al. // Appl. Microbiol. Biotechnol. 2009. V. 82. P. 1067–1078.
41. Karimi M., De Meyer B., Hilson P. // Trends. Plant. Sci. 2005. V. 10. P. 103–105.
42. Karimi M., Depicker A., Hilson P. // Plant. Physiol. 2007. V. 145. P. 1144–1154.
43. Sando T., Takeno S., Watanabe N. // Biosci. Biotechnol. Biochem. 2008. V. 72. P. 2903–2917.
44. Cao Y., Lu Z., Sun P., et al. // Wei Sheng Wu Xue Bao. 2008. V. 48. P. 312–316.
45. Khodabandehloo M., Shamsi Shahrabadi M., Keyvani H. // Iran Biomed. J. 2009. V. 13. P. 9–18.
46. Liu Z., Yang G., Li B. // Mol. Biotechnol. 2003. V. 24. P. 21–26.
47. Liu Z., Yang G.Z., Chi C.W., et al. // Protein. Pept. Lett. 2002. V. 9. P. 419–426.
48. Farrokhi N., Hrmova M., Burton R.A., et al. // Methods Mol. Biol. 2009. V. 513. P. 175–198.
49. Schwarz D., Dotsch V., Bernhard F. // Proteomics. 2008. V. 8. P. 3933–3946.
50. Schwarz D., Junge F., Durst F., et al. // Nat. Protoc. 2007. V. 2. P. 2945–2957.
51. Spirin A.S. // Bioorgan. khimia (Bioorganic Chemistry). 1992. V. 18. P. 1394–1402.
52. Oh S.J., Kim Y.C., Park Y.W., et al. // Gene. 1987. V. 56. P. 87–97.
53. Chou C.P., Kuo B.Y., Lin W.J. // J. Biosci. Bioeng. 1999. V. 88. P. 160–167.
54. Ohashi H., Katsuta Y., Nagashima M., et al. // Appl. Environ. Microbiol. 1989. V. 55. P. 1351–1356.
55. Daumy G.O., Williams J.A., McColl A.S., et al. // J. Bacteriol. 1986. V. 168. P. 431–433.
56. Barbero J.L., Buesa J.M., Gonzalez de Buitrago G., et al. // Gene. 1986. V. 49. P. 69–80.
57. Meevootisom V., Somsuk P., Prachaktam R., et al. // Appl. Environ. Microbiol. 1983. V. 46. P. 1227–1229.

REVIEWS

58. Dai M., Zhu Y., Yang Y., et al. // *Eur. J. Biochem.* 2001. V. 268. P. 1298–1303.
59. Ignatova Z., Mahsunah A., Georgieva M., et al. // *Appl. Environ. Microbiol.* 2003. V. 69. P. 1237–1245.
60. Jiang Y.M., Tong W.Y., Wei D.Z. // *Biotechnol. Prog.* 2007. V. 23. P. 1031–1037.
61. Norris V., Grant S., Freestone P., et al. // *J. Bacteriol.* 1996. V. 178. P. 3677–3682.
62. Scherrer S., Robas N., Zouheiry H., et al. // *Appl. Microbiol. Biotechnol.* 1994. V. 42. P. 85–91.
63. Xu Y., Rosenkranz S., Weng C.L., et al. // *Appl. Microbiol. Biotechnol.* 2006. V. 72. P. 529–536.
64. Berks B.C., Sargent F., Palmer T. // *Mol. Microbiol.* 2000. V. 35. P. 260–274.
65. Narayanan N., Xu Y., Chou C.P. // *Biotechnol. Prog.* 2006. V. 22. P. 1518–1523.
66. Xu Y., Weng C.L., Narayanan N., et al. // *Appl. Environ. Microbiol.* 2005. V. 71. P. 6247–6253.
67. Maresova H., Stepanek V., Kyslik P. // *Biotechnol. Bioeng.* 2001. V. 75. P. 46–52.
68. Sriubolmas N., Panbangred W., Sriurairatana S., et al. // *Appl. Microbiol. Biotechnol.* 1997. V. 47. P. 373–378.
69. Vohra P.K., Sharma R., Kashyap D.R., et al. // *Biotechnol. Lett.* 2001. V. 23. P. 531–535.
70. De Leon A., Breceda G.B., Barba de la Rosa A.P., et al. // *Biotechnol. Lett.* 2003. V. 25. P. 1397–1402.
71. Huang S.W., Lin Y.H., Chin H.L., et al. // *Biotechnol. Prog.* 2002. V. 18. P. 668–671.
72. Deak P.M., Lutz-Wahl S., Bothe H., et al. // *Biotechnol. Lett.* 2003. V. 25. P. 397–400.
73. Oinonen C., Rouvinen J. // *Protein. Sci.* 2000. V. 9. P. 2329–2337.
74. Švyadas V.K., Margolin A.L., Sherstuk S.F., et al. // *Bioorgan. khimia (Bioorgan. Chemistry)*. 1977. V. 3. P. 546–554.
75. Brannigan J.A., Dodson G., Duggleby H.J. // *Nature*. 1995. V. 378. P. 416–419.
76. Alkema W.B., Hensgens C.M., Snijder H.J., et al. // *Protein. Eng. Des. Sel.* 2004. V. 17. P. 473–480.
77. Alkema W.B., Prins A.K., de Vries E., et al. // *Biochem. J.* 2002. V. 365. P. 303–309.
78. McVey C.E., Walsh M.A., Dodson G.G., et al. // *J. Mol. Biol.* 2001. V. 313. P. 139–150.
79. Alkema W.B., Hensgens C.M., Kroezinga E.H., et al. // *Protein. Eng.* 2000. V. 13. P. 857–863.
80. Done S.H., Brannigan J.A., Moody P.C., et al. // *J. Mol. Biol.* 1998. V. 284. P. 463–475.
81. Gololobov M.Yu., Borisov I.L., Švedas V.K. // *J. Theor. Biol.* 1989. V. 140. P. 193–204.
82. Youshko M.I., Bukhanov A.L., Švedas V.K. // *Biochemistry (Moscow)*. 2003. V. 68. № 3. P. 334–338.
83. Jager S.A.W., Jekel P.A., Janssen D.B. // *Enzyme Microb. Technol.* 2007. V. 40. P. 1335–1344.
84. Gabor E.M., Janssen D.B. // *Protein Eng. Des. Sel.* 2004. V. 17. P. 571–579.
85. Margolin A.L., Svedas V.K., Berezin I.V., et al. // *Biochim. Biophys. Acta*. 1980. V. 616. P. 283–289.
86. Guranda D.T., van Langen L.M., van Rantwijk F., et al. // *Tetrahedron. Asym.* 2001. V. 12. P. 1645–1650.
87. Oh B., Kim K., Park J., et al. // *Biochem. Biophys. Res. Commun.* 2004. V. 319. P. 486–492.
88. Alkema W.B., Dijkhuis A.J., De Vries E., et al. // *Eur. J. Biochem.* 2002. V. 269. P. 2093–2100.
89. Jager S.A., Shapovalova I.V., Jekel P.A., et al. // *J. Biotechnol.* 2008. V. 133. P. 18–26.
90. Shapovalova I.V., Alkema W.B.L., Jamskova O.V., et al. // *Acta Naturae*. 2009. V. 1. № 3. P. 94–98.
91. Flores G., Soberon X., Osuna J. // *Protein Sci.* 2004. V. 13. P. 1677–1683.
92. Wang T., Zhu H., Ma X., et al. // *Protein. Pept. Lett.* 2006. V. 13. P. 177–183.

PEDF – A Noninhibitory Serpin with Neurotrophic Activity

N.I. Minkevich, V.M. Lipkin, I.A. Kostanyan*

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences

*E-mail: iakost@mx.ibch.ru

Received 29.06.2010

ABSTRACT The pigment epithelium-derived factor (PEDF) is a glycoprotein with a molecular weight of 50 kDa belonging to the noninhibitory serpin family. It regulates several physiological processes, such as stimulation of retinoblastoma cell differentiation into neuron cells, and facilitation of the growth and viability of photoreceptor cells and neurons of the central nervous system. Moreover, this factor protects neuronal cells against apoptosis. PEDF is not only a neurotrophic factor, but also a natural angiogenesis inhibitor. This protein, as well as its biologically active fragments, possesses significant neuroprotective, neurotrophic, and antiangiogenic capabilities. The precise molecular mechanisms underpinning the effects of PEDF are still not quite clear. However, this protein generates great interest as a promising drug for the therapy of a wide range of neurodegenerative, ophthalmological, and oncological diseases. This review is a summary of what is known today about the structural features, biochemical properties, and multimodal functions of PEDF.

KEYWORDS PEDF (pigment epithelium-derived factor), serpin, neurotrophic factor, angiogenesis.

INTRODUCTION

PEDF (Pigment Epithelium-Derived Factor) was identified first in 1989 by a group of American researchers led by L.V. Johnson. It was purified from the growth medium of a culture of retinal pigment epithelium cells extracted from a human fetus [1, 2]. This growth medium was found to stop unlimited cell division and growth in Y-79 retinoblastoma cells. Moreover, 90% of the cells acquired the morphological and biochemical characteristics of mature neurons [1]. Electrophoretic and chromatographic means allowed to purify a 50 kDa protein from the growth medium, which was responsible for the observed effects on Y-79 cells. As was shown later, human PEDF consists of a single polypeptide chain that is 418 a.a. long (Fig. 1) [3].

Practically at the same time, PEDF was extracted from a culture of human lung fibroblast-like cells (WI-38) [4]. The murine analog of PEDF, caspin, was also discovered [5]. Human PEDF and caspin cDNA are 82.7% identical, whereas the amino acid sequences of the two are 85.6% identical.

Current research shows that PEDF is expressed in almost all mammalian and avian tissues and has numerous important functions. These include a differentiating effect on embryo and tumor cells, a protective effect on mature neurons and other neural tissue cells, and antiangiogenesis activity – the ability to inhibit the formation of new blood vessels. This is why this protein is of major interest as a regulatory factor of cell and tissue growth.

PEDF PRIMARY STRUCTURE AND ITS ENCODING GENE

The human PEDF gene and the encoded protein

The *pedf* gene is localized on the short arm of chromosome 17 in its distal region – 17p13.3 [6, 7]. It is approximately 15 kbp long and includes 8 exons interspersed with 7 introns. The exon-intron contacts are all in accordance with the GC/AG consensus. The length of the exons varies from 92 to 377 bp. The promoter region of the gene is located in the 5'-flanking region, and the CAAT-box is located in the -43 position from the transcription initiation site. A cluster of Alu-repeats which includes 8 complete and 3 partial Alu-repeats is located in the 5'-terminal region of *pedf*, up to -5 kbp [8].

The *pedf* gene encodes a sequence of 418 a.a. [3]. The N-terminal region of PEDF, right after the initiatory methionine, includes a 17 a.a. stretch of hydrophobic amino acids, with a sequence characteristic of a signaling fragment that targets the protein for secretion [3, 9]. However, according to some reports, the sequence of mature PEDF secreted from producing cells begins with Asn21, and according to these data the polypeptide chain consists of 398 a.a. [10, 11]. Posttranslational modifications of blood serum PEDF include an N-terminal residue of pyroglutamic acid, which is formed via deamination of the Gln2 residue, and the canonic N-glycosylation site Asn-Leu-Thr in positions 285-287. The added polysaccharides are heterogenic and have a molecular weight of approximately 4 kDa; their presence explains the difference between the calculated

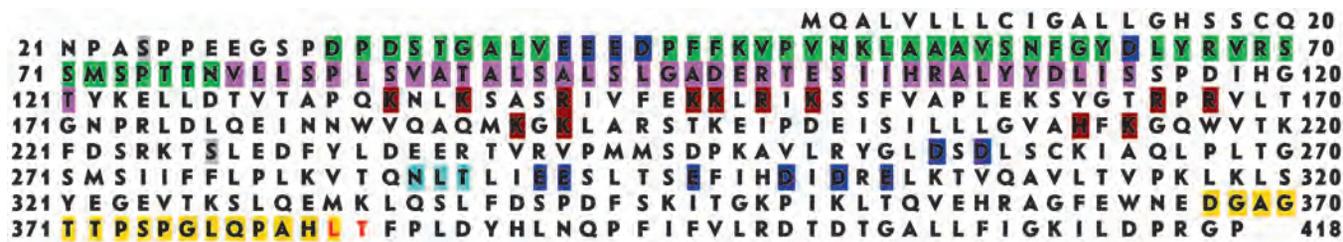


Fig. 1. Amino acid sequence of the PEDF protein. The bond Leu382-Thr383 is marked in bold red. Yellow highlights mark the serpine loop, and light blue marks the N-glycosylation site. Green – peptide responsible for the antiangiogenic function; pink – peptide responsible for the neurotrophic function. Grey – phosphorylation sites. Blue – cluster of negatively charged amino acids, brown – positively charged amino acids.

and actual masses of the factor [12]. According to 2D electrophoresis data, human PEDF secreted by pigment epithelium cells has 4 isoforms, whose isoelectric points are 6.0; 6.2; 6.4 and 6.6 [13].

PEDF is a member of the noninhibitory serpine family

Based on the analysis of the primary structure of the protein and corresponding cDNA, PEDF was included into the serpine family (serine proteases inhibitors) [14]. Amino acid sequences of the factor and α 1-antitrypsin are 27% identical (42% if equivalent substitutions are factored in), and 27 and 26% identical with α 1-antichemotrypsin and α 2-plasmin inhibitor, respectively (44 and 43% if equivalent substitutions are factored in) (Fig.2) [3]. It is known that 51 a.a. residues are necessary to form a spatial structure of the serpine type [15], in PEDF there are 39 coincidences (76%) [16]. Despite the high level of similarity with the primary and spatial structures of serine protease inhibitors, PEDF does not inhibit these enzymes [17]. This may possibly be due to the structural characteristics of the exposed reactive loop of the factor. Even though it is equal in length to other serpines (17 a.a. residues Glu366-Leu382) [18,19], its content lacks some elements that are characteristic of inhibitor serpines – Ala-rich sequences (in the 9th and 12th loop positions) and Thr in positions 8 and 14, while also possessing proline residues, which is not characteristic of serpines [16].

The structure of PEDF is similar to other serpines in that it possesses a bond that is especially prone to proteolysis in a loop exposed into the cytoplasm [14], as do other serpines. Limited proteolysis of PEDF with various serine proteases (chemotrypsin, elastase, subtilisin, etc.) reduces its molecular mass by approximately 4 kDa. It has been determined that these proteases hydrolyze the Leu382-Thr383 bond, which results in the release of a 36-a.a. residue fragment from the C-terminus. Unlike inhibitory serpines, no covalent bond is formed between the PEDF molecule and the active site

of the proteinase. PEDF with a truncated C-terminus conserves its ability to differentiate Y-79 cells into mature neurons [20].

After the release of the 36-a.a. fragment, the general ellipticity of the molecule is reduced, indicating that the molecule experiences a reduction in the α -helix structure. Unlike inhibitory serpines, this process lowers the thermal stability of the PEDF molecule. However, these conformational alterations are not widespread and involve only the C-terminal region located near the exposed loop. The functional role of the latter and the



Fig. 2. Comparison of the amino acid sequences of human PEDF and α 1-antitrypsin (HUMA1AT). Colons (:) depict identical amino acids, and periods depict conserved substitutions [3].

released C-terminal fragment of PEDF remain largely unstudied. Verified data demonstrate that the residues present in this region of the molecule, in conjunction with the N-terminal signaling peptide, are needed for the secretion of PEDF from the cell.

Removal of the Pro415-Pro418 and Pro373-Ala380 fragments, as well as point substitutions of Gly376, Leu377, Pro393, Phe394, or Phe396 (the first two are located in the exposed loop), either results in changes of the spatial structure of the molecule or alters its interaction with other proteins, presumably transporters. PEDF transfer from the endoplasmic reticulum into the Golgi apparatus is disrupted, which prevents PEDF secretion [16, 21].

Spatial structure of PEDF. Asymmetrical localization of positive and negative charges in the PEDF molecule

A mature molecule of the factor is a globule with a radius of <3.05 nm. Approximately 60% of PEDF amino acid residues are involved in the formation of 10 α -helices and 3 β -sheets [22]. Ser24, Ser114, and Ser227 are phosphorylation sites. The factor includes two important amino acid sequences – a 34 a.a stretch (Asp34-Asn77) and a 44 a.a. stretch (Val78-Thr121) – which are responsible for the antiangiogenic and neurotrophic interactions, respectively (Fig. 3) [23].

An interesting feature of the PEDF molecule is the presence of positively and negatively charged amino acid clusters, located on opposite sides of the globule. These clusters allow the factor to bind with structural elements of the extracellular matrix, namely glycosaminoglycans and collagen, via electrostatic bonds [24, 25].

The cluster of negatively charged residues (Asp and Glu) is located in two regions of the polypeptide chain, which are exposed on the surface of the globule, adjacent to each other: Glu41-Asp64 (residues Glu41, Glu42, Glu43, Asp44, and Asp64) and Asp256-Glu304 (residues Asp256, Asp258, Glu290, Glu291, Glu296, Asp300, and Glu304). The presence of this cluster determines PEDF affinity to type I, II, and III collagens [25]. PEDF binding to collagens, as well as to glycosaminoglycans, is weakened in high ionic strength solutions, since the bonds are electrostatic in their nature. The peptide fragments Val40-Arg67 and Phe277-Ile301, which are also involved in the formation of this cluster, are conserved from fish to mammals [16].

The cluster of positively charged amino acids, which includes 12 residues of Arg, Lys, and His, is located in the Lys 134-Lys214 region. Residues Lys134, Lys137, Lys189, Lys191, His212, and Lys214 form the “core” of the cluster, while residues Arg141, Lys146, Lys147, Arg149, Arg165, and Arg167 are located on the periphery. This local positive charge allows the binding

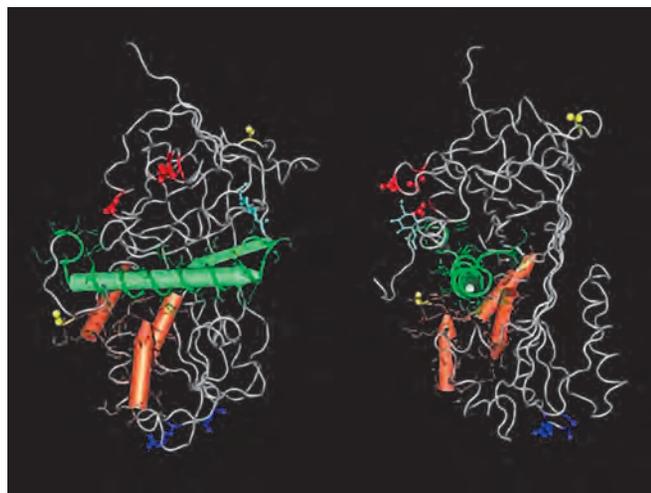


Fig. 3. 3D crystal structure of PEDF. This figure shows two projections on the vertical axis with a shift of 90°. The peptide responsible for the antiangiogenic function is marked in green, and the one responsible for the neurotrophic function is marked in orange. Red marks the negatively charged amino acids which form the collagen-binding cluster, and purple marks the positively charged amino acids which form the “core” of the glycosaminoglycan-binding cluster. Phosphorylation sites (Ser114 and Ser227) are marked in yellow, and the glycosylation site at Asn285 with the bound sugar residues is marked in light blue [23].

of PEDF to polyanionic molecules which are rich in charged sulfate groups, such as heparin, chondroitin sulfate, etc. This cluster is also involved in the interaction between PEDF and the glycosaminoglycans in the interphotoreceptor matrix of the retina [24].

Apart from the cytoplasm, PEDF can also be found inside the nucleus. It seems that the positively charged Lys146-Arg149 fragment can act as a nuclear localization sequence. Notably, only 10 of the studied 6,000 proteins, including PEDF, possess a C-terminal glycosylation site; all of these proteins are localized in the nucleus [16]. In its role as a nuclear protein, PEDF may possibly be involved in cell cycle regulation.

Charges on the surface of the PEDF molecule are positioned in an asymmetric fashion – the clusters of basic and acidic amino acids are on opposite sides of the globule. This allows PEDF to simultaneously interact with glycosaminoglycans and collagen. Interaction with glycosaminoglycans increases PEDF affinity towards collagen and vice versa [25]. Thus, PEDF can modulate cell adhesion processes that involve integrin-collagen interactions. Since cell adhesion is one of the key stages of angiogenesis, this may indeed be the basis for the antiangiogenic activity of the factor [25].

Phosphorylation of PEDF influences its antiangiogenic and neurotrophic activity

PEDF purified from human plasma is a phosphoprotein. There are three phosphorylation sites on the surface of the PEDF molecule - Ser24, Ser114, and Ser227. PEDF possesses and other potential phosphorylation sites; however, in normally folded molecules these sites are unavailable for kinases. In the seum factor is phosphorylated mainly by casein kinase 2 on 2 main residues, Ser24 and Ser114, but also by protein kinase A on Ser227 [26]. PEDF is functionally modulated by extracellular phosphorylation. The casein kinase 2 phosphorylated PEDF has a reduced neurotrophic activity, while its antiangiogenic activity is significantly increased. On the other hand, protein kinase A phosphorylation reduces the PEDF antiangiogenic activity but has only a slight effect on its neurotrophic activity. A fully phosphorylated factor displays both high antiangiogenic and neurotrophic activities.

Casein kinase 2 phosphorylation of Ser24 and Ser114 prevents further phosphorylation of PEDF by protein kinase A, and partial denaturation of protein restores its sensitivity to protein kinase A phosphorylation. Casein kinase 2 phosphorylation is followed by a conformational change in the PEDF molecule, thereby making Ser227 inaccessible to protein kinase A phosphorylation. On the other hand, protein kinase A phosphorylation of PEDF does not affect the conformational state of the casein kinase 2 phosphorylation sites and allows their further phosphorylation. The inhibitory effect of the casein kinase 2 phosphorylated PEDF on its protein kinase A phosphorylation may serve as a regulatory mechanism of PEDF function under conditions where the highly antiangiogenic activity of PEDF should be preserved while its neurotrophic activity should be eliminated. On the other hand, the protein kinase A phosphorylation of PEDF reduces PEDF antiangiogenic function. PEDF could be further phosphorylated by casein kinase 2, and that converts it from a poor to a very potent antiangiogenic factor that maintains its neurotrophic activity [27]. This is a novel role of extracellular phosphorylation that completely changes the nature of the physiologic activity of a circulating protein. Thus, differential phosphorylation induces variable effects of PEDF and therefore contributes to the complexity of PEDF action.

Occurrence and evolutionary conservation of PEDF

The *pedf* gene is conserved and can be found in the genomes of various animal species, ranging from fish to humans (Fig. 4, 5) [16]. Through DNA-RNA hybridization methods, researchers discovered that PEDF mRNAs were present in nearly all 44 of the examined tissues of adult humans and human embryos. It is pre-

sumed that PEDF is expressed mainly by cells that have not lost their ability to divide *in vivo* [28]. In various regions of the eye and layers of the retina in human adults and embryos, antibodies to the PEDF polypeptide interacted with the cytoplasm of developing photoreceptors, the glial layer, single cells in the neuroblast layer, and pigment granules in pigment epithelium cells (starting from the 8th week of embryo development). In the organism of an adult, the antibodies interacted with the nuclei of rods (but not with those of cones), with the cytoplasm of certain cells in the nuclear and glial layers, with pigment epithelium cells, choriocapillaris of the cornea, the pupil, and ciliated epithelium [29].

The most conserved species-specific peptide fragments are the following sequences [16]. The leading N-terminal region of the factor responsible for the secretion of PEDF from the cell was conserved during the course of evolution [30]. Another highly conserved sequence, Asn285-X-Thr287, is an N-glycosylation site. Four other peptide fragments are also conserved: two of which are unique to PEDF and form a cluster of negatively charged amino acids, and two others (in positions Val78-Gly95 and Phe384-Pro415) which are similar to other serpins. It seems that the functions of the latter two are similar for all serpins.

EFFECTS OF PEDF ON CELLS AND TISSUES OF VARIOUS ORIGINS

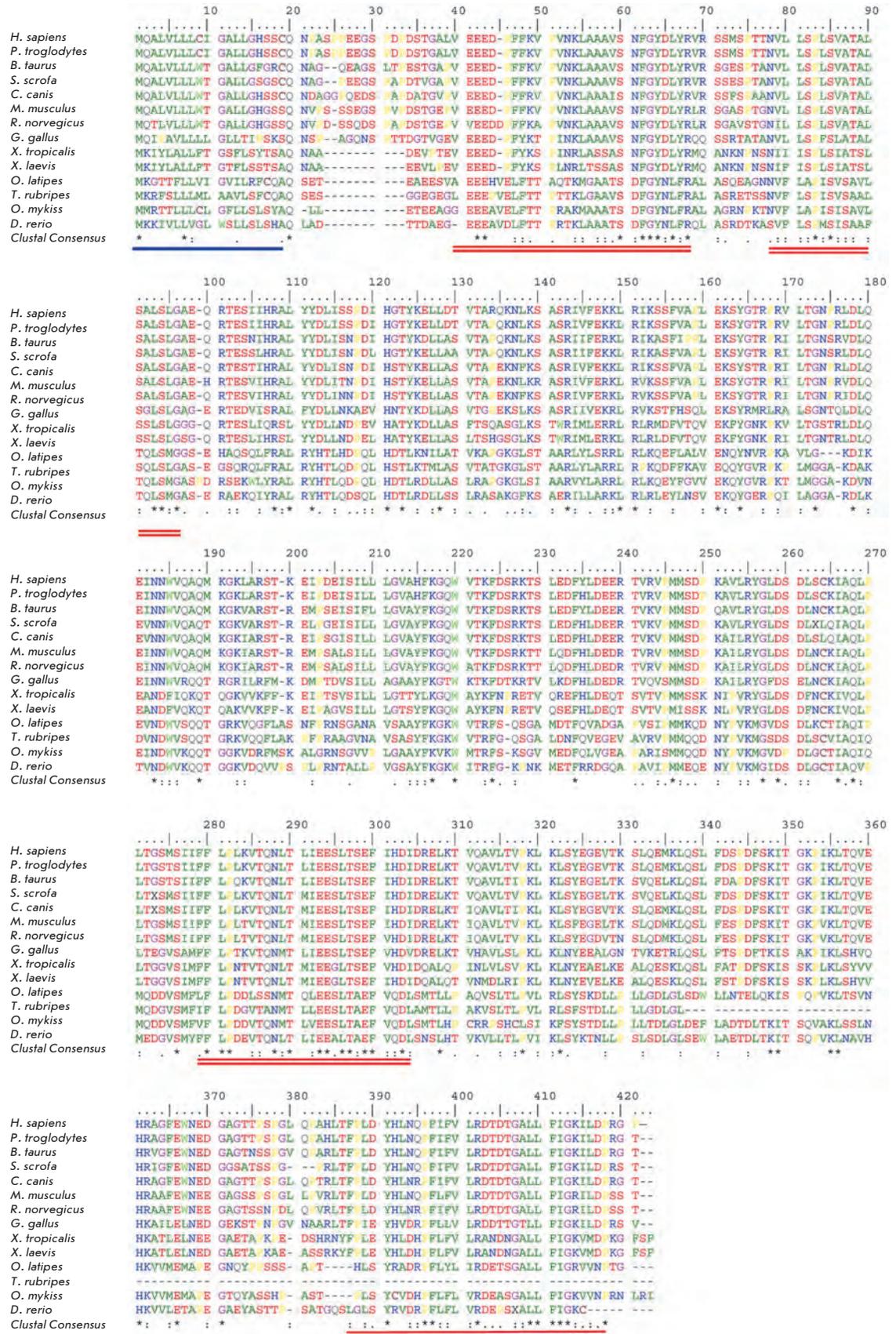
Differentiation effect of PEDF and its effect on metabolism

The differentiation effect of the neurotrophic PEDF factor was shown and studied on neuronal cells – both on embryonic (immature neurons, developing photoreceptors) [31,32] and tumor cells (neuroblastoma, retinoblastoma, and carcinoma) [3,33,34]. Human and bovine PEDF has a differentiation effect on *in vitro* cell cultures of immature motor neurons obtained from chicken embryo (5.5 days old) spinal cords [35]. PEDF is the main factor that controls normal morphogenesis and metabolism of photoreceptor and glial cells during the embryonic development of the retina [32, 36, 37].

PEDF has differentiation and antiproliferative effects on Y-79 retinoblastoma cells [1, 3]. The antitumor effect can also be observed on neuroblastoma cells. For this type of tumour, clinical prognosis is usually better when the population of cells is rich in differentiated neurons and Schwann cells, whose cultural medium has antitumor properties [38]. PEDF, which is one of the secreted products of Schwann cells, mediates the antitumor effects observed for these cultural media [33].

PEDF also has an antitumor effect on ovarian epithelial carcinoma, which is the most lethal type of can-

Fig. 4. Comparison of the amino acid sequences of PEDF proteins of different origins. 14 sequences of the factor from different biological species (human, *Homo sapiens*; chimpanzee, *Pan troglodytes*; domestic cow, *Bos taurus*; pig, *Sus scrofa*; dog, *Canis canis*; mouse, *Mus musculus*; rat, *Rattus norvegicus*; chicken, *Gallus gallus*; Western clawed frog, *Xenopus tropicalis*; African clawed frog, *Xenopus laevis*; Japanese ricefish, *Oryzias latipes*; rainbow trout, *Oncorhynchus mykiss*; zebra fish, *Danio rerio*) were analyzed using ClustalW software. Asterisks mark the positions of fully conserved residues. Colons indicate that a "strong" functional group is conserved; periods indicate that a "weak" group is conserved. The blue line marks the PEDF leader sequence. The double red line marks regions that are highly conserved in all the PEDF sequences. Dashes are used for the maximal alignment of the sequences [16].



	Hs	Chimp	Bt	Pig	Dog	Mm	Rt	Chick	Xt	Xl	Fugu	Trout	Zebra
Hs	100												
Chimp	99	100											
Bt	87	88	100										
Pig	88	88	90	100									
Dog	90	90	87	88	100								
Mm	86	85	84	84	87	100							
Rt	83	83	82	82	87	93	100						
Chick	63	63	63	63	62	63	63	100					
Xt	55	55	55	56	56	56	56	55	100				
Xl	56	57	57	58	57	57	56	53	89	100			
Fugu	34	35	36	36	36	36	35	35	34	35	100		
Trout	37	37	39	38	36	37	37	39	38	38	64	100	
Zebra	40	40	41	41	41	40	39	42	41	40	56	65	100

Fig. 5. Interspecies identity of the amino acid sequences of PEDF in percentage points. The table was generated using PSI-Blast software [16].

cers. The level of PEDF expression in ovarian tumors and their cell lines is significantly lower in comparison with normal ovarian epithelium. Exogenous PEDF inhibits the growth of both normal and tumour ovarian epithelium, while a decreased activity of endogenous PEDF accomplished through the injection of inhibitors results in increased proliferation of these cells.

The effect of PEDF on ovarian epithelial cells is mainly regulated by estrogens. Treatment of cell lines with 17β-estradiol lowers the production of PEDF and its mRNA by repressing the transcription of the *pedf* gene. Estrogen receptors play a role in this regulation. 17β-estradiol promotes the growth of normal and cancerous tumour ovarian epithelium cell lines, while simultaneous treatment with 17β-estradiol and PEDF negates the stimulating effect [34]. Some lines of ovarian carcinoma bear a deletion in the 17 chromosome at locus 17p13.3. This deletion is 15 kbp long and includes most of the *pedf* gene [39]. Proliferation of these cell lines cannot be stimulated by 17β-estradiol, which seemingly indicates that its effector region in the 5'-flanking region of *pedf* is also deleted. Epigenetic modifications, such as hypermethylation of the promoter, may act as alternative mechanisms for regulating *pedf* expression [34].

In pigment epithelium cells, PEDF stimulates the synthesis and accumulation of melanin [40]. The importance of melanin in the development of retinal neurons is suggested by the fact that, irrespective of the causative factors for albinism, all mammal albinos display

impaired visual signals [41]. The factor boosts the activity of the promoter of the tyrosine kinase gene, whose product is involved in the synthesis of melanin [42].

Neuroprotective effect of PEDF

PEDF was shown to prevent apoptosis of cells in a culture of retinal neurons after H₂O₂ treatment [43]. Pre-treatment with PEDF protects the pigment epithelium from the disruptions in the barrier function caused by H₂O₂. It also protects the cells from other effects of hydrogen peroxide: redistribution of the synaptic proteins occludin and N-cadherin in cell membranes of the pigment epithelium; reorganizations of actin and activation of the signaling cascade mediated by the p38/27 kDa heat chock protein [44].

The factor also has a protective effect on cultures of primary cerebellum granular cells extracted from baby rats (8 days old). It considerably slows spontaneous and serum deprivation-mediated apoptosis in this culture [45, 46]. PEDF also protects cerebellum granular cells [47], hippocampus neurons [48], and spinal motor neurons [49] from the toxic effects of glutamate. According to certain data, the main mechanism responsible for the protective effect of PEDF on granular cells is the activation of the NF-κB transcription factor [50].

Antiangiogenic effect of PEDF

PEDF is one of the most potent antiangiogenic factors. It inhibited vessel growth in rat corneas that were stim-

ulated with bFGF, which is a very potent angiogenic factor. The effective concentration of PEDF is much lower than that of the most potent, previously known antiangiogenic factors – angiostatin, endostatin, and thrombospondin [51]. The factor is involved in the complex and balanced control of angiogenesis, acting in opposition to angiogenic factors such as VEGF ((vascular endothelial growth factor) and bFGF (basic fibroblast growth factor) [52–55].

Saturation of tissues with oxygen affects the concentration of PEDF. Hyperoxygenation causes the formation of retinas with decreased amounts of blood vessels, while (though) the level of PEDF expression in these conditions was shown to be high [56]. Hyperoxygenation stimulates factor production; hypoxia, on the other hand, has a repressive effect.

The mechanism responsible for the antiangiogenic effect of PEDF is currently under study. Endothelium cells display positive chemotaxis towards angiogenic factors. Thus, one of the mechanisms of PEDF action is the inhibition of the chemotaxis of endothelial cells, which form the walls of blood vessels. This effect inhibits chemotaxis towards all of the studied angiogenic factors – VEGF, PDGF, IL-8, etc. [51].

According to other data, the reason for the antiangiogenic effect of PEDF is the apoptosis of endothelial cells [57]. A characteristic trait of the factor is its selective effect on growing vessels; it does not damage any functioning vessels [58]. The reason for such a selective effect is the mode of apoptosis induction, mediated by the Fas-ligand [59]. Angiogenic agents stimulate endothelial cells towards producing a range of anti-apoptotic molecules, which increase their viability [60], and they also stimulate Fas-receptors. Vessels that are already functioning have their endothelium cells in a state of “arrested growth,” with the cells in tight contact with each other. Fas-receptor expression is also decreased [61]. These factors protect cells from the apoptotic effect of PEDF.

Nonetheless, besides PEDF having a stimulating effect on Fas/FasL expression and Fas-mediated apoptosis, there must be alternative pathways for the apoptotic activity of the factor, since Fas and FasL knockout mice are still sensitive to PEDF-mediated apoptosis [57, 62, 63]. It has been shown that PEDF can activate p38 MAPK in a culture of epithelial cells extracted from a human umbilical vein (HUVEC) by promoting their phosphorylation. Activation of p38 results in the activation of caspases 3, 8, 9 and, ultimately, in cell apoptosis [64].

It is notable that VEGF also promotes p38 phosphorylation, and this effect is potentiated during co-treatment with PEDF. One of the most important functions of VEGF is to protect endothelial cells from apoptosis

in various conditions, including medium depletion. Activation of the p38 kinase by VEGF is accomplished by regulation of its transcription [65, 66] and cytoskeleton reorganization, as well as cell migration [67]. It seems likely that the threshold concentration of phosphorylated p38 MAPK needed for the proangiogenic effect to manifest is lower than the concentration needed for the activation of the apoptosis pathway. The PEDF-mediated increase in the level of activated p38 can result in concentrations of p38 that exceed the threshold concentration and, thus, cause apoptosis [64].

Even though PEDF stimulates the apoptosis of endothelial cells, it prevents the apoptosis of various cells of neural origin. PEDF can have the opposite effects on endothelial cells of differing phenotypes [68]. In these cases, it is possible that the factor’s action is effected via different unrelated mechanisms.

PEDF RECEPTOR

The multiple functions of PEDF are obviously realized through different mechanisms. However, it seems likely that in most cases the interaction between PEDF and cells takes place via a ligand-receptor mechanism. The effects of the factor can be blocked by antibodies which prevent the factor from binding to the cell’s surface [47,69,70].

Not long ago, researchers managed to identify a gene whose product acted as a PEDF receptor. The gene was thus named PEDF-R. This discovery was made using a two-hybrid system in retinal cells [72]. The gene is localized in region 11p15.5 in chromosome 11 and consists of 10 exons and 9 introns. The mRNA transcript of *pedf-r* is 2122 bases long. It encodes a protein molecule which consists of 504 amino acids (molecular mass of 55.315 kDa). The sequence contains 4 consensus N-glycosylation sites. When expressed in eukaryotic systems, the receptor has a mass of approximately 81 kDa, which is similar to the mass of PEDF-binding proteins on the cell’s surface. The latter were identified previously in cell lines of human retinoblastoma (Y-79), in rat cerebellum granular neurons, and in cells extracted from bovine retinas [70,71].

The receptor is mainly produced in retinal pigment epithelium cells, the inner segments of photoreceptors, and neuronal cells of the retina. However, the receptor is not detected in the outer segments of rods. Apart from normal pigment epithelium cells, PEDF-R is also present in a number of other tissues and organs – in the optical nerve, uvea, retinoblastoma cells, other tumors of various origins, as well as animal brain cells that were in a state of serotonin and catecholamine depletion. The gene encoding this receptor has been evolutionarily conserved in mammals. The highest degree of similarity was observed between human PEDF-R

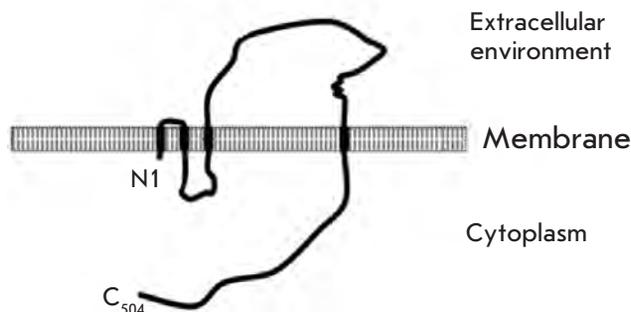


Fig. 6. Topology of PEDF-R predicted according to its amino acid sequence. The thick lines indicate the transmembrane domains [72].

and that of mice (89% identical) and rats (89% identical). Genes homologous to PEDF-R have also been found in the genomes of nematodes (*Caenorhabditis elegans*) and mosquitoes (*Anopheles gambiae*).

PEDF-R is a transmembrane protein with four transmembrane, two extracellular and three intracellular domains; the N- and C-termini of the molecule are exposed into the intracellular environment (Fig. 6) [72].

PEDF-R can bind PEDF with a $K_d \sim 3.03 \pm 0.716$ nM both in solution, as well as after immobilization of one of the proteins. It retains this ability even after denaturation and refolding. Posttranslational modifications, including N-glycosylation, are not required for the factor to bind to the receptor – PEDF-R expressed in bacteria interacts with PEDF with a K_d nearly identical to that observed under normal conditions. The factor-binding domain of PEDF-R has been identified as Gln250-Arg383 and is called 12c. This fragment was shown to be highly similar to type I rat collagen. It seems likely, however, that this is not the only fragment that plays a role in PEDF binding, since the factor's affinity towards a full-sized receptor is higher. The dissociation constant of the latter is nearly two orders of magnitude larger than that of the 12c fragment ($K_d.12c \sim 134$ nM).

The factor-receptor interaction involves the Val78-Thr121 fragment of the PEDF molecule. A synthetic analog (44-a. a. residues long) retains the ability to bind to the receptor [70].

Apart from the ability to bind PEDF, PEDF-R also manifests phospholipase activity and has been classified as a member of the calcium-independent, A_2 phospholipase family. Another feature of the PEDF-R sequence is two spatially adjacent patatin-like fragments (Asn39-Ala54 and Ser158-Tyr177) that form a Ser47/Asp166 catalytic dyad, which is present in the active sites of patatin B2 and human cytosolic A_2 phospholipases [73].

A_2 Phospholipases release lysophospholipids, which play a role in cell signaling and affect the development

and functioning of all mammalian organ systems [74]. A_2 phospholipases also release biologically active fatty acids from the cell's membrane. These acids act as messengers or as precursors of eicosanoids, which mediate signal transduction [75]. Arachidonic and docosahexaenoic acids, which are abundant in the membranes of cells in the retina and the central nervous system, can influence the viability of retinal and central nervous system cells [76, 77] and demonstrate antitumor and antiangiogenic effects, similar to PEDF [78,79].

PEDF-R can hydrolyze arachidoinoyl-*sn*-glycerin-3-phosphocholine, thus releasing arachidonic acid [80]. Even though there is currently no data on the release of docosahexaenoic acid by PEDF-R, it has been shown that PEDF in subnanomolar concentrations can activate the synthesis and release of its derivatives from the ARPE-19 human cell line [81, 82]. These data indicate that PEDF can facilitate the survival of retinal cells by influencing the lipid signaling pathways that are connected with PEDF-R. Interaction of PEDF with its receptor, PEDF-R, results in the release of fatty acids and lysophospholipids from cell membranes. These entities act both para- and autocrinally, by raising the viability and differentiation potential of neuronal cells, or by causing the death of tumor and endothelial cells [72].

PEDF AS A SUBSTRATE OF MMP-2 AND MMP-9

As previously mentioned, regulation of angiogenesis is very dependent on the quantitative ratio between the differentiation factors VEGF and PEDF. It has been shown that increasing the concentrations of exogenous PEDF in a model of choroidal neovascularization stimulates the production of VEGF by endothelial cells. In turn, increased concentrations of VEGF induce the production of matrix metalloproteinases MMP-2 and -9 (or the corresponding gelatinases A and B). The activated proteinases hydrolyze the components of the extracellular matrix and the PEDF previously bound to these components. However, they do not influence VEGF. Thus, increased concentrations of PEDF result in elevated inactivation of this factor by MMP, which increases the VEGF/PEDF ratio and stimulates angiogenesis.

The region of PEDF most accessible for proteolysis by MMP-2 and -9 is the exposed loop that contains the Leu382-Thr383 bond. However, apart from this bond, the PEDF sequence also has several sites for MMP-2 and -9 proteolysis, which are partially or completely "masked" by a tertiary structure. These sites are less accessible to the activity of matrix metalloproteinases. It is possible that the activity of MMP-2 and/or -9 results in a gradual loss of its tertiary structure by the PEDF molecule, making the inner bonds more accessible to metalloproteinases. Complete proteolysis creates

a multitude of peptide fragments, which manifest no antiangiogenic or neurotrophic activity even though they retain the ability to bind components of the extracellular matrix. Thus, matrix metalloproteinases can inactivate PEDF through the same mechanism as the one effected by inhibitory serpins (such as antitrypsin and antithrombin III). They affect the reactive site of the loop and are thought to form a stable inactive complex with the factor molecule (although the existence of such a complex has yet to be proven). However, there is still a possibility that the loss of biological activities by PEDF is achieved by the complete gradual proteolysis of the protein [83].

CONCLUSION

The PEDF protein has been the subject of intensive studies for several years, and the amount of PEDF-

related research increases with every year. It has been shown that the protein regulates a wide range of processes in the cells and tissues of various eukaryotic organisms, including in those of humans. PEDF is a key player in high-level intracellular interactions, though it is possible that some of its functions in the metabolic pathways of the cell have yet to be identified. The antiproliferative, antiangiogenic, protective, differentiative and other activities of PEDF are very promising not only from the viewpoint of basic scientific research, but also for the use of this protein in the treatment of various diseases, such as neurodegenerative damage to the optical and nervous systems and cancerous processes. ●

This work was supported by the program of the RAS Presidium Molecular and Cell Biology

REFERENCES

- Tombran-Tink J., Johnson L.V. // Invest. Ophthalmol. Vis. Sci. 1989. V. 30. P. 1700–1707.
- Tombran-Tink J., Chader G.J., Johnson L.V. // Exp. Eye Res. 1991. V. 53. P. 411–414.
- Steele F., Chader G.J., Johnson L.V., Tombran-Tink J. // Proc. Natl. Acad. Sci. USA. 1993. V. 90. P. 1526–1530.
- Pignolo R.J., Cristofalo V.J., Rotenberg M.O. // J. Biol. Chem. 1993. V. 268. P. 8949–8957.
- Kozaki K., Miyaishi O., Koiwai O., et al. // J. Biol. Chem. 1998. V. 273. P. 15125–15130.
- Tombran-Tink J., Pawar H., Swaroop A., et al. // Genomics. 1994. V. 19. P. 266–272.
- Goliath R., Tombran-Tink J., Rodriguez I.R., et al. // Mol. Vis. 1996. V. 2. P. 5.
- Tombran-Tink J., Mazuruk K., Rodriguez I., et al. // Mol. Vis. 1996. V. 2. P. 11.
- von Heijne G. // Nucleic Acids Res. 1986. V. 14. P. 4683–4690.
- Stratikos E., Alberdi E., Gettins P.G., Becerra S.P. // Protein Sci. 1996. V. 5. P. 2575–2582.
- Ortego J., Escribano J., Becerra S.P., et al. // Invest. Ophthalmol. Vis. Sci. 1996. V. 37. P. 2759–2767.
- Petersen S.V., Valnickova Z., Enghild J.J. // Biochem. J. 2003. V. 374. P. 199–206.
- Tombran-Tink J., Shivaram S.M., Chader G.J., et al. // J. Neurosci. 1995. V. 15. P. 4992–5003.
- Carrell R.W., Pemberton P.A., Boswell D.R. // Cold Spring Harbor Symp. Quant. Biol. 1987. V. 52. P. 527–535.
- Irving J.A., Pike R.N., Lesk A.M., Whisstock J.C. // Genome Res. 2000. V. 10. P. 1845–1864.
- Tombran-Tink J., Aparicio S., Xu X., et al. // J. Struct. Biol. 2005. V. 151. P. 130–150.
- Becerra S.P., Palmer I., Kumar A., et al. // J. Biol. Chem. 1993. V. 268. P. 23148–23156.
- Huber R., Carrell R.W. // Biochemistry. 1989. V. 28. P. 8951–8966.
- Zhou A., Carrell R.W., Huntington J.A. // J. Biol. Chem. 2001. V. 276. P. 27541–27547.
- Becerra S.P., Sagasti A., Spinella P., Notario V. // J. Biol. Chem. 1995. V. 270. P. 25992–25999.
- Shao H., Schvartz I., Shaltiel S. // Eur. J. Biochem. 2003. V. 270. P. 822–831.
- Simonovic M., Gettins P.G.W., Volz K. // Proc. Natl. Acad. Sci. USA. 2001. V. 98. P. 11131–11135.
- Becerra S.P. // Exp. Eye Res. 2006. V. 82. P. 739–740.
- Alberdi E., Hyde C.C., Becerra S.P. // Biochemistry. 1998. V. 37. P. 10643–10652.
- Meyer C., Notari L., Becerra S.P. // J. Biol. Chem. 2002. V. 277. P. 45400–45407.
- Maik-Rachline G., Shaltiel S., Seger R. // Blood. 2005. V. 105. P. 670–678.
- Gettins P.G., Simonovic M., Volz K. // Biol. Chem. 2002. V. 383. P. 1677–1682.
- Pignolo R.J., Rotenberg M.O., Cristofalo V.J. // J. Cell. Physiol. 1995. V. 162. P. 110–118.
- Karakousis P., John C., Behling K., et al. // Mol. Vis. 2001. V. 7. P. 154–163.
- Tombran-Tink J., Shivaram S.M., Chader G.J., et al. // J. Neurosci. 1995. V. 15. P. 4992–5003.
- Isobe M., Emanuel B.S., Givol D., et al. // Nature. 1986. V. 320. P. 84–86.
- Jablonski M.M., Tombran-Tink J., Mrazek D.A., Iannaccone A. // J. Neurosci. 2000. V. 20. P. 7149–7157.
- Crawford S.E., Stellmach V., Huang X., et al. // Cell Sci. 2001. V. 114. P. 4421–4428.
- Cheung L.W.T., Au S.C.L., Cheung A.N.Y., et al. // Endocrinology. 2006. V. 147. P. 4179–4191.
- Houenou L.J., D'Costa A.P., Linxi L., et al. // J. Comp. Neurol. 1999. V. 412. P. 506–514.
- Jablonski M.M., Tombran-Tink J., Mrazek D.A., Iannaccone A. // Glia. 2001. V. 35. P. 14–25.
- Sugita Y., Becerra S.P., Chader G.J., Schwartz J.P. // J. Neurosci. Res. 1997. V. 49. P. 710–718.
- Brodeur G.M. // New Engl. J. Med. 1996. V. 334. P. 1537–1539.
- Phillips N.J., Ziegler M., Radford D.M., et al. // Cancer Res. 1996. V. 56. P. 606–611.
- Malchiodi-Albedi F., Feher J., Caiazza S., et al. // Int. J. Dev. Neurosci. 1998. V. 16. P. 423–432.
- Guillery R.W. // Trends Neurosci. 1986. V. 9. P. 364–367.
- Abul-Hassan K., Walmsley R., Tombran-Tink J., Boulton

REVIEWS

- M. // *Pigment Cell Res.* 2000. V. 13. P. 436–441.
43. Cao W., Tombran-Tink J., Chen W., et al. // *J. Neurosci. Res.* 1999. V. 57. P. 789–800.
44. Ho T.C., Yang Y.C., Cheng H.C., et al. // *Biochem. and Biophys. Res. Commun.* 2006. V. 342. P. 372–378.
45. Taniwaki T., Becerra S.P., Chader G.J., Schwartz J.P. // *J. Neurochem.* 1995. V. 64. P. 2509–2517.
46. Araki T., Taniwaki T., Becerra S.P., et al. // *J. Neurosci. Res.* 1998. V. 53. P. 7–15.
47. Taniwaki T., Hirashima N., Becerra S.P., et al. // *J. Neurochem.* 1997. V. 68. P. 26–32.
48. DeCoster M.A., Schabelman E., Tombran-Tink J., Bazan N.G. // *J. Neurosci. Res.* 1999. V. 56. P. 604–610.
49. Bilak M.M., Corse A.M., Bilak S.R., et al. // *J. Neuropathol. Exp. Neurol.* 1999. V. 58. P. 719–728.
50. Yabe T., Wilson D., Schwartz J.P. // *J. Biol. Chem.* 2001. V. 276. P. 43313–43319.
51. Dawson D.W., Volpert O.V., Gillis P., et al. // *Science.* 1999. V. 285. P. 245–248.
52. Ferrara N., Davis-Smyth T. // *Endocr. Rev.* 1997. V. 18. P. 4–25.
53. Klagsbrun M., Vlodavsky I. // *Prog. Clin. and Biol. Res.* 1988. V. 266. P. 55–61.
54. O'Reilly M.S., Holmgren L., Shing Y., et al. // *Cold Spring Harbor Symp. Quant. Biol.* 1994. V. 59. P. 471–482.
55. DiPietro L.A., Nebgen D.R., Polverini P.J. // *J. Vasc. Res.* 1994. V. 31. P. 178–185.
56. Smith L.E., Wesolowski E., McLellan A., et al. // *Invest. Ophthalmol. Vis. Sci.* 1994. V. 35. P. 101–111.
57. Stellmach V., Crawford S., Zhou W., Bouck N. // *Proc. Natl. Acad. Sci. USA.* 2001. V. 98. P. 2593–2597.
58. Bouck N. // *Trends Mol. Med.* 2002. V. 8. P. 330–334.
59. Volpert O.V., Zaichuk T., Zhou W., et al. // *Nat. Med.* 2002. V. 8. P. 349–357.
60. Gerber H.P., Dixit V., Ferrara N. // *J. Biol. Chem.* 1998. V. 273. P. 13313–13316.
61. Aoudjit F., Vuori K. // *J. Cell. Biol.* 2001. V. 152. P. 633–644.
62. Barreiro R., Schadlu R., Herndon J., et al. // *Invest. Ophthalmol. Vis. Sci.* 2003. V. 44. P. 1282–1286.
63. Mori K., Duh E., Gehlbach P., et al. // *J. Cell Physiol.* 2001. V. 188. P. 253–263.
64. Chen L., Zhang S., Barnstable C.J., Tombran-Tink J. // *Biochem. and Biophys. Res. Commun.* 2006. V. 348. P. 1288–1295.
65. Mayo L.D., Kessler K.M., Pincheira R., et al. // *J. Biol. Chem.* 2001. V. 276. P. 25184–25189.
66. Read M.A., Whitley M.Z., Gupta S., et al. // *J. Biol. Chem.* 1997. V. 272. P. 2753–2761.
67. Rousseau S., Houle F., Landry J., Huot J. // *Oncogene.* 1997. V. 15. P. 2169–2177.
68. Hutchings H., Maitre-Boube M., Tombran-Tink J., Plouët J. // *Biochem. and Biophys. Res. Commun.* 2002. V. 294. P. 764–769.
69. Wu Y.Q., Notario V., Chader G.J., Becerra S.P. // *Protein Expr. Purif.* 1995. V. 6. P. 447–456.
70. Alberdi E., Aymerich M.S., Becerra S.P. // *J. Biol. Chem.* 1999. V. 274. P. 31605–31612.
71. Aymerich M.S., Alberdi E.M., Martinez A., Becerra S.P. // *Invest. Ophthalmol. Vis. Sci.* 2001. V. 42. P. 3287–3293.
72. Notari L., Baladron V., Aroca-Aguilar J.D., et al. // *J. Biol. Chem.* 2006. V. 281. P. 38022–38037.
73. Gettins P.G., Simonovic M., Volz K. // *Biol. Chem.* 2002. V. 383. P. 1677–1682.
74. Goetzl E.J., Tigyi G. J. // *Cell. Biochem.* 2004. V. 92. P. 867–868.
75. Balsinde J., Winstead M.V., Dennis E.A. // *FEBS Lett.* 2002. V. 531. P. 2–6.
76. SanGiovanni J.P., Chew E.Y. // *Prog. Retin. Eye Res.* 2005. V. 24. P. 87–138.
77. Kim H.Y., Akbar M., Kim K.Y. // *J. Mol. Neurosci.* 2001. V. 16. P. 279–284.
78. Rose D.P., Connolly J.M. // *Pharmacol. Ther.* 1999. V. 83. P. 217–244.
79. Rose D.P., Connolly J.M. // *Int. J. Oncol.* 1999. V. 15. P. 1011–1015.
80. Jenkins C.M., Mancuso D.J., Yan W., et al. // *J. Biol. Chem.* 2004. V. 279. P. 48968–48975.
81. Bazan N.G., Marcheselli V.L., Hu J., et al. // *Invest. Ophthalmol. Vis. Sci.* 2005. V. 46. P. 167.
82. Marcheselli V.L., Bazan N.G., Hu J., et al. Program 148.9. Washington, D.C.: Society for Neuroscience, 2005.
83. Notari L., Miller A., Martinez A., et al. // *Invest. Ophthalmol. Vis. Sci.* 2005. V. 46. P. 2736–2747.

Association of ERAP1 Allelic Variants with Risk of Ankylosing Spondylitis

I. V. Zvyagin^{*1#}, V. Yu. Dorodnykh^{1#}, I. Z. Mamedov¹, D. B. Staroverov¹, A. G. Bochkova², D. V. Rebrikov^{3,4}, Y. B. Lebedev¹

²Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences

²Institute of Rheumatology, Russian Academy of Medical Sciences

³Vavilov Institute of General Genetics, Russian Academy of Sciences

⁴DNA-Technology JSC

Authors contributed equally to this work

*E-mail: izvyagin@gmail.com

Received 03.09.2010

ABSTRACT Ankylosing spondylitis (AS) belongs to a group of autoimmune diseases affecting the axial skeleton. Beside the *hla-b*27* allele, several other human genes that control the variety processes of immune homeostasis are considered to be associated with AS manifestation in different human populations. Among strong associated non-MHC genes *erap1* encoding the endoplasmic reticulum aminopeptidase 1 isoform was recently identified by single nucleotide polymorphisms (SNPs) meta analysis. In our study we inspected the genetic association of five non-synonymous coding SNPs from *erap1* with AS in Caucasians. We implemented the SSP-PCR system for precise genotyping of 87 *hla-b*27* positive AS patients and 77 *hla-b*27* healthy donors from the Russian population. Considerable differences in allele's frequencies within patients vs control cohort were shown for 3 of 5 SNPs under investigation. Using the EM-algorithm we reconstructed 3-marker haplotypes that distinguish with high probability two cohorts due to differences in the haplotypes frequencies. In such a way both the sensitive, CCT, haplotype and the protective, TTC, one were predicted. To verify the calculation we determined genuine frequencies of 5-marker haplotypes in AS cohort by haplotyping of individual cDNA samples using improved SSP-PCR primer set. We demonstrated that the frequencies of *in silico* reconstructed haplotypes and the frequencies of experimentally detected haplotypes are in a good agreement. Frequency of the risk haplotype CCT (rs17482078/10050860/2287987) detected within AS cohort reaches 88%, as well as the frequency calculated by EM-algorithm.

KEYWORDS aminopeptidase ERAP1, human single nucleotide polymorphism, allele-specific PCR, human haplotypes, ankylosing spondylitis.

ABBREVIATIONS AS - ankylosing spondylitis, ER - endoplasmic reticulum, MHC - major histocompatibility complex, TNF - tumor necrosis factor, TNFR - tumor necrosis factor receptor, IL - interleukin.

INTRODUCTION

Ankylosing spondylitis (AS) is one of autoimmune diseases belonging to spondyloarthropathy. The disease is characterized by an inflammation of axial skeleton joints and a resulting formation of ossified regions in the joints. The first symptoms of the disease appear at the age of 20-25, and then it slowly progresses. Recently several genetic loci have been found to be associated with the disease. The products of these genes are involved in various stages of the immune response. According to results of twins studies the concordance in monozygote twins varies from 26 to 60%, whereas for dizygote twins the concordance varies from 4 to 20%. This evidence underlies the importance of the genetic background for AS development [1-3].

The strongest association was observed for the *hla B*27* allelic variant of the *MHC-I* gene. Approximately 90% of patients of European descent suffering from AS bear this allele; however, only 5% of Europeans carrying the allele are affected by the disease [4]. Large-

scale full-genome comparisons of diagnosed patients cohorts with healthy people show that several other SNPs in non-MHC-I genomic loci are also associated with the disease. These loci have a weaker association with the risk of developing AS. Among the non-MHC loci, the strongest association was observed for the *erap1* aminopeptidase gene [5, 6].

ERAP1 aminopeptidase acts in a variety of immune response stages. The difference of ERAP1 functioning probably exists due to its different intracellular localizations. ERAP1 is mainly located in the endoplasmic reticulum (ER) in human and murine cells. However, the protein also might be expressed on the outer surface of the cytoplasmic membrane or be secreted into the extracellular environment [7]. One of the functions of ERAP1 is the N-terminal proteolysis of peptides, which are formed during the proteasome-dependent degradation of cellular proteins. The formed peptides are then exposed on the cell's surface by MHC-I molecules. Also ERAP1 might be involved in the formation of the solu-

ble form of the type I tumor necrosis factor receptor (TNFR-1), along with the soluble forms of interleukin 1 and 6 (IL) receptors. ERAP1 thus plays an important role in the regulation of the immune response [8-10].

Studies of cohorts of different ethnicities and varying genetic homogeneities in terms of the *hla-b* allele content revealed in identification of several non-synonymous SNPs in the coding region of the *erap1* gene which are associated with an increased risk of AS [11-15]. The sets of AS-associated polymorphisms detected in these studies are different. It is probably due to the genetic properties of the cohorts under investigation.

The presence of amino acid substitutions and/or their combinations in ERAP1 variants may result in changes in the aminopeptidase activity. These substitutions may also result in the deregulation of peptide processing and control over immune cells via TNF or IL-1 and -6. The putative ability of ERAP1 variants bearing substitutions to form inappropriate peptides on MHC-I is considered as the aminopeptidase impact on AS progression. The hypothesis assumes a misrecognition of the presented peptides as alien signals, and it leads, in turn, to an increase of empty MHC-I or its dimmers consisting exclusively of heavy chains on cell's surface. Finally these processes could result in the formation of nonspecific immune responses against cells that expose the misfolded MHC-I [16]. Dysfunction of ERAP1 activity aimed at IL receptors could be another explanation of the mutant aminopeptidase involvement in AS progression.

In our research, we have determined the frequencies of *erap1* allele variants and corresponding genotypes for 5 non-synonymous SNPs in cohorts of AS patients and healthy individuals from the Russian population. Both the AS patients cohort and healthy group consist of *hla-B*27* positive individuals. Markers rs2287987 (Met349Val), rs30187 (Lys528Arg), rs10050860 (Asp575Asn), rs17482078 (Arg725Gln), and rs27044 (Glu730Gln) were selected for the genotyping. Each of the markers have been previously associated with AS by other authors [11-15]. The location of these non-synonymous substitutions within the coding region of the *erap1* gene allows us to imply the existence of functional effects of the corresponding amino acid substitutions.

We used the genotyping results for haplotypes reconstruction and for association analysis of the *erap1* allele variants with AS.

MATERIALS AND METHODS

Genomic DNA and RNA samples

The samples of genomic DNA were purified from the peripheral blood of 77 *hla-B*27* positive healthy donors aged 21-63, and of 84 *hla-B*27* positive patients with

the axial form of AS. Genomic DNAs were extracted from the peripheral blood mononuclear cells using a Diatom Prep 100 kit (Izogen Laboratory, Moscow) according to the manufacturer's protocol.

Samples of total RNA were extracted from peripheral blood mononuclear cells using TRIzol® Reagent (Invitrogen Lifetechnologies, USA), according to the manufacturer's protocol.

cDNA synthesis

1-2 micrograms of total RNA were used for cDNA synthesis. The first strands of cDNAs were synthesized using the "Mint-Universal cDNA synthesis kit" (Eurogen, Russia) according to the manufacturer's protocol.

Geno- and haplotyping

Genotyping was performed using a set of allele-specific PCR primers (Table 1). Each pair consisted of an allele-specific primer complementary to the allele variant and a primer specific to the nearest intron. The presence of the distinct allele in a sample was determined by the presence/absence of the specific PCR products. Allele-specific PCR on cDNA templates using subsets of four allele-specific primers corresponding to each pairs of SNPs was performed for haplotyping.

To enhance the allele discrimination ability a mismatch was introduced in the third position of the primer 3'-end of each of the allele-specific primers. Geno- and haplotyping results have been independently reproduced at least twice for each sample. Additionally concordance of the geno- and haplotyping results was inspected.

"HS Taq-DNA-polymerase kit" (Eurogen, Russia) was used for PCR amplification; the final volume of the reaction mixture was 15 microliters; the annealing tempera-

Table 1. Sequence of the primers used in this study

Primer	Sequence of nucleotides (5'-3')
30-A For	ATGAACACTTGGACACTGCACAA
30-G For	ATGAACACTTGGACACTGCACAG
27-C Rev	CACACAGGCGAGGAGTAGTAGATC
27-G Rev	CACACAGGCGAGGAGTAGTAGATG
100-G For	CATTCATCACCAGCAAATGCG
100-A For	CATTCATCACCAGCAAATGCA
174-T Rev	TAGTAGTTTACTCCGCAGCAATT
174-C Rev	TAGTAGTTTACTCCGCAGCAATC
228-A For	ATCAAGTAAGCTTGGCATCAGAA
228-G For	ATCAAGTAAGCTTGGCATCAGAG
228 int	AACAAATTAACCTCAAATGTGAAG
30 int	CCTCCTTAATCCTACTGGGAAGAT
100 int	GGCCATACATATGATATAACCCAGTA
174&27 int	CTGGGACTCTTCATGGTACTTGGAG

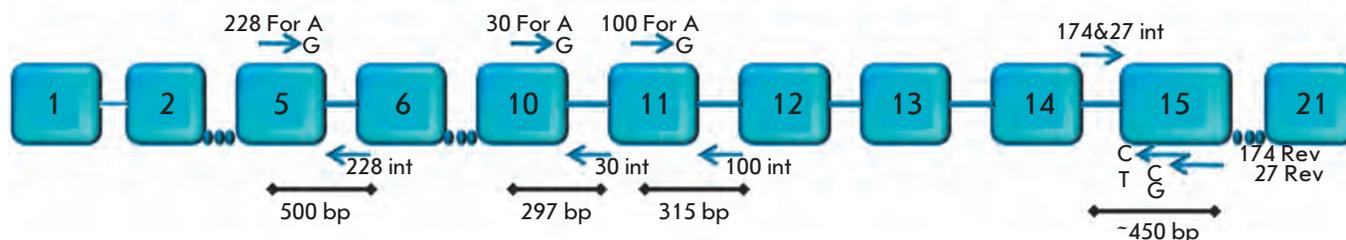


Fig. 1. Scheme of sequence-specific primer positions for *erap1* allele detection. Introns and exons of *erap1* and size of PCR-products are shown.

ture was set at 65°C; 30 and 33 cycles of PCR were implemented to amplify DNAs and cDNAs, respectively.

Statistical analysis

Haploview 4.1 (<http://www.broadinstitute.org/haploview/haploview>) and Genetic Data Analysis (Lewis, P. O., and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. <http://lewis.eeb.uconn.edu/lewishome/software.html>) software was used to analyze the results of geno- and haplotyping.

Hardy-Weinberg equilibriums for *erap1* alleles were tested using Fischer exact test, as well as the exact test described by G. Abecasis and J. Wigginton [17].

The statistical significance of the differences in the allele frequencies were analyzed using the χ^2 test with Yate's correction. The value of the risk factor (OR) and the confidence interval (CI 95%) were calculated using an online calculator: www.openepi.com.

Expectation-maximization algorithm (EM-algorithm) of Haploview 4.1 software was used to detect linkage disequilibrium (LD) of tested SNPs and to calculate the frequencies of reconstructed haplotypes. Strong LD was defined as $D' \geq 0.8$ CI95% 0.7–0.99 [18].

Bayesian algorithm of PHASE 2.1 software was implemented to confirm the frequencies of reconstructed haplotypes.

RESULTS AND DISCUSSION

The frequencies of minor allelic variants of the *erap1* gene are decreased in AS patients

By using a set of allele-specific primers (Table 1, Fig. 1), we identified the frequencies of 10 allelic variants of the *erap1* gene and the corresponding genotypes in *hla-B27* positive cohorts of 84 AS patients and 77 healthy donors. The allelic variants of the *erap1* gene were distinguished by the set of non-synonymous SNPs: rs2287987 (C/T, Met349Val), rs30187 (C/T, Lys528Arg), rs10050860 (C/T, Asp575Asn), rs17482078 (C/T, Arg725Gln), and rs27044 (C/G, Glu730Gln). These pol-

ymorphisms code the amino acid substitutions in ERAP1. The protein variants bearing the substitutions might be functionally different and therefore play a positive or protective role in AS progression. Association of each of 5 markers with ankylosing spondylitis was published recently by several authors [11–15]. Two of the five SNPs are “located” near the putative active sites of the aminopeptidase; rs2287987 – is located near the Zn²⁺ ion-binding site [11], while rs30187 is located near the putative substrate-binding pocket.

Distribution of 5 allele pairs in both the AS patients and the control cohorts showed no significant deviations from the Hardy-Weinberg Equilibrium.

Analysis of the genotyping results showed that the frequencies of three minor alleles of rs2287987 ($p < 0.002$, OR = 0.35), rs10050860 ($p < 0.004$, OR = 0.39), and rs17482078 ($p < 0.05$, OR = 0.52) markers are significantly lower among AS patients compared to healthy donors in the Russian population (Table 2). An associative analysis test shows that three allelic variants

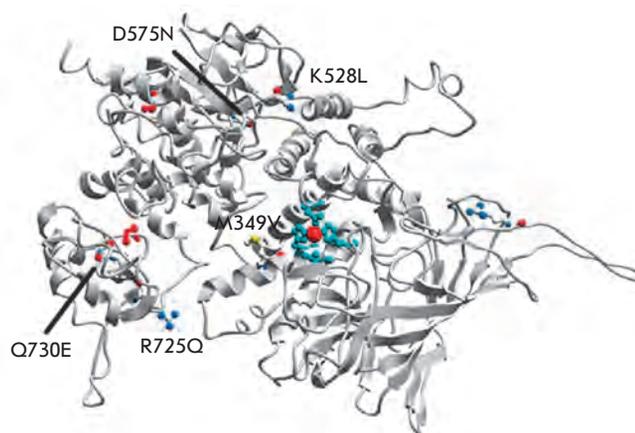


Figure 2. Ribbon model of putative human ERAP1 structure (modified from [11]). Amino acid substitutions encoded by SNPs studied in this work are highlighted. Corresponding amino acid residues are marked by blue balls. The putative active site of ERAP1 is denoted by cyan circle with a red sphere representing a Zn atom.

Table 2. Frequencies of allelic variants and genotypes for the studied genetic markers in two cohorts

Marker, polymorphic nucleotide	AS patients, n = 84			Healthy donors, n = 77			p	OR (95%CI)
	MAF* [nucleotide in minor allele]	Number of genotypes (frequency)		MAF* [nucleotide in minor allele]	Number of genotypes (frequency)			
rs2287987 [C/T]	0.09 [C]	CC	1 (0.01)	0.22 [C]	CC	3 (0.04)	0.002	0.35 (0.18–0.68)
		CT	13 (0.15)		CT	26 (0.36)		
		TT	70 (0.84)		TT	44 (0.60)		
rs30187 [C/T]	0.37 [T]	CC	29 (0.35)	0.33 [T]	CC	35 (0.45)	0.55	1.18 (0.75–1.87)
		CT	48 (0.57)		CT	33 (0.43)		
		TT	7 (0.08)		TT	9 (0.12)		
rs10050860 [C/T]	0.11 [T]	CC	66 (0.78)	0.23 [T]	CC	46 (0.60)	0.004	0.39 (0.21–0.73)
		CT	18 (0.21)		CT	26 (0.34)		
		TT	-		TT	5 (0.06)		
rs17482078 [C/T]	0.13 [T]	CC	65 (0.77)	0.21 [T]	CC	47 (0.61)	0.046	0.52 (0.29–0.95)
		CT	17 (0.20)		CT	27 (0.35)		
		TT	2 (0.02)		TT	3 (0.04)		
rs27044 [C/G]	0.33 [G]	GG	7 (0.08)	0.29 [G]	GG	9 (0.12)	0.47	1.23 (0.76–1.97)
		GC	42 (0.5)		GC	26 (0.34)		
		CC	35 (0.42)		CC	42 (0.55)		

*MAF – minor allele frequency

of *erap1* are associated with an increased risk of AS (rs2287987 [T] (OR = 2.86), rs10050860 [C] (OR = 2.54), and rs17482078 [C] (OR = 1.91). However, our results also show a lack of a significant association of rs30187 and rs27044 polymorphisms with AS in the Russian population.

Decreased frequencies of minor alleles of rs2287987, rs10050860, and rs17482078 markers are accompanied by a decreased number of heterozygous individuals and an increased number of homozygous for the major allele individuals in the AS patient group as compared to the control group. Heterozygosity for the rs30187 and rs27044 markers, which are not associated with AS in the Russian population, increases in AS patients, although the minor allele frequencies do not differ significantly AS patients from healthy individuals.

In general our data are in a good agreement with those reported by other researchers. Several studies have shown that all five of the inspected markers are associated with the risk of AS [5, 6, 11], although the cohorts of AS patients and healthy individuals included a considerable number of *hla-B27*-negative donors. Two other studies [14, 15] were performed on cohorts consisting of *hla-B27*-positive donors exclusively or mostly (77–85%). No association with AS was found for the rs30187 and rs17482078 markers, while rs2287987, rs10050860, and rs27044 proved to be associated with the AS risk [14]. In another study [15] it was demonstrated that al-

lelic variants for the markers rs10050860 and rs30187, but not rs27044, were associated with the AS risk.

Statistically significant association of rs27044 and rs30187 but not of rs2287987, rs10050860, and rs17482078 markers with AS was reported for the Korean population [12].

The differences between the results of the associative analyses are most likely due to considerable differences between the studied populations in *erap1* allele frequencies. Another reason for the incongruence between the cited results might be the different presentation of *hla-B27* positive individuals in the cohorts under comparison. In our study, special attention was focused on the genetic homogeneity of the cohorts in terms of the *hla-b27* carriership and on the uniformity of clinical manifestations of the disease. Such considerations helped us avoid risk of false positive associations.

Haplotype reconstruction based on obtained genotype data

The genotyping results for both cohorts were used to calculate haplotype frequencies. We searched for haplotypes whose frequencies increased statistically significantly in AS patients and could, thus, be associated with AS risk. We used Haploview 4.1 software to identify the linkage disequilibrium between the pairs of the studied markers. Polymorphisms rs2287987, rs10050860, and rs17482078 proved to be in statistically

Table 3. Calculated frequencies of reconstructed haplotypes including markers rs17482078/10050860/2287987, associated with risk of AS in Russian population.

Haplotype	Frequency in AS patients, <i>n</i> = 84	Frequency in healthy donors, <i>n</i> = 77	P	OR (95%CI)
CCT	0.86	0.75	0.026	1.96 (1.12–3.46)
TTC	0.08	0.2	0.003	0.33 (0.17–0.67)

significant pairwise linkage disequilibrium ($D' \geq 0.85$, 95%CI 0.76–0.99, according to [18]) and were grouped into a single block. The theoretical haplotype frequencies in both AS patients and healthy cohorts were calculated for markers belonging to this block. Haplotypes with theoretical frequencies below 5% were discounted from further analysis.

The CCT haplotype was found to be statistically significantly associated with AS risk ($p < 0.03$, OR = 1.96). The calculated frequency of the TTC haplotype was to be lower in AS patients as compared to the healthy individuals ($p < 0.003$, OR = 0.33) (Table 3).

Haplotyping of AS patients in Russian population

Using paired combinations of allele-specific primers and cDNA samples we identified *erap1* gene haplotypes in the same cohort of AS patients. Table 4 lists the frequencies of the identified 5-marker haplotypes in the patients cohort. Haplotypes consisting of only 3 markers associated with AS in the Russian population were listed specifically. The table does not list haplotypes with frequencies of less than 5%, with the exception of TTC haplotype, which was determined as a putative protective haplotype based on theoretical calculations.

The determined frequency of the putative risk haplotype CCT among AS patients was approximately 88%. The percentage is in agreement with the calculated frequency value. The frequency of the putative protective TTC haplotype is approximately 1%, while the calculated value is approximately 8% (Table 3). This difference could be attributed to the insufficient size of the AS patients cohort for a statistically representative identification of the actual frequency of the protective haplotype.

The identified risk haplotype includes mostly those SNP variants that were demonstrated to be risk-associated in other cohorts of Europeans (Table 5).

Notably recent researchs has yielded controversial data regarding the risk-associated haplotypes for the markers rs27044 and rs30187 in different populations. Maksymowych *et al.* [14] reported a risk association for the C-allele (marker rs27044), while a study by Pazar *et al.* [15] showed that both identified risk haplotypes include G-allele of the same marker. Moreover, the two

risk haplotypes identified in the cited study [15] are invariant for the rs30187 marker. These results, as well as the controversial data on the association of *erap1* gene allele variants for markers rs27044 and rs30187 with AS risk, may indicate that the amino acid substitutions encoded by these polymorphisms do not affect the activity of the aminopeptidase, nor its role in AS. The described associations may be due to the heterogeneity of the patient cohorts or/and genetic linkage of these polymorphisms with other non-synonymous substitutions which result in changes of aminopeptidase function that is essential for AS.

The published results of the few *in vitro* studies indicate that a Lys528Arg substitution, encoded by the SNP rs30187, has a strong effect on ERAP1 aminopeptidase activity. Goto *et al.* [19] demonstrated that the ERAP1 variant, which has Arg in position 528, displays low peptidase activity to synthetic substrates as well to the natural substrates (angiotensin II and kallidin). In the cited study [19] it was also shown that substitution Gln730Glu, encoded by the polymorphism rs27044, and substitution Asp575Asn, encoded by the polymorphism rs10050860, have almost no effect on peptidase activity to synthetic and natural substrates. Transient ex-

Table 4. Distribution of identified haplotypes in a AS patient cohort of the Russian population

Haplotype*	Number of haplotypes in AS patient (frequency), <i>n</i> = 69
rs27044/17482078/10050860/30187/2287987	
CCCCT	70 (0.51)
GCCCT	36 (0.26)
GCCCT	8 (0.06)
CCCTT	7 (0.05)
GTTGT	7 (0.05)
CTTCC	2 (0.01)
rs17482078/10050860/2287987	
CCT	121 (0.88)
TTT	7 (0.05)
TTC	2 (0.01)

*Nucleotides which are a part of the predicted protective and risk –associated haplotypes are underlined and bold.

Table 5. Allelic variants of *erap1*, associated with increased risk of AS development in various populations

rs27044	rs17482078	rs10050860	rs30187	rs2287987	reference
C	-	C	T	-	[14]
G	C	C	T	T	[15]
G	C	C	C	T	[15]
G	C	C	T	-	[12]
G	-	-	T	-	[13]
-	C	C	-	T	this study

pression of *erap1* in a melanoma cell line demonstrated that the amino acid substitution Met349Val (polymorphism rs2287987) increased ERAP1 peptidase activity to a synthetic substrate [20]. Comparison of amino acid substitutions characterized in the previously cited articles with our associative analysis data, and other similar findings by other authors (Table 5), suggests that some of the AS-associated non-synonymous substitutions can alter ERAP1 peptidase activity. However, most of the disease-associated alleles are likely to have other functional manifestations, since the corresponding amino acid substitutions do not affect the enzyme's activity on synthetic substrates.

Seemingly disagreement between the results of functional tests and the associative analysis can be overcome by identification of haplotypes specific for healthy cohorts of patients. In our study, we have identified risk-associated CCT and protective TTC haplotypes for the markers rs17482078/10050860/2287987. The frequency of both haplotypes is statistically significantly increased in the corresponding cohort (Table 4). One of the substitutions encoded by the protective haplotype (Met349Val) increases ERAP1 activity on sev-

eral substrates [20]. This fact suggests that increased ERAP1 activity might prevent AS. The protective effect of the TTC haplotype may be due to the activity of ERAP1 towards receptors of proinflammatory cytokines or due to the correct presentation of peptides on MHC-I molecules. In contrast, the non-synonymous substitutions included in the risk haplotype CCT could lead to increased levels of membrane-bound forms of TNFR-1 and receptors for IL-6 and -1.

Considering that AS is a multigene disease, further research for the ERAP1 role in AS should be focused on both identification of the function features of different *erap1* allele variants and study of the putative interactions between the enzyme and the products of other genes associated with the disease. These interactions may indeed be affected by the non-synonymous substitutions in the risk haplotype of the *erap1* gene which we have identified in the Russian population. ●

This study was supported by the Molecular and Cellular Biology program of Presidium of RAS and by grants from the Russian Foundation for Basic Research (№ 10-04-01771-a and 11-04-01585-a).

REFERENCES

- van der Linden S.M., Valkenburg H.A., de Jongh B.M., Cats A. // *Arthritis Rheum.* 1984. V. 27. № 3. P. 241–249.
- Pedersen O.B., Svedensen A.J., Ejstrup L., et al. // *Scand. J. Rheumatol.* 2008. V. 37. № 2. P. 120–126.
- Brown M.A., Laval S.H., Brophy S., Calin A. // *Ann. Rheum. Dis.* 2000. V. 59. № 11. P. 883–886.
- Reveille J.D., Maganti R.A. // *Adv. Exp. Med. Biol.* 2009. V. 649. P. 159–176.
- The Australo-Anglo-American Spondyloarthritis Consortium (TASC) // *Nature Genetics.* 2010. V. 42. № 2. P. 123–127.
- Wellcome Trust Case Control Consortium & The Australo-Anglo-American Spondylitis Consortium (TASC) // *Nature Genetics.* 2007. V. 39. № 11. P. 1329–1337.
- Toshihiro T., Hattori A., Shinako M., et al. // *J. Biol. Chem.* 2003. V. 278. № 34. P. 32275–32283.
- Cui X., Hawari F., Alsaaty S., et al. // *J. Clin. Invest.* 2002. V. 110. № 4. P. 515–526.
- Cui X., Rouhani N.F., Hawari F., et al. // *J. Immunology.* 2003. V. 171. № 12. P. 6814–6819.
- Cui X., Rouhani N.F., Hawari F., Levine J.S. // *J. Biol. Chem.* 2003. V. 278. P. 28677–28685.
- Harvey D., Pointon J.J., Evans M.D., et al. // *Hum. Mol. Genet.* 2009. V. 18. № 21. P. 4204–4212.
- Choi C.-B., Kim T.-H., Jun J.-B., et al. // *Ann. Rheum. Dis.* 2010. V. 69. № 3. P. 582–584.
- Tsui F.W.L., Haroon N., Reveille D.J., et al. // *Ann. Rheum. Dis.* 2010. V. 69. № 4. P. 733–736.
- Maksymowych W.P., Inman R.D., Gladman D.D., et al. // *Arthr. & Rheum.* 2009. V. 60. № 5. P. 1317–1323.
- Pazar B., Safrani E., Gergely P., et al. // *J. Rheum.* 2010. V. 37. № 2. P. 379–384.
- Haroon N., Inman D.R. // *Nat. Rev. Rheum.* 2010. *Advanced online publication.*
- Wigginton J., Cutler D., Abecasis G.A. // *Am. J. Hum. Genet.* 2005. V. 76. № 5. P. 887–883.
- Gabriel S.B., Schaffner S.F., Nguyen H., et al. // *Science.* 2002. V. 296. № 5576. P. 2225–2229.
- Goto Y., Hattori A., Ishii Y., Tsujimoto M. // *FEBS Lett.* 2006. V. 580. P. 1833–1838.
- Kamphausen E., Kellert C., Abbas T., et al. // *Cancer Immun. Immunother.* 2010. V. 59. № 8. P. 1273–1284.

Adaptive Changes in *Mycobacterium avium* Gene Expression Profile Following Infection of Genetically Susceptible and Resistant Mice

D. V. Ignatov¹, T. A. Skvortsov¹, K. B. Majorov², A. S. Apt², T. L. Azhikina^{1*}

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

²Central Research Institute for Tuberculosis, Russian Academy of Medical Sciences

*E-mail: tatazhik@ibch.ru

Received 03.08.2010.

ABSTRACT We performed a comparative analysis of *Mycobacterium avium* transcriptomes (strain 724R) in infected mice of two different strains- resistant and susceptible to infection. Sets of mycobacterial genes transcribed in lung tissue were defined, and differentially transcribed genes were revealed. Our results indicate that *M. avium* genes coding for enzymes of the Krebs cycle, oxidative phosphorylation, NO reduction, fatty acid biosynthesis, replication, translation, and genome modification are expressed at high levels in the lungs of genetically susceptible mice. The expression of genes responsible for cell wall properties, anaerobic nitrate respiration, fatty acid degradation, synthesis of polycyclic fatty acid derivatives, and biosynthesis of mycobactin and other polyketides is increased in the resistant mice. In the resistant host environment, *Mycobacterium avium* apparently transitions to a latent state caused by the deficiency in divalent cations and characterised by anaerobic respiration, degradation of fatty acids, and modification of cell wall properties.

KEYWORDS *Mycobacterium avium*, transcriptome analysis *in vivo*, coincidence cloning, RNA-seq.

INTRODUCTION

Infectious diseases caused by intracellular pathogenic bacteria represent a significant challenge in health care. The course of the infection depends not only on the protective mechanisms (native and acquired immune response, and mucous barriers), but also on the specific expression of bacterial genes. Altered expression as a response to the immune reaction of the host organism is critical for the survival and functioning of pathogenic bacteria. An analysis of these changes is important for understanding how infectious diseases proceed and developing effective approaches towards their treatment.

Mycobacterium avium are widespread mycobacteria that become intracellular pathogens in humans in the absence of normal T-cell-mediated immunity [1, 2]. These bacteria are found in approximately 70% of incurable AIDS patients and are believed to be the main cause of death in such patients [3]. In patients with weakened immunity (older people and children), *M. avium* may cause chronic lung diseases [4–6]. Experiments modelling the infection in mice of the C57BL/6 (B6) and derivative strains with knockout mutations in genes essential for immunity showed that T-cell-mediated immune response to *M. avium* had both defensive,

as well as pathogenetic functions. In such an infection, the balance between the immune response and pathogenic processes in lung tissue is very similar to that of tuberculosis [7–9]; therefore, we can assume that the diseases caused by these mycobacteria are similar not only in their immune system mechanisms, but also in the mechanisms employed by the pathogens to overcome this defence.

It has been shown that mice of the I/St (I/StSnEgY-Cit) and B6 strains differ in their ability to resist an *M. avium*-induced infection [10]. Respiratory infection in B6 mice leads to a prolonged infiltration of lung tissue by macrophages and neutrophils, leading to the formation of necrotic lung granulomas and death. In contrast, in the I/St mice the infection is controllable, produces moderate infiltration of lung tissue, leading to small and medium granulomas without a necrotic centre, and the animals survive. The susceptibility of B6 mice to the *M. avium*-induced infection was shown to result from the presence of the nonfunctional allele of the *Nramp1* (natural resistance-associated macrophage protein-1) gene in their genome. The protein coded for by that gene consists of 12 transmembrane domains and is expressed at the membranes of late lysosomes and phagosomes. *Nramp1* functions by removing divalent cations

Table 1. Oligonucleotides and primers used for coincidence cloning.

	Name	5'-3' structure
Suppressive adapter 1A (resulted from anneal of equimolar mixture of 1A long and 1A short)	1A long	GTAATACGACTCACTATAGGGCAGCGTGGTCGCGCCGAGAG
	1A short	CTCTCGGCCG
Suppressive adapter 1B	1B long	GTAATACGACTCACTATAGGGCAGGGCGTGGTTCGCGAGGGCGGC
	1B short	GCCGCCCTCC
Suppressive adapter 2A	2A long	GTAATACGACTCACTATAGGGCAGGCAGGCGGTGGTGGGCAGGC
	2A short	GCCTGCCAC
Suppressive adapter 2B	2B long	GTAATACGACTCACTATAGGGCAGCGGAGGCGGTAGGAGGCGGA
	2B short	TCCGCCTCCT
External primer	T7	GTAATACGACTCACTATAGGGC
Internal primers	pr 1A	AGCGTGGTCGCGCCGAGAG
	pr 1B	AGGGCGTGGTTCGCGAGGGCGGC
	pr 2A	AGGCAGGCGGTGGTGGGCAGGC
	pr 2B	AGCGGAGGCGGTAGGAGGCGGA

(Fe²⁺, Mn²⁺, etc.) from phagosomes, thus depriving the mycobacteria of important metabolites [10].

The B6 immune response is characterized by an increased production of IFN- γ , TNF- α , and, especially, IL-12. We suppose that the differences in the immune response to *M. avium* infection are manifested in the differences in the pathogen expression in the lungs and lymphoid organs of mice of the susceptible strain versus those of the resistant strain, showing that the mechanisms essential for resistant host survival may not activate during infection of the susceptible host.

This work endeavoured to study the biochemical processes involved in the adaptation of *M. avium* to genetically different host organisms. We compared sequences transcribed in mice of the I/St and B6 strains in the 13th week of infection.

EXPERIMENTAL

Standard DNA and RNA procedures were carried out according to ref. [11]. Genomic DNA of the *M. avium* 724R strain was isolated according to the procedure described in ref. [12].

Infection

Mice of inbred strains I/StSnEgYCit (I/St) and C57BL/6YCit (B6) were bred and maintained under conventional, non-specific-pathogen-free conditions at the Animal Facilities of the Central Institute for Tuberculosis (Moscow, Russia) in accordance with guidelines from the Russian Ministry of Health (guideline 755) and the NIH Office of Laboratory Animal Welfare (assurance A5502-06). Female mice, 2.5–3.0 months old, were infected by the respiratory route with 1–2 $\times 10^3$ viable CFU of *M. avium* 724R strain, characterized in ref. [13], using an inhalation exposure system (Glas-Col, USA) according to the procedure described in ref. [10].

RNA isolation and cDNA synthesis

RNA was isolated from the lungs of mice of both strains in the 13th week after infection, using the RNA Isolation System kit (Promega, USA). RNA samples were treated with DNase I (MBI Fermentas, Lithuania) to remove DNA traces. The first cDNA strand was constructed using oligonucleotide primers BR (5'-AAGCAGTGGTATCAACGCAGAGTAC(N)₉) and SMART

(5'-AAGCAGTGGTATCAACGCAGAGTACGrGrG). Both primers (at 12 μ M) were annealed with 2 μ g of total RNA in 11 μ l of solution. The mixture was incubated for 2 minutes at 70°C and then placed in ice for 10 minutes. cDNA was synthesised using reverse transcriptase PowerScript II (Clontech, USA). In parallel with reverse transcription (RT+), a reaction used as a control (RT-) without reverse transcriptase was performed. The RT+ and RT- mixtures were incubated at 37°C for 10 minutes, then for 40 minutes at 42°C. cDNA was synthesised in 30 PCR cycles (95°C for 20 sec, 64°C for 20 sec, and 72°C for 2 min) using 5S primers (5'-GTGGTATCAACGCAGAGT). Then, cDNA was purified using QIAquick PCR Purification kit (Qiagen, USA).

Coincidence cloning was carried out following the procedure described in ref. [14]. Genomic DNA of the *M. avium* 724R strain and total cDNA samples (synthesised using total RNA) were fragmented with restrictases RsaI and AluI. The obtained genomic DNA fragments were ligated with suppressive adaptors 1A for hybridisation with the I/St cDNA sample, and adaptors 1B for hybridisation with the B6 cDNA sample (Table 1). Suppressing adaptors 2A and 2B were ligated to cDNA fragments from the lung tissues of I/St and B6 mice, respectively. A mixture of 100 ng of the genomic DNA sample and 100 ng of one of the cDNA samples in 2 μ l of the hybridisation buffer (50 mM HEPES, pH 8.3; 0.5 M NaCl; 0.02 mM EDTA, pH 8.0) was incubated at 99°C for 5 minutes (denaturation) and then at 68°C for 18 hours (renaturation). After this procedure, 100 μ l of the hybridisation buffer at 68°C was added to the mixture, and 1 μ l of the resulting solution was used as a template for PCR. The first PCR stage was performed in a 25 μ l reaction volume containing 10 pmol of T7 primer. After incubation for 5 minutes at 72°C (filling-in sticky ends), 20 amplification cycles were carried out (94°C for 30 sec, 66°C for 30 sec, and 72°C for 90 sec). The second stage of amplification was performed with 10 pmole of internal primers pr1A/pr1B and pr2A/pr2B, and it consisted of 25 cycles (94°C for 30 sec, 68°C for 30 sec, and 72°C for 90 sec), using the PCR product of the first stage, diluted ten-fold. The amplification product was purified using the QIAquick PCR Purification kit (Qiagen) and then used for 454 sequencing.

454 SEQUENCING

Nucleotide sequences of cDNA libraries were determined by massive parallel pyrosequencing using the genetic analyser GS FLX (Roche, Germany) and a 20 x 75 cm picotitration plate. The sequences of 83,000 independent reactions were determined. The sequences were mapped to the genome sequence of the *M. avium* strain 104, since the *M. avium* 724R genome has not yet been sequenced. The number of cDNA fragments

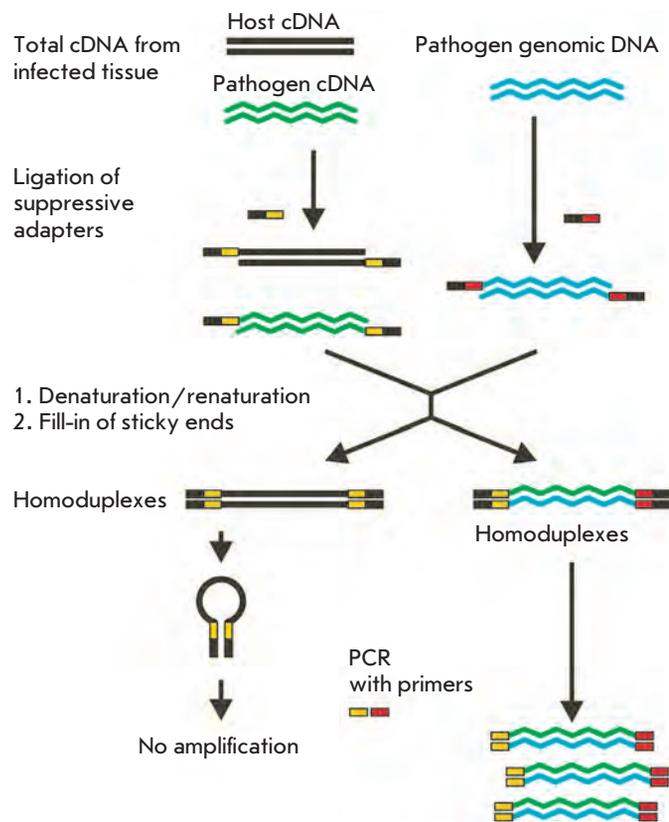


Fig. 1. Coincidence cloning. Suppressing oligonucleotide adaptors are ligated to fragments of bacterial genomic DNA and total cDNA. The samples are mixed, denatured, and slowly renatured, which leads to the formation of two types of duplexes. Due to selective suppression of PCR, only heteroduplexes containing fragments of bacterial genomic DNA and bacterial cDNA are amplified.

corresponding to each gene was determined using the BLASTn algorithm (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). A sequence was considered to belong to a certain gene if a fragment of that sequence had more than 95% homology with the gene segment longer than 40 nucleotides. The *M. avium* genes, the expression of which in samples I/St and B6 is significantly different, were determined following the procedure described in ref. [15].

RESULTS AND DISCUSSION

The course of the pathology and immune response to *M. avium* infection in mice from the susceptible (B6) and resistant (I/St) strains are discussed in detail in refs. [10] and [16]. The airborne infected susceptible B6 mice died after 7 months, while the resistant I/St mice survived for longer than 11 months. In the susceptible B6 mice, the lung pathology developed quickly,

Table 2. *M. avium* genes differentially transcribed in the lungs of infected I/St and B6 mice.

Gene	Coded protein
Increased expression in the lungs of the I/St mice	
MAV_2015	MbtG; mycobactin lysine-N-oxygenase
MAV_1696	Glutamate dehydrogenase
MAV_1304	NarH; nitrate reductase, β -subunit
MAV_2379	MetH; vitamin B12-dependent methionine synthase
MAV_2385	Mce protein
MAV_2063	Mce protein
MAV_2386	Mce protein
MAV_0118	PPE protein
MAV_3109	RifB; polyketide synthase 7
MAV_0880	3-Ketosteroid- δ -1-dehydrogenase
MAV_3000	Acyl-CoA dehydrogenase
MAV_4019	Assumed acyl-CoA dehydrogenase
MAV_4679	Cyclopropane fatty acid synthase
Increased expression in the lungs of the B6 mice	
MAV_2514	PPE protein
MAV_2924	PPE protein
MAV_2926	PPE protein
MAV_2244	GlnA; glutamine synthetase
MAV_4011	NO-reductase, β -subunit
MAV_1074	SucC; succinyl-CoA-synthase, β -subunit
MAV_3303	AcnA; aconitate hydratase
MAV_1130	NADH-dehydrogenase I, H-subunit
MAV_4040	NADH-dehydrogenase I, H-subunit
MAV_1524	ATP-synthase F_1F_0 , δ -subunit
MAV_5034	Transposase
MAV_1059	Transposase

accompanied by enhanced infiltration of lung tissue by immune system cells and increased production of pro-inflammatory cytokines IFN- γ , TNF- α , IL-6, and IL-12. Two parameters showing the susceptibility to infection correlated well: in the susceptible B6 mice, *M. avium* grew faster in lungs, and the lung pathology was deeper than in the resistant I/St mice.

We studied transcribed sequences *in vivo* using the coincidence cloning method we had developed previously [14, 17]. From the lungs of infected mice, we isolated total RNA, a mixture of the mice and bacterial RNA, with the amount of bacterial RNA being very small (less than 0.1–0.2% according to [18]). Using the total RNA from the I/St and B6 mice, total cDNA was synthesized. In the coincidence cloning method (Fig. 1), total cDNA and *M. avium* genomic DNA were denatured and renatured in one mixture. After a two-step selective PCR amplification, a set of fragments enriched with the bacterial cDNA fragments was obtained.

Qualitative (determination of nucleotide sequences of specifically expressed genes) and quantitative (the level of their expression) analyses of the sets were performed using parallel pyrosequencing.

The sequencing produced two libraries of *M. avium* cDNA sequences expressed in the lung tissue of infected I/St and B6 mice. We selected a series of genome loci, the expression of which was higher in sample I/St than in sample B6, and a series of loci, the expression of which was higher in sample B6 than in sample I/St (Table. 2). Locus annotation was performed using the KEGG (www.genome.jp/kegg) database. We anticipated that differential gene expression in the samples could be a manifestation of the microorganism's environmental adaptation; therefore, the products of the genes we found could be potential virulence factors.

We found the differential expression of the PPE gene family (*MAV_0118*, *MAV_2514*, *MAV_2924*, and *MAV_2926*). These proteins play an important role in the course of the mycobacterial infection because of both their antigen and immune functions. These acidic proteins, rich in glycine, are identified by the specific Pro-Pro-Glu (PPE family) and Pro-Glu (PE family) domains; they often contain polymorphic GC-rich sequences (PGRSs) and multiple copies of basic polymorphic tandem repeats. It is believed that these proteins are expressed on the cell's surface and are responsible for antigen variability, inducing different immune responses depending on the type of PE/PPE proteins expressed on the cell's surface [19]. Thus, the *MAV_0118* gene is expressed in the resistant mice, while the *MAV_2514*, *MAV_2926*, and *MAV_2924* genes are expressed in the susceptible mice. Since the mechanism of PPE protein action remains unknown, the above-mentioned observation is hard to explain; however, it

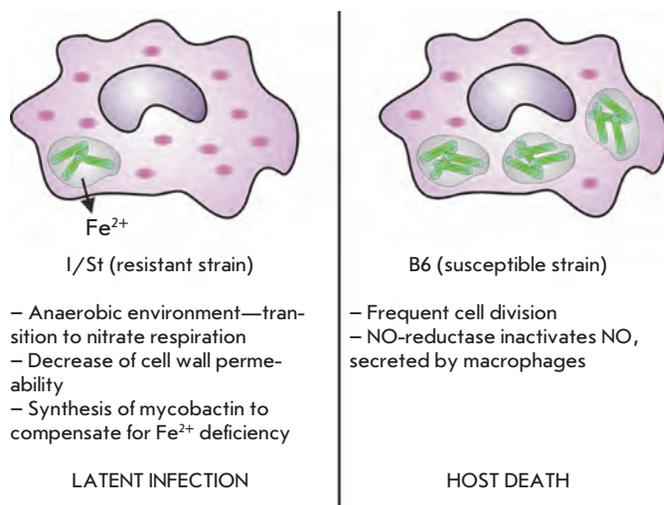


Fig. 2. *M. avium* metabolic state in the lungs of the I/St and B6 mice

is possible that the differential PPE expression is due to the differences in the immune responses.

The expression of the *MAV_2244* locus is increased in the B6 sample. This gene is an ortholog of the *glnA1* *M. tuberculosis* gene; it codes for glutamine synthetase, a key enzyme for nitrogen assimilation. It has been shown that this enzyme is important for *M. tuberculosis* persistence in macrophages. It is possible that in infected B6 mice *M. avium* enters into an environment auxotrophic for *L*-glutamine [20]. The *MAV_4011* locus that codes for the cytochrome b-containing subunit of NO-reductase is also worth mentioning. This enzyme reduces NO to N₂O, and it participates in the denitrification process in some soil microorganisms. However, no denitrification is observed in the *M. avium* from the susceptible B6 mice. It is speculated that *M. avium* use NO-reductase to get rid of the NO released by the macrophage into the endosomes and thus avoid the harmful effects of NO [21]. This could be the reason why *M. avium* are resistant to NO [22]. The expression of NO-reductase in *M. avium* from the lungs of the susceptible mice could be a result of the stronger immune response and increased NO production by macrophages.

In sample B6, we observed a more active and diverse expression of genes coding for the Krebs cycle enzymes: *MAV_1074* and *MAV_3303* coding for succinyl-CoA-synthetase and aconitate hydratase, respectively; and of genes coding for the proteins important for oxidative phosphorylation, as well as the respiratory electron-transport chain proteins: *MAV_1130*, *MAV_4040*, and *MAV_1524*. It is likely that in the susceptible mice, respiration is increased during persistent infection in or-

der to supply the pathogen. The *MAV_4040* locus codes for one of the NADH-dehydrogenase I subunits, which is typical for the *M. tuberculosis* virulent form, during exponential proliferation of the pathogen [23].

The expression of the *MAV_5034* and *MAV_1059* genes coding for transposases in sample B6 indicates an enhanced level of gene rearrangements. Also, in sample B6, we detected an increased expression of the *MAV_5024* and *MAV_5027* genes coding for type II restriction-modification enzymes that protect cells from foreign DNA.

The increased expression of the *MAV_0382* (subunits of DNA-polymerase III) and *MAV_4450* (ribosomal protein) genes in sample B6 indicates an increase in the DNA replication level due to a more frequent mitosis, as well as a higher translation level.

In sample I/St, there was a very high level of expression of the *MAV_2015* gene that codes for mycobactin lysine-N-oxygenase (MbtG). This enzyme is responsible for one of the last stages of mycobactin synthesis; it is an iron-chelating agent that supplies the microorganism with iron from the environment [24]. It has been shown for *M. tuberculosis* that the activation of the *mbt B-H* gene cluster involved in mycobactin synthesis occurs either when the environment is depleted in iron [25] or in an anaerobic environment [26]. The expression of this gene is high in *M. avium* from the resistant I/St mice, but it is very low in the B6 mice. As mentioned above, these two mice strains differ in the *Nramp1* gene allele that codes an ionic pump which is assumed to pump out divalent cations from the endosomal region, where *M. avium* is located [16]. There is the functional allele of this gene in the I/St mice, as opposed to a nonfunctional one in the B6 mice. Apparently, the *M. avium* endosomes from the resistant mice are iron-deficient, and the microorganism synthesizes vast amounts of mycobactin in order to compensate for the deficiency.

In *M. avium* from the I/St mice, an increased expression of the *MAV_1696* gene coding for NAD⁺-dependent glutamate dehydrogenase is observed. It is believed that, in contrast to NADP⁺-dependent glutamate dehydrogenase, which is responsible for nitrogen assimilation, in microorganisms the former enzyme takes part in the glutamate catabolism, and this gene expression is independent of the NH₄⁺ concentration. On the other hand, it has been shown recently that in *M. smegmatis* the expression of the *msmeg_4699* gene, an ortholog of *MAV_1696*, increases in response to NH₄⁺ deficiency [27]. In addition, there is no gene coding for NADP⁺-dependent glutamate dehydrogenase in the *M. avium* genome [28]. Some researchers speculate that, in mycobacteria, nitrogen assimilation involving NAD⁺-dependent glutamate dehydrogenase may be more energy-efficient than via the GS/GOGAT pathway; this

being important, for example, when the pathogen is in a latent state [27].

The *MAV_2379* gene coding for B₁₂-dependent methionine synthase MetH is expressed at a high level in *M. avium* from the I/St mice. This protein is involved in the final stage of methionine synthesis. In the *M. avium* genome, this reaction is controlled by MetE-B₁₂-independent methionine synthase, which is not expressed in the presence of vitamin B₁₂ [29]. The regulation of the *metH* gene expression has not been studied in detail, so the reason for the increase in its expression in the resistant mice is not quite clear.

The *MAV_1304* locus coding for the β-subunit of nitrate reductase is of particular interest. This gene is orthologous to the *narH* gene of *M. tuberculosis*. Its product is a subunit of anaerobic nitrate reductase NarGHJI, an enzyme enabling nitrate respiration in the absence of oxygen. Mutants lacking NarH cannot reduce nitrogen under anaerobic conditions [30]. When this gene was deleted in *M. bovis*, BCG bacteria demonstrated normal growth *in vitro* with sufficient oxygen supply; however, they appeared significantly less virulent when used for infecting mice [31]. The expression of the *MAV_1304* gene in *M. avium* from the lungs of the I/St mice might be an indication that, due to the harmful effects of the host's defence systems, the microorganism is subjected to anaerobic conditions and has to switch to nitrate respiration.

The *MAV_2063*, *MAV_2385*, and *MAV_2386* genes coding for proteins from the Mce family are expressed in the resistant mice. The function of the Mce proteins has not been clarified, although it is known that they enable invasiveness. These proteins supposedly represent a new group of ABC-transporters participating in the remodeling of the cell's membrane [32].

The *MAV_4679* locus coding for an enzyme involved in the synthesis of mycolic acids is expressed at a high level in *M. avium* from the I/St mice. An ortholog of this gene in *M. tuberculosis* is important for persistence in mice lungs. Mutants of this gene cannot cause lung infection in mice [33].

The *MAV_3109* locus codes for the RifB protein and is an ortholog of the *M. tuberculosis pks7* gene. An increased expression of the gene is observed in infected mice of the resistant strain. The protein product of this gene codes for an enzyme involved in the synthesis of phthiocerol dimycocerosate, one of the components of the mycobacterial cell wall, which ensures its impermeability [34].

The *MAV_0880* locus codes for 3-ketosteroid-δ-1-dehydrogenase, one of the enzymes involved in cholesterol catabolism. During the *M. tuberculosis*-induced in-

fection, cholesterol provides the pathogen with energy for persistence in macrophages [35]. In *M. avium* from the I/St mice, we observed an increased expression of the *MAV_3000* and *MAV_4019* genes coding for enzymes degrading fatty acids: acyl-CoA-dehydrogenase and acyl-CoA-synthase. During persistence in macrophages, the catabolism of fatty acids is the primary energy source for *M. tuberculosis* [36].

CONCLUSION

This paper contains the first description of the *M. avium* transcriptome during infection *in vivo*. Until now, only a single publication on the *M. avium paratuberculosis* transcriptional response to various factors *in vitro* [37] has been available.

We employed the model of genetic control of susceptibility to *M. avium* infection and disease severity in mice in order to detect the sequences that are transcribed differently in infected mice from the genetically resistant and genetically susceptible strains, i.e. when the pathogen persists in genetically different microenvironments. We obtained data on the qualitative and quantitative differences in the transcription profiles of genes of bacteria persisting in the resistant and susceptible mice, which indicate some changes in the metabolism of *M. avium* (Fig. 2).

In the course of the infection in the genetically susceptible organism (B6 strain), we found an increased expression of several genes responsible for nitrogen assimilation, NO reduction, the Krebs cycle, and oxidative phosphorylation, as well as replication and translation. The infection proceeds with active division of the mycobacteria and death of the host organism.

In the course of infection in the genetically resistant organism (I/St strain), we found an increased expression of several genes responsible for the modification of the cell surface's properties, switching to anaerobic nitrate respiration, degradation of fatty acids, synthesis of polycyclic derivatives of fatty acids, and biosynthesis of mycobactin and other polyketides. In general, the changes in the *M. avium* metabolism are an indication that, in the resistant mice, the bacterial pathogen transitions to the latent state, because of the deficit in divalent metal ions. ●

This work was supported by the Russian Foundation for Basic Research (grant No. 08-04-01053), the Programme for Support of Leading Russian Scientific Schools (project NSh-5638.2010.4), the Programme "Molecular and Cell Biology" of the Presidium of RAS, and the National Institute of Health, USA (NIH; grant № AI078864 for A.S. Apt).

REFERENCES

1. Horsburgh C.R., Jr. // *N. Engl. J. Med.* 1991. V. 324. P. 1332-1338.
2. Inderlied C.B., Kemper C.A., Bermudez L.E. // *Clin. Microbiol. Rev.* 1993. V. 6. P. 266-310.
3. Nightingale S.D., Byrd L.T., Southern P.M., et al. // *J. Infect. Dis.* 1992. V. 165. P. 1082-1085.
4. Benson C.A. // *Clin. Infect. Dis.* 1994. V. 18. Suppl 3. P. S218-222.
5. Griffith D.E. // *Curr. Opin. Pulm. Med.* 1997. V. 3. P. 139-145.
6. Nolt D., Michaels M.G., Wald E.R. // *Pediatrics.* 2003. V. 112. P. e434.
7. Benini J., Ehlers E.M., Ehlers S. // *J. Pathol.* 1999. V. 189. P. 127-137.
8. Rook G.A., Hernandez-Pando R. // *Annu. Rev. Microbiol.* 1996. V. 50. P. 259-284.
9. Schluger N.W., Rom W.N. // *Am. J. Respir. Crit. Care Med.* 1998. V. 157. P. 679-691.
10. Kondratieva E.V., Evstifeev V.V., Kondratieva T.K., et al. // *Infect. Immun.* 2007. V. 75. P. 4762-4768.
11. Sambrook J., Fritsch E.F., Maniatis T. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, N.Y.; Cold Spring Harbor Laboratory Press, 1989.
12. Torrea G., Levee G., Grimont P., et al. // *J. Clin. Microbiol.* 1995. V. 33. P. 1899-1904.
13. Pedrosa J., Florido M., Kunze Z.M., et al. // *Clin. Exp. Immunol.* 1994. V. 98. P. 210-216.
14. Azhikina T., Skvortsov T., Radaeva T., et al. // *Biotechniques.* 2010. V. 48. P. 139-144.
15. Audic S., Claverie J. // *Genome Research.* 1997. V. 7. P. 986.
16. Kondratieva E., Logunova N., Majorov K., et al. // *PLoS One.* 2010. V. 5. P. e10515.
17. Skvortsov T., Azhikina T. // *Russ. J. Bioorg. Chem.* 2010. V. 36. P. 550-559.
18. Banaiee N., Jacobs W.R., Jr., Ernst J.D. // *Infect. Immun.* 2006. V. 74. P. 6449-6457.
19. Ramakrishnan L., Federspiel N.A., Falkow S. // *Science.* 2000. V. 288. P. 1436-1439.
20. Tullius M.V., Harth G., Horwitz M.A. // *Infect. Immun.* 2003. V. 71. P. 3927-3936.
21. Hendriks J., Oubrie A., Castresana J., et al. // *Biochim. Biophys. Acta.* 2000. V. 1459. P. 266-273.
22. Appelberg R. // *Immunol. Research.* 2006. V. 35. P. 179-190.
23. Shi L., Sohaskey C.D., Kana B.D., et al. // *Proc. Natl. Acad. Sci. USA.* 2005. V. 102. P. 15629-15634.
24. Ratledge C. // *Tuberculosis.* 2004. V. 84. P. 110-130.
25. Gold B., Rodriguez G.M., Marras S.A.E., et al. // *Molecular Microbiology.* 2001. V. 42. P. 851-865.
26. Bacon J., James B.W., Wernisch L., et al. // *Tuberculosis.* 2004. V. 84. P. 205-217.
27. Harper C., Hayward D., Kidd M., et al. // *BMC Microbiology.* 2010. V. 10. P. 138.
28. Amon J., Titgemeyer F., Burkovski A. // *J. Mol. Microbiol. Biotechnol.* 2009. V. 17. P. 20-29.
29. Warner D.F., Savvi S., Mizrahi V., Dawes S.S. // *J. Bacteriol.* 2007. V. 189. P. 3655-3659.
30. Malm S., Tiffert Y., Micklinghoff J., et al. // *Microbiology.* 2009. V. 155. P. 1332-1339.
31. Weber I., Fritz C., Ruttkowski S., et al. // *Mol. Microbiol.* 2000. V. 35. P. 1017-1025.
32. Casali N., White A.M., Riley L.W. // *J. Bacteriol.* 2006. V. 188. P. 441-449.
33. Glickman M.S., Cox J.S., Jacobs W.R. // *Mol. Cell.* 2000. V. 5. P. 717-727.
34. Waddell S.J., Chung G.A., Gibson K.J., et al. // *Letts. Appl. Microbiol.* 2005. V. 40. P. 201-206.
35. van der Geize R., Yam K., Heuser T., et al. // *Proc. Natl. Acad. Sci. USA.* 2007. V. 104. P. 1947-1952.
36. Muñoz-Eliás E., McKinney J. // *Nature Medicine.* 2005. V. 11. P. 638-644.
37. Wu C.W., Schmoller S.K., Shin S.J., Talaat A.M. // *J. Bacteriol.* 2007. V. 189. P. 7877-7886.

Fingerprint-like Analysis of “Nanoantibody” Selection by Phage Display Using Two Helper Phage Variants

S. V. Tillib*, T. I. Ivanova, L. A. Vasilev

Institute of Gene Biology, Russian Academy of Sciences

*E-mail: tillib@genebiology.ru

Received 06.07.2010

ABSTRACT This paper discusses the selection of mini-antibody (nanoantibody, nanobody® or single domain antibody) sequences of desired specificity by phage display-based method using a generated library of antigen-binding domains of special heavy-chain only antibodies (single-stranded antibodies) of immunized camel. A comprehensive comparison of the efficiency of parallel selection procedures was performed by using the traditional (M13K07) and modified (with N-terminal deletion in the surface gIII protein) helper phages. These two methods are partly complementary, and by using them in parallel one can significantly improve the selection efficiency. Parallel restriction analysis (fingerprinting) of PCR-amplified cloned sequences coding for mini-antibodies (HMR-analysis) is proposed for identifying individual clones, as a replacement to sequencing (to a certain extent). Using this method, unique data were collected on the selection of mini-antibody variants with the required specificity at various stages of a multi-stage selection procedure. It has been shown that different sequences coding for mini-antibodies are selected in different ways, and that, if this feature is not taken into account, some mini-antibody variants may be lost.

KEYWORDS immunisation, phage display, helper phage, recombinant mini-antibody, fingerprinting.

ABBREVIATIONS ELISA—enzyme-linked immunosorbent assay, BSA—bovine serum albumin, hp—helper phage, PCR—polymerase chain reaction, HA-tag – antigenic determinant, a fragment of 9 amino acids (YPYDVPDYA) from the hemagglutinin protein of the influenza virus, this epitope can be efficiently detected by commercially available antibodies, HMR-analysis—a method proposed in this paper of parallel restriction analysis (using three restriction enzymes - HinfI, MspI and, and RsaI) of PCR-amplified cloned mini-antibody sequences, TNF (TNF- α)—tumour necrosis factor, a multifunctional anti-inflammatory cytokine, PBS—phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4).

INTRODUCTION

This work was performed in the framework of projects aimed at developing new methods for passive immunisation for preventive care and immunotherapy of infectious diseases, particularly rabies and anthrax, in humans and animals. In this paper, we put special emphasis on an analysis of sequences encoding specific mini-antibodies at various stages of selection that is based on the phage display technology [1–3].

Starting from the establishing in 2003 of collaboration work with Prof. S. Muyldermans' laboratory at Vrije Universiteit Brussel, Belgium, we have been developing and using a novel technology for generating specific camel single domain mini-antibodies (also called nanoantibodies or nanobodies®) in the Institute of Gene Biology, Moscow, Russia.

Nanoantibodies are antigen recognition molecules that are fragments (variable domains) of unusual antibodies found in nature, along with classic antibodies, only in *Camelidae* family animals and in some cartilaginous fishes such as sharks and ratfish [4, 5]. These

“camel-specific” antibodies consist of dimers of a single, short (without the first CH1 domain) immunoglobulin heavy chain and are fully functional in the absence of the immunoglobulin light chain. Only a single variable domain of this antibody is necessary and sufficient in order to specifically recognize an antigen and bind to it. In contrast to the majority of recombinant antibodies, generated single-stranded antibodies usually demonstrate a rather high affinity to a given antigen, because the first stage of their formation takes place in the animal's organism (*Camelidae*), via affinity maturation *in vivo*. Nanoantibodies have several advantages over traditional antibodies and may have great prospective use in the future in research and in new biotechnological devices, as well as in clinical diagnostics and treatment. The advantages of nanoantibodies include their smaller size, new structural features (better penetration into tissues and the ability to recognize epitopes hidden, inaccessible to conventional antibodies), the possibility to be economically and efficiently mass-produced (in bacteria and yeast), good solubility, resistance to sig-

nificant changes in temperature and pH, and simplicity when used for various genetic engineering operations [5, 6]. It is important to note that these camel-variable domains do not cause an evident immune response in primates, and their structure is closely homologous to the variable domains of one subclass of human IgG immunoglobulins (IgG3). It has been shown that these camel mini-antibodies can be “humanized” without significant loss of their specific activity, via few point amino acid substitutions [7]. This indicates a potential for the broad use of mini-antibodies as a passive immunisation treatment to prevent various dangerous infections [5, 7–9].

In our method, the first critical stage of nanoantibody production is inducing specific antibodies in the camel’s (or llama) organism by immunisation, and the second critical stage is selecting the clones of the nucleotide sequences of the required nanoantibodies by phage display from the generated library of the entire repertoire of variable domains of special antibodies of the immunised animal. The latter stage is quite nontrivial and has not been studied extensively. In this work, we have carefully studied the selection of the required sequences and compared the techniques using two different helper phages. Recently, we proposed a modified helper phage for a more efficient selection of the antigen recognition domains of the special single-stranded camel antibodies (nanoantibodies) by phage display based on the filamentous phage M13 [10]. The use of a mutant M13KO7 phage called hpΔMBpIII (with N-terminal deletion in the M13 phage gIII protein, which makes the phage unable to infect the bacterial cells but does not stop the formation of the phage particle) as a helper phage can in some cases significantly increase the selection efficiency. This has been demonstrated on a model system in the final (third) selection stage of nanoantibodies binding to the tumour necrosis factor, TNF- α . The nanoantibodies were selected from a library that had been specifically pre-enriched in two traditional selection procedures. In this work, we performed a comprehensive comparison of the selection efficiency using both the traditional and mutant helper phages.

EXPERIMENTAL

Antigens and camel immunisation

A preparation of the Rabies virus and a recombinant protein corresponding to the *Bacillus anthracis* lethal factor synthesized in a bacterial expression system were used as antigens for camel immunisation. The anthrax lethal factor was kindly provided by Dr. A. Kolesnikov, Laboratory of Biocatalysis, Institute of Bioorganic Chemistry, RAS. A rabies vaccine based

on an inactivated attenuated RB 71/10 strain of Rabies virus, produced at the Pokrov bioplant (Vladimir region), was kindly provided by Prof. B. S. Naroditsky (N.F.Gamaleya Research Institute of Epidemiology and Microbiology). This vaccine was used to make the Rabies virus preparation. Virus particles were separated from the culture medium proteins by ultrafiltration through a membrane that is permeable to particles smaller than 300 kDa (Vivaspin 20 ml Concentrator, VS 2051, Vivascience, USA) but not to much larger virus particles. In parallel, the virus preparation was concentrated (approximately to 10^7 virus particles in 10 ml of standard PBS solution). The obtained preparation was divided into six equal parts and frozen at -70°C for storage (five parts were used for five successive immunisation injections, and the sixth part was used for the selection and analysis of mini-antibodies). Before each immunisation stage, one of the aliquots stored at -70°C was thawed, then combined with 650 μg of the *Bacillus anthracis* lethal factor and mixed with an equal amount of Freund adjuvant. Thus, a five-stage immunisation of a 6-year-old female *Camelus bactrianus* was performed by using the two antigen preparations described above, mixed with equal amounts of Freund adjuvant (complete—only for the first injection, and the other 4 times—with incomplete). The second injection was made 4 weeks after the first, while the other three injections were made in 10-day intervals, one after another. Blood samples were taken 5 days after the last injection.

Construction of a nucleotide sequence library of nanoantibodies

DNA cloning of the entire set of variable domains of specific single-stranded antibodies from the B-lymphocytes of immunised camel peripheral blood was performed in accordance with the standard procedure (RNA isolation, cDNA synthesis, two-stage PCR with specific primers, insert restriction and purification, and insertion into phagemid vector [11–14]). This procedure resulted in a new specific nucleotide sequence library of mini-antibodies in the pHEN4 phagemid vector [11]. This library consisted of two sub-libraries, which were different in the restriction sites used for insertion into the phagemid vector of sequences coding for antigen recognition domains, (i) NcoI + NotI and (ii) PstI + NotI. *Escherichia coli* cells (strain TG1) were transformed by electroporation with recombinant plasmid DNA obtained by ligation of the two insert sub-libraries and two corresponding vector sequences. Standard cloning, bacterial and bacteriophage M13 procedures were performed following the techniques described in ref. [15, 16]. The “Basic library of immunised camel nanoantibodies” satisfied the criteria of 10^7 clones and about 80%

of inserts of the expected size. Nanoantibodies with the required specificity were selected from the library.

Selection of nanoantibodies by phage display

The selection of camel mini-antibodies (nanoantibodies) by phage display, production of nanoantibodies in the bacterial periplasm, and the analysis of the nanoantibodies' ability to recognize a given antigen were performed using the described techniques [11–14], with the following modifications. For selection and ELISA, MICROLON 600 (Greiner Bio-One) high binding polystyrene microtiter plates were used. Five percent Skim milk (Bio-Rad) and 1% BSA (Sigma-Aldrich) in PBS were used as blocking agents. The working concentrations of ampicillin and kanamycin were 100 μ g/ml and 70 μ g/ml, respectively.

E. coli K12 / TG1 {*supE thi-1 (lac-proAB) (mcrB-hsdSM)5(rK- mK-)* [*F' traD36 proAB lacIqZ M15*]} (Stratagene) and *E. coli* K12 / WK6 Δ (*lac-pro*), *galE, strA, nal*; *F' lacIqZ* Δ M15, *pro+*) strains were used [17], as well as M13K07 bacteriophage (New England Biolabs) [18] and pHEN4 phagemid [11]. The *E. coli* strains and phagemid were provided by Prof. S. Muyl-dermans (Vrije Universiteit Brussel, Belgium). We had previously prepared hp Δ MBpIII modified bacteriophage [10] by deleting a region between the MspAI (1634) and BamHI (2221) sites from the M13K07 genomic DNA. The helper phage corresponding to the mutant bacteriophage was produced by a two-stage sequential transformation of *E. coli* cells (strain TG1). At the first stage, the cells were transformed with pHEN4 Δ f1ori mutant DNA with a deleted fragment coding for replication initiation of the f1 filamentous phage (between the EcoRI (1650) and Alw44I (2345) sites). This deletion preserves the plasmid's ability to replicate as a two-stranded DNA (by still having the corresponding section for replication initiation), but it is no longer capable of either phagemid single-stranded replication nor of packing to a phage particle. At the second stage, the cells containing pHEN4 Δ f1ori, and therefore resistant to ampicillin, were transformed with mutant genomic DNA (hp Δ MBpIII) of the M13K07 bacteriophage (replicative two-stranded form) containing the kanamycin resistance gene. Then, the cells were cultured with both antibiotics. One of the cell colonies was used to produce the modified helper phage. The phage was grown overnight in 2xTY medium containing kanamycin and 1 mM of isopropyl- β -D-1-thiogalactopyranoside (used for activating the synthesis of the wild-type gIII protein, coded for by pHEN4). Then, the phage was purified via the traditional PEG precipitation technique [15, 16]. The resulting phage is effective in infecting bacterial cells; however, its genomic DNA only codes for the gIII mutant protein. As

a result, the daughter phages cannot infect bacteria in the absence of an additional plasmid coding for either the wild-type gIII or gIII with a mini-antibody inserted sequence.

Nanoantibodies binding to the Rabies virus preparation or anthrax lethal factor were selected from the same library using two different methods in parallel. The first method was traditional phage display with the M13K07 helper phage in three selection/amplification cycles. The blocking agents used for the cycles were 5% Skim milk (Bio-Rad Laboratories, USA) in PBS in the first cycle, 1% BSA (Amresco, USA) in PBS in the second cycle, and the 5% milk again in the third cycle. On average, $\sim 10^{11}$ phage particles from the original library were introduced into a well with an immobilized antigen; after incubation and washing, $\sim 10^6$ phage particles were eluted and remained active after the first selection stage, while 10 times more ($\sim 10^7$ out of $\sim 10^{11}$) amplified phage particles remained active after the second selection stage. As a result of the third stage, $\sim 10^8$ out of loaded $\sim 10^{11}$ phage particles were selected, which is, similarly, about 10 times more than after the previous stage. This is usually an indication that the selection proceeds well and that specifically binding phage particles proliferate, with mini-antibodies getting exposed on the particle's surface, presumably ensuring its specific binding properties. After the last elution and neutralisation, the phage particle solutions at various dilution levels were used to infect *E. coli* cells (WK6 strain), then separate colonies were prepared on a Petri dish for the final analysis.

In parallel to the traditional procedure, nanoantibodies able to bind to the Rabies virus preparation or anthrax lethal factor were selected from the same library, using the mutant hp Δ MBpIII helper phage with N-terminal deletion on the surface gIII protein [10]. The selection procedure was analogous to that described for the traditional helper phage. We should note that, in this case, the number of colonies grown after each selection stage (which corresponds to the number of active phage particles with nanoantibodies) was significantly smaller than in the case when the traditional helper phage was used (100 times less after the first stage and 10 times less after the second stage). Helper phage was less reliable for subsequent infection of bacteria and amplification of eluted phage particles, which sometimes resulted in the loss of selected clones. Infecting the WK6 strain cells with the modified phage after elution is apparently ineffective (and almost impossible); therefore, in the case of the modified helper phage, *E. coli* cells of the TG1 strain were only used for infection with eluted phage particles, after all selection stages, including the final one.

A periplasmic extract containing a mini-antibody with a C-terminal HA tag was used to estimate the specificity and efficiency of the nanoantibody binding to the preparation of the antigen immobilised in the immunological plate with traditional ELISA [13, 14]. Anti-HA-monoclonal antibodies (CHGT-45P-Z, ICL, Inc., USA) conjugated with horseradish peroxidase were used as secondary antibodies to the HA-tag. Horseradish peroxidase activity was determined by using ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonate) as a chromogenic substrate. The absorbance was measured with a plate fluorimeter at 405 nm. There was no antigen in the control wells, but they were also blocked and processed in parallel with the experimental wells containing the antigen. The number of "+" signs in the "ELISA signal" column in Tables 1 and 2 corresponds to the relative ELISA absorbance value (reflects the summarized efficiency of expression/availability and binding of the nanoantibody from the periplasmic extract to the antigen immobilised in the well).

RESULTS AND DISCUSSION

HMR analysis of nanoantibody cloned nucleotide sequences

For interim analysis of the mini-antibody clone samples, PCR-amplification (on a given colony directly, colony PCR) is usually performed. A nucleotide sequence coding for the mini-antibody with small adjacent sections of the pHEN4 vector phagemid is to be amplified. The PCR product is then subjected to restriction endonuclease (restrictase) treatment, usually HinfI. In the sequences coding for the nanoantibodies, there are both relatively conservative regions, which contain conservative HinfI recognition sites, and rather long hypervariable regions (the third region, CDR3, is the longest one). In the hypervariable regions, the number and location of the sites recognized by restrictases vary remarkably, hence the strong variation in the number and size of the split DNA fragments. We have found that in order to identify a given clone (a mini-antibody sequence variant), parallel digestion of the PCR product with three different restrictases (HinfI, MspI, RsaI) yields the most reliable results.

A forward primer, RP (5'-cacacaggaacagctatgac-3'), and a reverse primer, GIII (5'-ccacagacagcctcatag-3'), were used for PCR amplification of the cloned sequence coding for the nanoantibody with small adjacent segments of the pHEN4 phagemid. PCR was carried out in a volume of 20 μ l, then the mixture was distributed into three tubes (6 μ l of mixture into each tube), where parallel treatment of the PCR product was done (in 15 μ l total volume) with three different restrictases, one in each tube: HinfI (Fermentas, Lithuania), MspI

(Fermentas, Lithuania), and RsaI (SibEnzyme, Russia). The three hydrolysates were loaded into adjacent 2.5–3% agarose gel wells for electrophoresis. The "low range" mixture (Fermentas) was used as a DNA fragment marker. The resulting electrophoretogram (three gel lanes with DNA treated with the HinfI, MspI, and RsaI restrictases, respectively, and the fourth lane with marker DNA) is a reliable fingerprint image of a given mini-antibody coding sequence cloned in the pHEN4 phagemid. We called this extended fingerprint analysis technique, which can identify a variant of a cloned mini-antibody sequence, "HMR analysis." We used the HMR analysis to study the selection results (samples of mini-antibody clones).

Figure 1 shows the HMR analysis data of some (presumably the most common) clones of the initial mini-antibody library obtained as a result of immunisation and subsequent cloning. We selected and analysed all 94 clones grown after 10^{-5} dilution of the base library consisting of two sub-libraries. We found 85 fingerprints, with only 9 of them repeating twice in this sample, which meant that the library was very heterogeneous without dominating sequences. The majority of the fingerprints (61) corresponded to the clones in which the sequences coding for the nanoantibody were inserted into the PstI–NotI vector restriction sites. The minority of the fingerprints (33) denoted with the letter "N" corresponded to the clones in which the sequences coding for the nanoantibody were inserted into the NcoI–NotI vector restriction sites. Very similar fingerprints found in both sub-libraries are underlined. *Figure 1* confirms that the parallel use of three restrictases instead of one reveals the differences in many more clones. Thus, if we had used only HinfI, we would not have been able to distinguish the following clone groups: 1) 3, 38, 52, 53, and 66 and 2) 55, 56N, 60N, 62N, 74, and 75N. Analogously, much fewer clones could have been differentiated using only one of the other two restrictases: MspI or RsaI.

This result indicates a rather wide diversity of clones in the original library and the high-resolution capability of the HMR analysis. We should note that when only one restrictase is used (out of the three: HinfI, MspI, and RsaI), the number of different fingerprints is much smaller and, hence, the analysis less reliable; therefore, we decided that it was essential to use the full version of the HMR analysis at all further stages of the mini-antibody clone analysis.

We used the following algorithm for nanoantibody sampling. With the HMR analysis, we studied a series of the most represented clones (those that formed no fewer than 24 colonies at the highest dilution of the library or of the eluted phage particles). All clones with unique fingerprints were used to produce the

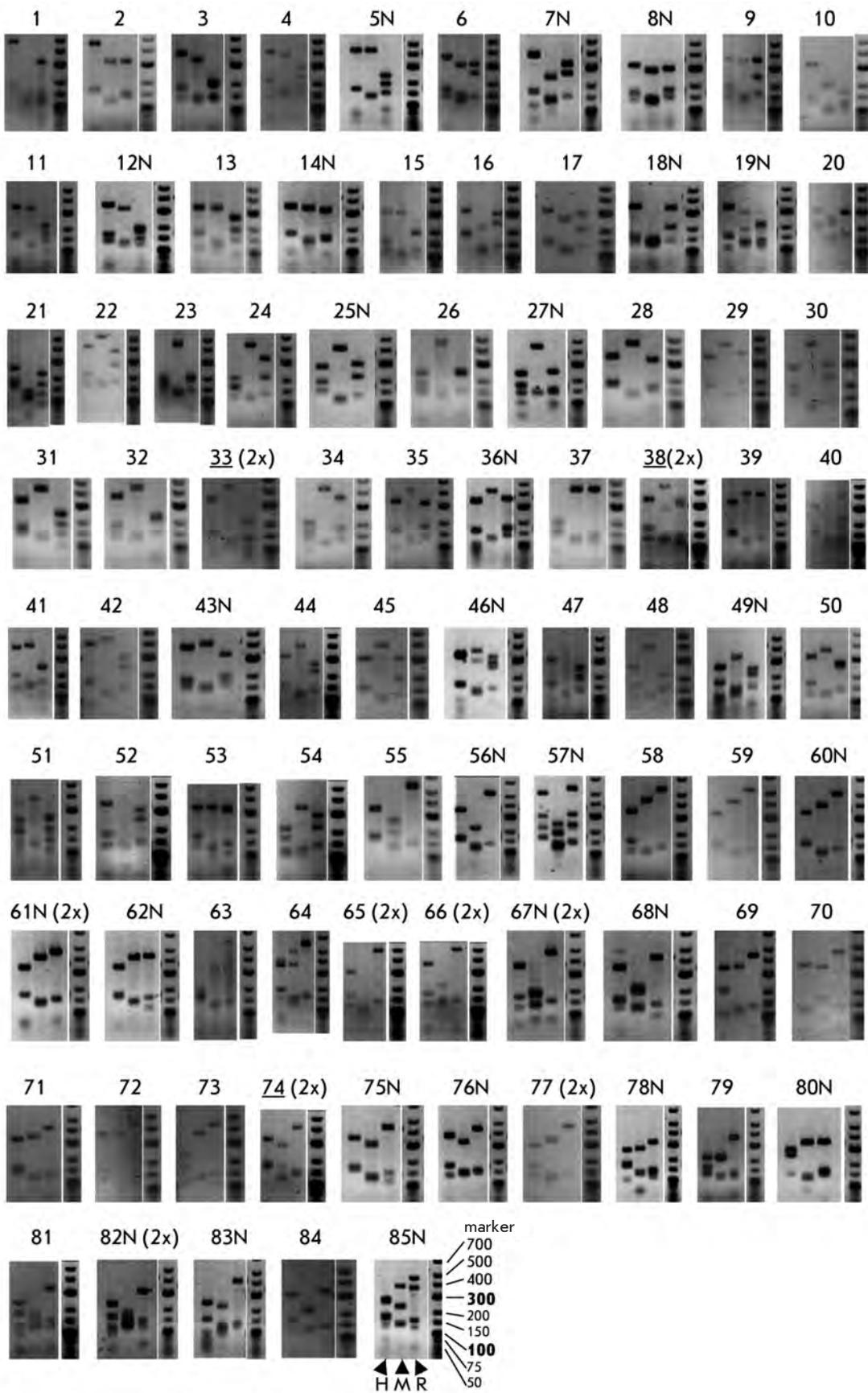


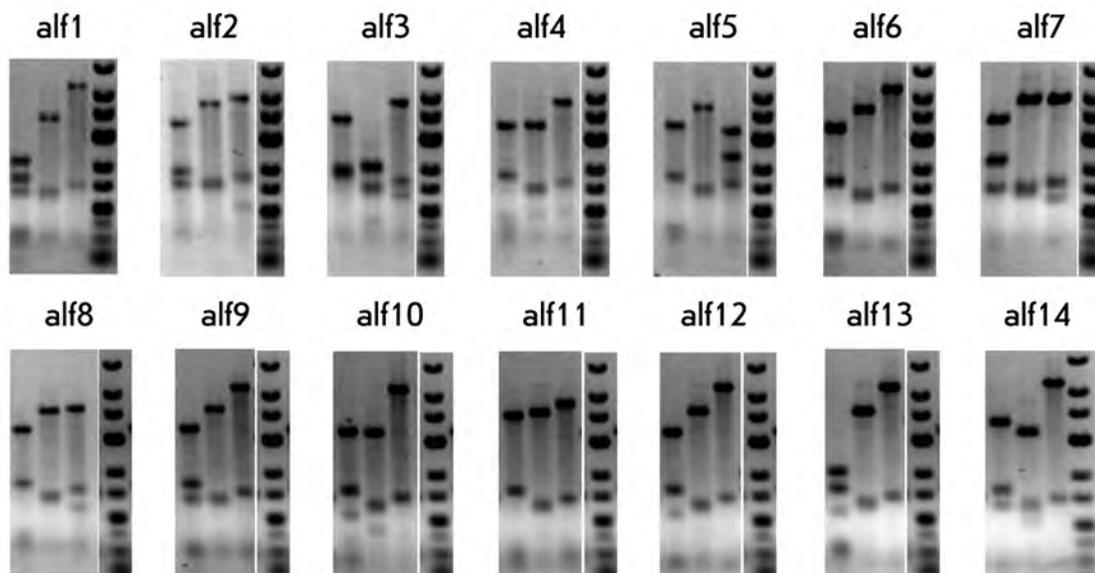
Fig. 1. HMR analysis data (HMR fingerprints) of clones from the original library of cloned mini-antibody nucleotide sequences. Ninety-four clones were selected and analysed. Eighty-five different fingerprints were identified with only 9 of them twice as repeats in this set (marked as 2x). The majority of fingerprints correspond to clones with PstI-NotI insertions of adapted nanoantibody sequences. The smaller part of fingerprints (marked by the letter N) corresponds to clones with NcoI-NotI insertions of adapted nanoantibody sequences into the plasmid vector. Fingerprints found in both sub-libraries are underlined. Each HMR fingerprint is an electrophoretogram consisting of three gel lanes with separated DNA fragments obtained after parallel treatment of the PCR product (an amplified nanoantibody sequence) with one of three restrictases, Hinfl (H), MspI (M) or RsaI (R), and the fourth lane with marker DNA (the sizes of the marker DNA fragments are shown at the bottom of the figure).

Table 1. Relative representation of fingerprint variants of selected nanoantibody clones recognizing the anthrax lethal factor (alf), at various selection stages.

alf clone number	1 st stage, normal helper phage	1 st stage, mutant helper phage	2 nd stage, normal helper phage	2 nd stage, mutant helper phage	3 rd stage, normal helper phage	3 rd stage, mutant helper phage	ELISA signal
1	3/28	0/27	3/30	0/43	3 /42	5 /43	++
2	0/28	1/27	3 /30	0/43	0/42	0/43	+++
3	1/28	0/27	0/30	0/43	3 /42	3 /43	++
4	2/28	0/27	0/30	0/43	1 /42	0/43	+
5	2/28	0/27	0/30	1/43	0/42	1 /43	+
6	3/28	0/27	12 /30	0/43	12 /42	13 /43	++
7	0/28	0/27	0/30	0/43	0/42	1 /43	+++
8	0/28	0/27	0/30	0/43	0/42	2 /43	++
9	1/28	0/27	0/30	1/43	1 /42	1 /43	++
10	1/28	0/27	3 /30	0/43	2 /42	1 /43	+++
11	0/28	0/27	0/30	0/43	0/42	1 /43	++
12	0/28	0/27	0/30	0/43	1 /42	1 /43	++
13	1/28	0/27	2/30	0/43	6 /42	2 /43	++
14	0/28	0/27	2/30	1/43	2 /42	2 /43	+++

Note: Here and in Table 2 the clones used for the production and ELISA of corresponding nanoantibodies are in bold font. The number of “+” signs correlated with the relative increase in the ELISA signal (absorbance).

Fig. 2. HMR fingerprints of the selected nanoantibody clones that bind in ELISA to immobilised recombinant anthrax lethal factor. These clones are designated as “alf” with a number.



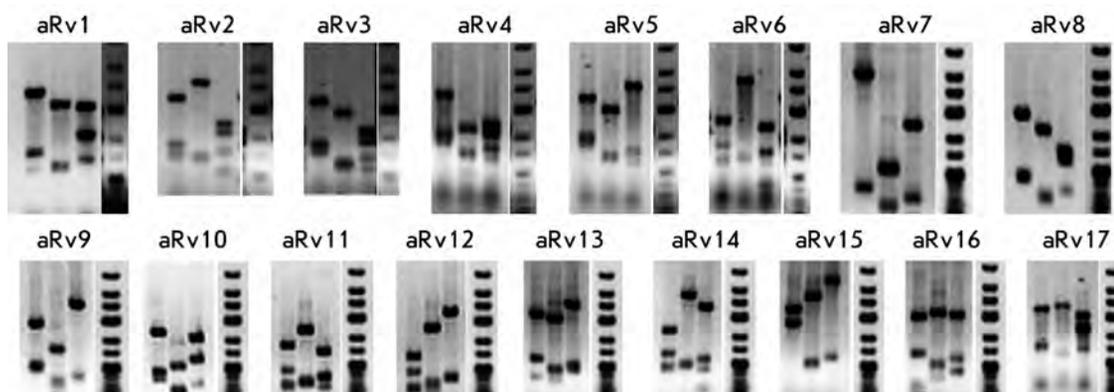
corresponding nanoantibodies in the host bacteria's periplasm. The periplasmic extracts were then tested with ELISA for the presence of a nanoantibody that binds to a given antigen. The same preparations of the Rabies virus and anthrax lethal factor that were used for camel immunisation were used as antigens for the selection procedures and ELISA. The increase in the absorbance value at 405 nm in the plate wells reflected two processes: the level of expression/availability of the nanoantibody and the strength of its interaction with the immobilized antigen. Information pertaining

to the selection of the final nanoantibody clones binding to a given antigen is presented in *Figs. 2 and 3* and in *Tables 1 and 2*.

It is necessary to exert caution when interpreting the data. It is possible that some clones differing in nucleotide and amino acid sequences will have similar fingerprints, especially when sampled clones are compared to random clones from the original library. Sequencing can, naturally, provide more information. Our experience, however, shows that in most cases of clone identification the HMR analysis can replace

Table 2. Relative representation of fingerprint variants of selected nanoantibody clones recognizing Rabies virus preparation (aRv), at various selection stages.

aRv clone number	2 nd stage, normal helper phage	2 nd stage, mutant helper phage	3 rd stage, normal helper phage	3 rd stage, mutant helper phage	ELISA signal
1	0/67	3/45	1/30	2/58	+
2	7/67	1/45	5/30	2/58	+
3	13/67	0/45	8/30	11/58	++
4	0/67	0/45	2/30	0/58	+++
5	4/67	1/45	1/30	0/58	+
6	0/67	0/45	1/30	0/58	+
7	0/67	0/45	0/30	1/58	+++
8	2/67	0/45	0/30	0/58	+
9	1/67	1/45	0/30	0/58	+
10	1/67	0/45	0/30	0/58	+
11	1/67	0/45	0/30	0/58	++
12	2/67	0/45	0/30	0/58	+++
13	4/67	0/45	0/30	2/58	++
14	0/67	2/45	0/30	0/58	++
15	1/67	0/45	0/30	0/58	+
16	1/67	0/45	0/30	0/58	+++
17	3/67	1/45	2/30	1/58	++

Fig. 3. HMR fingerprints of the selected nanoantibody clones that bind in ELISA to immobilised Rabies virus. These clones are designated as "aRv" with a number.

sequencing; therefore, it makes it a more economical technique.

HMR analysis of nanoantibodies recognizing anthrax lethal factor

Figure 2 shows the HMR fingerprints of 14 variants of the selected nanoantibody samples (alf) that recognize the anthrax lethal factor. There were only 5 out of the 14 for which we could find similar fingerprints among the 85 variants from the original library - the corresponding pairs being alf3—68N, alf6—59, alf8—62N, alf9—58, and alf10—64. Even though these 5 pairs are similar, that doesn't necessarily make them identical. Thus, the alf8 clone fingerprint variant was not found among 28–30 analyzed clones after the first and second selection stages; it appeared only after the third

selection stage and only when the mutant helper phage was used. On the other hand, the alf6 clone fingerprint variant may well correspond to one of the most representative variants in the original library, since this variant remains one of the most represented after each selection stage.

We would like to note here a rather unexpected phenomenon related to the modified selection method using the hpΔMBpIII helper phage. When nanoantibodies to each of the two antigens were being selected (Tables 1 and 2), enrichment of specific clone variants was observed during the first two selection stages, which were different from the 85 variants presented in Fig. 1 and those enriched using the traditional helper phage. Unfortunately, the nanoantibodies coded for by those clones did not bind to the corresponding antigens

in ELISA. Those clones were effectively sorted out at later selection stages. We believe that those temporarily selected clones bind to the blocking agent and are then sorted out when the blocking agent is replaced. Apparently, some nanoantibodies, more easily available on the surface of the mutant phage, can either bind to a component of Skim milk (at the first stage) or to BSA (at the second stage). At the third selection stage, when Skim milk was used again, a sharp clearing of the nonspecific background occurred. The nonspecific clones enriched at the first and second stages were disappearing, and the fingerprints of the sought specific clones appeared, many of them identical to the clones selected in the parallel traditional procedure (Table 1, clones alf1, alf3, alf6, alf9, alf10, alf12, alf13, and alf14). Interestingly, similar to our previous work [11], some nanoantibody variants could only be selected with the modified selection procedure (Table 1, clones alf5, alf7, alf8, and alf11). There were also variants that could only be selected with the standard procedure (clone alf2 at the second stage and clone alf4). Interestingly, the nanoantibody clones that produce the strongest ELISA absorbance signal were not necessarily the most represented among the selected clones and could even disappear after the subsequent selection stages (alf2). The most enrichment was observed for the clones that produced medium intensity ELISA absorbance signal.

HMR analysis of nanoantibodies recognizing Rabies virus preparation

The findings described above were reaffirmed during the selection of nanoantibodies (aRv) recognizing the Rabies virus preparation (Fig. 3, Table 2).

In that case, a similar fingerprint could only be found for 5 out of the 17 finally selected clones among the 85 variants in the original library: aRv2—32, aRv3—12N, aRv5—75N, aRv13—76N, and aRv17—46N. The aRv3 clone fingerprint may well be one of the most represented variants in the original library, since this variant remains the most represented after (repeating) each selection stage with the standard helper phage. When the mutant helper phage was used for the selection of this fingerprint variant, it was clear that, as in the case of selection of nanoantibodies recognizing the anthrax lethal factor, the nonspecific background cleared after the third stage with blocking agent replacement. In this case, the parallel use of the traditional and the modified selection methods also ensured a wider variety of selected nanoantibody clones. Thus, in using only the traditional procedure, the fingerprint variants corresponding to the following clones were selected: aRv4, aRv6, aRv8, aRv10, aRv11, aRv12, aRv15, and aRv16. But while using the modified procedure alone,

variants aRv7, aRv13, and aRv14 were selected. The aRv13 variant was also selected after the second stage of the traditional procedure, but it disappeared after the third stage.

It is interesting that in this case as well, the nanoantibody clones producing the strongest ELISA absorbance signal were not the most represented among the selected clones. Only a small number of the aRv4, aRv7, aRv12, and aRv16 clones were selected, and by using only one of two selection methods. Some of these clones (aRv12 and aRv16) can be easily lost at the third selection stage.

CONCLUSION

Parallel selection using the traditional helper phage and a modified helper phage with N-terminal deletion in the gIII surface protein is recommended for improving the efficiency of phage display selection of nanoantibodies with a required specificity. The modified procedure should take into account the higher nonspecific background, which is apparently due to the selection of phage particles containing nanoantibodies that bind to the blocking agents or to other nonspecific components. We cannot exclude a possibility of specific binding to the antigen of a special kind of nanoantibody, which is happening only when it is exposed on the phage's surface and which disappears in the case of stand-alone nanoantibody. The sequential use of different blocking agents in combination with the proposed mutant helper phage reduces the nonspecific background significantly after three selection stages. One should be aware that some important nanoantibodies might be lost during additional amplification/selection stages. We recommend analyzing selected clones after the second and third selection stages when performing the traditional procedure, and after the third stage when performing the modified procedure. Good results are also obtained when the modified method is used after the initial stage of the traditional selection procedure. ●

This work was supported by the Ministry of Education and Science of Russia (02.512.11.2320), by the Program № 27 of the Presidium of the Russian Academy of Sciences "Fundamentals of Basic Research of Nanotechnologies and Nanomaterials" (3.1.2.).

The authors are grateful to Prof. S. Muyltermans, Vrije Universiteit Brussel, Belgium, for providing the pHEN4 phagemid and the E. coli TG1 and WK6 strains, and to M. V. Rutovskaya and the staff members at Chernogolovka research-experimental base of A. N. Severtsov Institute of Ecology and Evolution (RAS) for their help in camel maintenance and immunisation.

REFERENCES

1. Brissette R., Goldstein N.I. // *Methods Mol. Biol.* 2007. V. 383. P. 203–213.
2. Sidhu S.S., Koide S. // *Curr. Opin. Struct. Biol.* 2007. V. 17. № 4. P. 481–487.
3. Hoogenboom H.R. // *Nat. Biotechnol.* 2005. V. 23. № 9. P. 1105–1116.
4. Hamers-Casterman C., Atarhouch T., Muyldermans S., et al. // *Nature*. 1993. V. 363. P. 446–448.
5. Ghassabeh G.H., Muyldermans S., Saerens D. // *Current Trends in Monoclonal Antibody Development and Manufacturing* / Eds Shire S.J., et al. New York: Springer, 2010. P. 29–48.
6. Deev S., Lebedenko E. // *Acta Naturae*. 2009. V. 1. P. 32–50.
7. Vincke C., Loris R., Saerens D., et al. // *J. Biol. Chem.* 2009. V. 284. № 5. P. 3273–3284.
8. Wesolowski J., Alzogaray V., Reyelt J., et al. // *Med. Microbiol. Immunol.* 2009. P. 157–174.
9. Muyldermans S., Baral T.N., Retamozzo V.C., et al. // *Vet. Immunol. Immunopathol.* 2009. V. 128. P. 178–183.
10. Vyatchanin A.S., Tillib S.V. // *Biotekhnologia (Biotechnology)*. 2008. № 4. P. 32–34.
11. Ghahroudi M.A., Desmyter A., Wyns L., et al. // *FEBS Lett.* 1997. V. 414. P. 521–526.
12. Nguyen V.K., Desmyter A., Muyldermans S. // *Adv. Immunol.* 2001. V. 79. P. 261–296.
13. Conrath K.E., Lauwereys M., Galleni M., et al. // *Antimicrob. Agents Chemother.* 2001. V. 45. P. 2807–2812.
14. Saerens D., Kinne J., Bosmans E., et al. // *J. Biol. Chem.* 2004. V. 279. P. 51965–51972.
15. Sambrook J., Fritsch E.F., Maniatis T. *Molecular Cloning: A Laboratory Manual*. Edn. 2. Cold Spring Harbor, N.Y.; Cold Spring Harbor Laboratory Press, 1989.
16. Barbas III C.F., Burton D.R., Scott J.K., Silverman G.J. *Phage Display: A Laboratory Manual*. Cold Spring Harbor, N.Y.; Cold Spring Harbor Laboratory Press, 2001.
17. Zell R., Fritz H.J. // *EMBO J.* 1987. V. 6. P. 1809–1815.
18. Vieira J., Messing J. // *Methods Enzymol.* 1987. V. 153. P. 3–11.

Regulation of Store-Operated Channels by Scaffold Proteins in A431 Cells

A. V. Shalygin¹, M. A. Ryazantseva¹, L. N. Glushankova¹, I. B. Bezprozvanny², G. N. Mozhayeva², E. V. Kaznacheeva¹

¹Institute of Cytology, Russian Academy of Sciences

²Department of Physiology, University of Texas Southwestern Medical Center, Dallas, USA

*E-mail: evkzn@hotmail.com

Received 29.08.2010

ABSTRACT Store-operated channels are major calcium influx pathways in nonexcitable cells. Homer scaffold proteins are well known for their role in regulating calcium signaling. Here we report on a detailed single-channel level characterization of native store-operated channels regulated by Homer scaffold proteins in A431 carcinoma cells. By applying the single-channel patch-clamp technique, we found that different types of store-operated calcium channels have different sensitivities to Homer proteins.

KEYWORDS SOC, Homer, A431

ABBREVIATIONS SOC - store-operated calcium channels, mGluR - metabotropic glutamate receptors, IP₃ - inositol 1,4,5-trisphosphate, IP₃R - the inositol 1,4,5-trisphosphate receptor, RyR - ryanodine receptors, UTP - uridine triphosphate, GST - glutathione S-transferase.

INTRODUCTION

Currently, calcium signaling in nonelectroexcitable cells is arousing substantial interest from researchers, because alterations of cytoplasmic Ca²⁺ regulate a plethora of intracellular events. The calcium concentration in cytosol increases upon the release of Ca²⁺ ions from intracellular stores either when extracellular Ca²⁺ enters the cell. In nonelectroexcitable cells, the entrance of Ca²⁺ is basically mediated by store-operated channels (SOC) [1]. These channels get activated upon the depletion of intracellular Ca²⁺ stores. SOCs from various tissues were shown to have different biophysical properties, indicating that there is a difference in their molecular composition [1]. Although the basic proteins which mediate store-operated Ca²⁺ influx are known, the mechanism underlying the colocalization of these proteins remains undiscovered.

Homer proteins that are found in neural tissue mediate the formation of the intermolecular complex, which includes the metabotropic glutamate receptor (mGluR) of the plasma membrane and inositol-1,4,5-trisphosphate receptor (IP₃R) [2]. It was shown later that Homer interacts with TRPC proteins [3,4], the Ca²⁺-ATPase, ryanodine receptor (RyR), and other proteins [5]. Moreover, in neural cells, Homer and Shank proteins were shown to form a meshlike structure which acts to organize postsynaptic proteins [6]. Homer proteins contain the EVH1 domain, which is located on the N-terminus of the polypeptide chain and accounts for the interaction between Homer and its targets by binding to the consensus amino acid sequence PPXXF (proline, proline, any two amino acids, and phenylalanine). There

are short and long isoforms of Homer proteins (Fig. 1A). The long isoform has a coiled-coil domain on its C-terminus which mediates oligomerization. The short isoform, which is produced by alternative splicing, lacks this domain [5]. The long isoforms were shown to form tetramers with the parallel arrangement of C-termini and four EVH1 domains, which makes it possible to mediate the colocalization of proteins of Ca²⁺ signaling [7]. Short Homer isoforms are believed to act as the negative regulators of the long one, since they are not able to form oligomers. It was shown that Homer proteins not only regulate protein colocalization, but they can also modulate the activity of mGluR [8], RyR [9], and TRPC [3, 4].

It was hypothesized that Homer proteins participate in the regulation of SOC. Previous studies confirming this suggestion were carried out by means of patch clamp in a whole-cell configuration and fluorescent imaging of Ca²⁺ signaling [3, 4]. However, these approaches make it possible to determine the net Ca²⁺ influx only. Since the cell contains various types of calcium-SOC, it has been unclear which of them are regulated by Homer proteins. In this study we set out to elucidate the role that Homer proteins play in the regulation of store-operated calcium channels in A431 cells.

EXPERIMENTAL PROCEDURES

Cells

Human epidermal carcinoma A431 cells (Cell lines collection, Cytology institute, Russian Academy of Sciences) were cultivated in a DMEM medium supplemented

with 10% fetal calf serum and antibiotics (100 µg/ml penicillin and 100 U/ml streptomycin). The cells were grown on microscope cover glasses 2–4 days prior to the experiment

Materials

DMEM medium (ICN), fetal calf serum (FCS, GIBCO BRL, United States), fetal bovine serum (FBS, GIBCO BRL, United States), and genitacin G-418 (Genitacin, GIBCO BRL, United States) were used in cell cultivation. Glutathion-sepharose and 1,5-isopropylthio-β-D-galactoside (IPTG) were from Pharmacia, Sweden; inositol-1,4,5-trisphosphate (IP₃) was from LC Laboratories, United States; uridine triphosphate (UTP) was from Cabiochem, Germany; EGTA was from Fluka, Switzerland; HEPES, Triton X-100 anti-GST antibodies, and secondary rabbit and mouse antibodies were purchased from Sigma, United States; and anti-Homer 1 bc antibodies were from Santa Cruz, United States. Anti-IP₃R T443 polyclone antibodies were described earlier [10]. PPKKFR and PPKKRR peptides were synthesized by Diapharm, Russia.

Patch Clamp

In all the experiments, the potential of extracellular solution was taken as zero.

In inside-out configuration experiments, the intracellular solution (chamber) contained (mM) 140 K-glutamate, 5 NaCl, 1 MgCl₂, 10 HEPES-K, pH 7.4, 2 EGTA-K, and 1.13 CaCl₂ (pCa 7). The pipette solution contained (mM) 105 BaCl₂ and 10 Tris-HCl (pH 7.3). The presence of Ba²⁺-ions led to the inhibition of voltage-dependent Ca²⁺ channels and Ca²⁺-dependent channels. The electric resistance of solution-filled pipettes was 8–20 MΩ.

In whole-cell configuration experiments, the pipette solution contained (mM) 145 NMDG aspartate, 10 Cs-EGTA, 10 Cs-HEPES, pH 7.3, 1.5 MgCl₂, and 4.5 CaCl₂ (pCa 7.0). The intracellular solution contained (mM) 140 NMDG aspartate, 10 BaCl₂, 10 Cs-HEPES, and pH 7.3. In these experiments we used pipettes which had resistances of 3–5 MΩ. In all the whole-cell experiments, the membrane potential was equal to 0 mV. The potential was changed according to the following scheme: in the beginning, the –100 mV potential was applied for 60 ms, which was followed by a voltage ramp in a –100- to 100-mV interval for 600 ms; after that, the 0 mV potential was restored. The whole-cell currents were normalized with respect to cellular capacity, which reflects the size of the cell. The average capacity value was 21±4 pF (25 experiments).

Currents were registered by means of an Axopatch 200B intensifier (Axon instruments, United States). The signal was digitized at a frequency of 5000 Gz with

ADC L305 board (L-Card, Russia). For an analysis and presentation of data on currents of individual channels of low conductivity, additional filtration was performed (80–100 Gz). Amplitudes of currents through individual channels of low conductivity were determined from registered currents and an amplitude histogram. The level of channel activity was characterized by the NP_o value, which is the product of the number of conducting channels (N) and the probability of the open state (P_o), which is equal to $P_o = I/(iN)$, where I is the average value of a current in the given membrane fragment at a certain time interval and i is the open channel current amplitude.

The registered data were digitized and analyzed by means of software developed by V.A. Alexeyenko, as well as pClamp 6.0.4, Microcal Origin 6.0, and Microsoft Excel.

Electrophoresis and Immunoblotting

Protein samples were analyzed in 8% polyacrylamide gels under denaturing conditions. Gels were stained with Coomassie or, alternatively, proteins were transferred to the membrane and visualized by specific antibodies against the proteins of interest.

Expression and Purification of Recombinant GST-Homer 1c and Homer 1a Proteins

Escherichia coli BL-21(DE3) cells were transformed with pGEX-2T-Homer1A and pGEX-2T-Homer1C plasmids (courtesy of M.M. Solovyev, University of Oxford, Great Britain). The expression of recombinant proteins was induced by the addition of 1 mM IPTG; then, bacterial cells were lysed, and chimeric GST-Homer proteins were purified on glutathion-sepharose. Proteins were stored at 4°C. The purity of protein preparations was checked by electrophoresis and immunoblotting with polyclone antibodies against GST and Homer.

Cellular Lysate

A431 cells were lysed for 10 min at 4°C in a 10-cm plate in 1 ml of solution containing 150 mM NaCl, 20 mM Tris-HCl (pH 7.6), 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.5% NP40, 10% glycerol, and 0.5 mM PMSF with the addition of a protease-inhibiting cocktail (PIC, Hoffmann-La Roche AG, Switzerland). The lysates were passed through a syringe three times and spun at 22 000g for 30 min at 4°C. The supernatant was withdrawn and used in subsequent experiments.

Pull-Down Assay (PD)

Glutathion-sepharose (25 µl) with bound chimeric GST-Homer was mixed with A431 cellular lysates and incubated on a rocking platform for 12–24 h. The re-

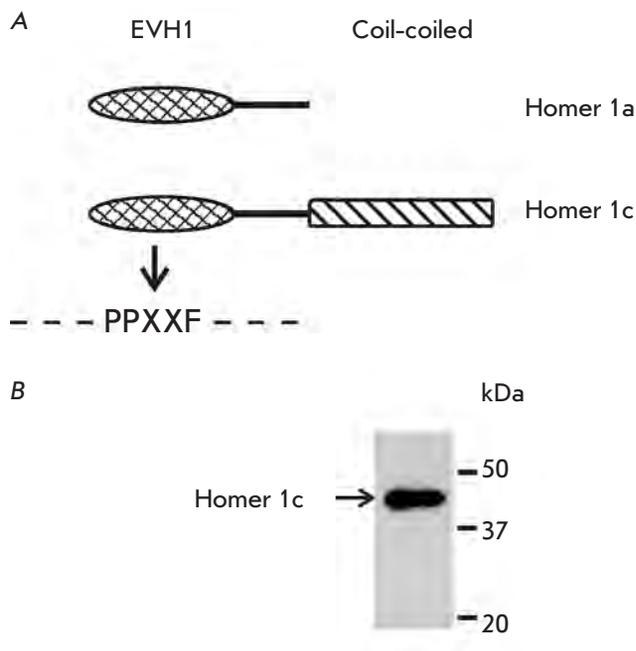


Fig. 1. Homer 1bc long isoform is expressed in the human epidermal carcinoma cell line A431. A – Domain organization of Homer 1 proteins. B – The immunoblotting of whole cell lysates from A431 cells was performed with polyclonal anti-Homer 1bc antibodies. The positions of molecular mass markers are shown on the right.

action was carried out in a PBS buffer containing 1% Triton X-100. In some cases, the incubation was conducted in the presence of IP_3 ; after that, sepharose was rinsed three times with 1 ml PBS containing 1% Triton X-100. The presence of IP_3R in the samples was confirmed by immunoblotting with polyclone antibodies against IP_3R .

RESULTS AND DISCUSSION

Uncoupling of Homer-Target Protein Interactions Caused by PPKKFR Interaction Activates Ca^{2+} Influx in A431 Cells

The results of immunoblotting show that A431 cells express long isoforms of Homer proteins (Fig. 1B). The EVH1 domain in Homer proteins recognizes the PPXXFR motif in target proteins, where X stands for any amino acid (Fig. 1A) [2]. In order to reveal the roles which Homer proteins play in regulating receptor-operated or store-operated calcium channels, we used a synthetic peptide PPKKFR, since it has been shown to promote the dissociation of the Homer-mGluR complex [2].

In control experiments we used peptide PPXXRR in which the substitution of phenylalanine with arginine

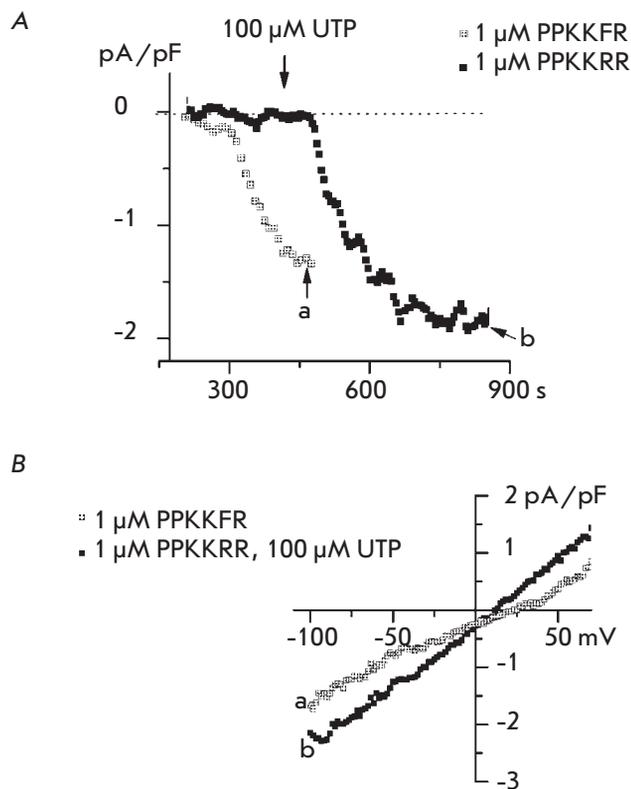


Fig. 2. Whole-cell recordings of currents induced by PPKKFR or PPKKRR peptides. A The development in time of a current from a cell dialyzed with 1 μ M PPKKFR peptide at -80 mV potential (gray squares) and with 1 μ M PPKKRR (black squares). The arrow at the top indicates when 100 μ M UTP was added. Arrows a and b indicate the maximum amplitudes of the current. B Average current–voltage relationships of currents induced by dialysis with PPKKFR peptide or by the addition of UTP. The current–voltage relationships were measured when the inward currents reached the maximum (indicated by arrows a and b in panel A).

cancelled the peptide's interaction with the EVH1 domain. In order to elucidate the role that Homer proteins play in the regulation of the Ca^{2+} influx in A431 cells, we applied a patch clamp technique in the whole-cell configuration. The net Ca^{2+} influx, which was measured under conditions of intracellular dialysis against PPKKFR and PPKKRR-containing solutions, was compared with the net store-operated Ca^{2+} influx, which was induced by uridine triphosphate.

Intracellular dialysis with a solution containing 1 μ M PPKKFR led to the selective activation of a Ca^{2+} current with an average maximal amplitude of 1.3 ± 0.1 pA/pF ($n = 5$) (Fig. 2A). The absence of PPKKRR in the pipette solution led to no current ($n = 10$), but the

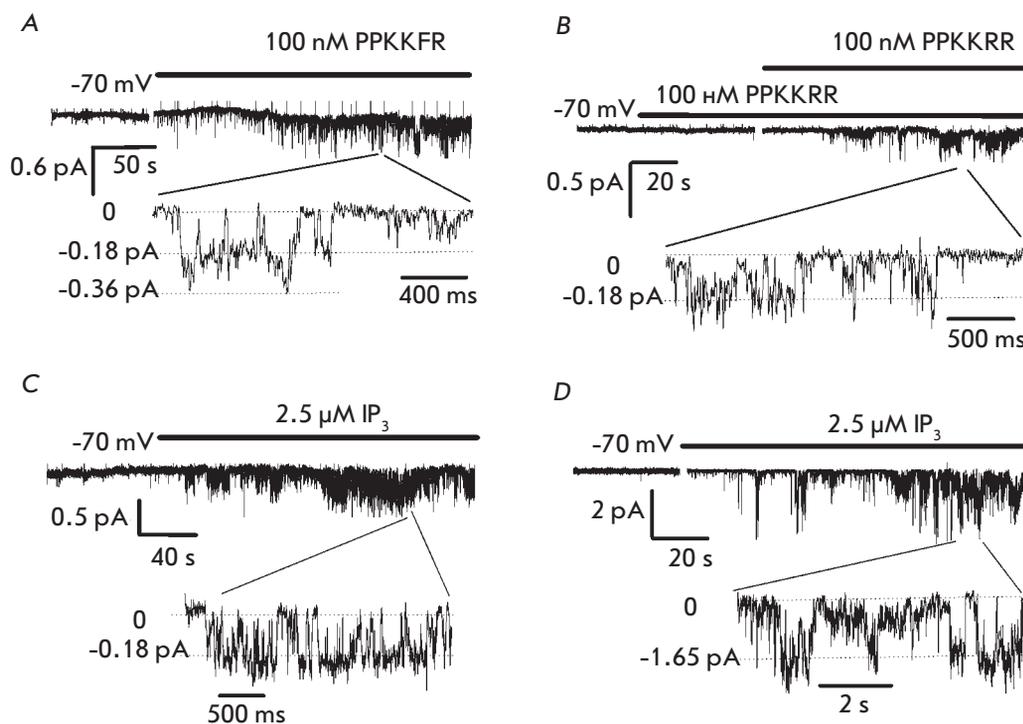


Fig. 3. Homer dissociation led to I_{\min} activity, but it didn't lead to I_{\max} activity. A – The application of 100 nM PPKKFR to the cytosolic surface of an inside-out patch held at a -70 mV membrane potential activated I_{\min} . The fragment of the current record is shown at the bottom on the expanded time scale. B – The application of 100 nM of the control peptide PPKRR did not activate I_{\min} , whereas the subsequent addition of 100 nM PPKKFR to the same patch induced I_{\min} activity. C, D – The application of $2.5 \mu\text{M}$ IP₃ induced (C) I_{\min} and (D) I_{\max} activity. Examples of single-channel recordings from the isolated membrane at -70 mV potential.

addition of $100 \mu\text{M}$ UTP caused the current to appear ($n = 5$) (Fig. 2A). The amplitude of the UTP-induced current was equal to approximately 1.8 ± 0.3 pA/pF and, therefore, was bigger than that of the PPKKFR-induced current (Fig. 2B). The reversal potential of PPKKFR-induced currents was more positive than that of UTP-induced currents. Thus, PPKKFR-activated channels are more selective for Ca^{2+} than UTP-activated channels.

Our data are in agreement with previously published works by other authors who suggested an interconnection between Homer proteins and store-operated Ca^{2+} influx in nonelectroexcitable cells [3, 4].

In previous works we described several types of UTP-sensitive Ca^{2+} channels with different reversal potentials of the currents [11]. We assumed that the difference in amplitudes of currents and channel selectivity in the response of PPKKFR and UTP has the following explanation: UTP treatment activates various types of channels, whereas PPKKFR activates particular channels.

I_{\min} (but Not I_{\max}) Channels are Sensitive to PPKKFR-Induced Homer Dissociation from Protein Targets

In order to investigate which Ca^{2+} channels from A431 cells account for the currents induced by the dissociation of Homer proteins from their targets, we performed patch clamp experiments in an inside-out configuration. These experiments show that, in A431 cells,

there are two types of store-operated channels (I_{\min} and I_{\max}) [11]. These channels have different electrophysiological characteristics, making it possible to identify them based on the registration of currents through them. I_{\min} channels have low conductivity (1.2 pS) and high selectivity for bivalent cations. In comparison with I_{\min} channels, I_{\max} have higher conductivity (18 pS) but lower selectivity.

In an inside-out configuration, the addition of a 100-nM PPKKFR peptide from the cytoplasmic side of the membrane fragment led to the activation of the influx current (Fig. 3A). It turned out that the basic electrophysiological properties of the activated channels (kinetics, conductivity, and reversal potential) coincide with those of previously described I_{\min} channels [11–18]. This observation leads to the conclusion that uncoupling Homer proteins and their protein targets activates I_{\min} channels in A431 cells, which is in agreement with the results obtained in our laboratory on HEK293 cells [12]. I_{\min} channels were activated by PPKKFR peptide in 43% of cases ($n = 60$) (Fig. 3A). Control peptide PPKRR (which should not uncouple interactions between Homer and target proteins) did not activate channels in any of the 36 experiments, though the subsequent addition of PPKKFR led to the activation of I_{\min} channels in 42% of experiments ($n = 26$) (Fig. 3B). PPKKFR did not activate I_{\max} channels, which are another group of store-operated channels in A431 cells ($n = 60$).

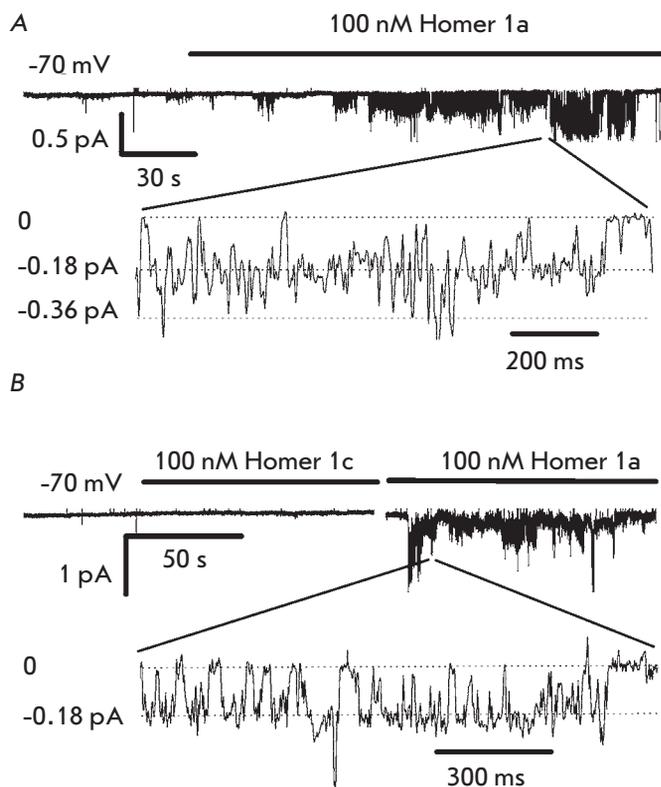


Fig. 4. Homer 1a protein activates I_{\min} channels. **A** The application of 100 nM Homer 1a to the cytosolic surface of an inside-out patch held at a -70 mV membrane potential activated I_{\min} channels. The fragment of the current record is shown at the bottom on the expanded time scale. **B** Homer 1c (100 nM) did not activate store-operated channels, whereas the subsequent application of Homer 1a was effective in the same experiment. The fragment of the current record is shown at the bottom on the expanded time scale.

The addition of $2.5 \mu\text{M}$ inositol-1,4,5-trisphosphate from the cytosolic side of the plasma membrane in inside-out experiments activated I_{\min} channels in 32% and I_{\max} channels in 8% of experiments ($n = 80$) (Figs. 3C, 3D), which is in agreement with our previous data [11]. At potential values of -70 mV, the amplitudes of I_{\min} currents were 0.18 pA, while in case of I_{\max} currents they were equal to 1.7 pA (Figs. 3A–3D).

These data allowed us to conclude that PPKKFR-induced uncoupling of interactions between Homer proteins and their protein targets leads to the activation of I_{\min} channels, while I_{\max} channels in A431 cells are insensitive to this peptide. It is unknown whether two other store-operated channels of A431 cells – I_{NS} and I_{CRAC} – are regulated by Homer proteins, since these channels are not present (or are indistinguishable) in

an inside-out configuration.

Data on the activity of single channels are in good agreement with the results of the whole-cell experiments. UTP activates all kinds of store-operated channels in A431 cells, and PPKKFR peptide does not activate at least I_{\max} channels, hence the peptide causes a partial activation of calcium channels in the whole-cell configuration. This fact explains the difference in values of the registered net currents. Since I_{\max} channels are less selective than I_{\min} channels, a UTP-induced net current exhibits lower selectivity.

I_{\min} Channels Are Activated by Homer 1a, but Not Homer 1c

Homer proteins fall into two different groups [5]. Long isoforms (e.g., Homer 1c) have a coiled-coil domain on their C-termini and so are able to form homo-oligomers. The lack of a coiled-coil domain in short isoforms (e.g., Homer 1a) prevents the formation of oligomeric complexes (Fig. 1A). In order to investigate the effects that long and short isoforms have on the activity of store-operated calcium channels in A431, we used recombinant Homer proteins produced in *E.coli* transformed with a GST-Homer plasmid. The functional activity of Homer 1a- and Homer 1c-purified proteins was judged by their ability to bind $\text{IP}_3\text{R1}$ in a pull-down assay.

The monomer isoform Homer 1c at a concentration of 100 nM activated Ca^{2+} channels in 30% of inside-out experiments ($n = 101$) (Fig. 4A). The current-voltage characteristics of Homer 1a-activated channels coincided with those of I_{\min} channels activated by UTP, store depletion, or IP_3 in A431 and HEK293 cells [11, 13–18]. The conductivity of Homer 1a-activated channels was equal to 1.3 pS. The long isoform Homer 1c at a concentration of 100 nM did not lead to channel activation ($n = 58$), while the subsequent addition of Homer 1a activated I_{\min} channels in 27% of cases ($n = 44$) (Fig. 4B). Neither Homer 1a nor Homer 1c activated I_{\max} channels.

The present data indicate that various Homer isoforms differ in their action on I_{\min} channels in A431 cells: the monomeric Homer 1a protein causes activation, while Homer 1c does not. Various Homer isoforms similarly affect mGluR [8] and TRPC channels [3, 4], but not RyR channels [19, 20], which are activated by both long and short isoforms of Homer proteins, long isoforms being stronger activators of type-I RyR than short ones.

One explanation, which was suggested in the case of TRPC channels, is probably applicable to I_{\min} channels as well. Since Homer 1c proteins can form oligomers and Homer 1a cannot because they are devoid of the coiled-coil domain, it was hypothesized that the oligomeric complexes block the channels, while uncoupling

with Homer oligomers activates them [3, 4]. Our experiments with peptides and recombinant proteins show that it is not Homer 1a itself but the uncoupling of the interaction of Homer oligomers and their protein targets that is necessary for the activation of I_{\min} -channels. We can suggest that the dissociation of Homer oligomeric complexes from their targets leads to the alteration of interactions of the Ca^{2+} channel and other proteins, including IP_3R . This uncoupling can be caused by the PPKKFR peptide, short isoform Homer 1a, and IP_3 (see below). It is known that Homer 1 proteins do not affect the phosphoinositide metabolism. In particular, they do not cause the elevation of the IP_3 concentration and cannot promote the release of Ca^{2+} from IP_3 -sensitive Ca^{2+} stores [21]. Therefore, we can conclude that Homer activates store-operated Ca^{2+} channels by means of direct interaction with the channel and not through releasing Ca^{2+} from the intracellular stores. Thus, despite the fact that I_{\min} channels are store-operated in A431 cells, they apparently can also exploit a store-independent mechanism, which happens if complexes of Homer and its protein targets are dissociated.

The Effects That IP_3 and the Uncoupling of Native Homer Proteins with Their Protein Targets have on I_{\min} Channels Are Nonadditive

Since I_{\min} channels are regulated by both IP_3 and Homer proteins, the question arises as to whether their action is additive. Adding $2.5 \mu\text{M}$ IP_3 to membrane fragments with PPKKFR-induced activity most of the time did not lead to any further increase in channel activity (in 10 out of 13 experiments) (Fig. 5A). In similar experiments with recombinant Homer 1a, the protein application of IP_3 also did not alter the I_{\min} channel activity (Fig. 5B). However, in several experiments, the addition of IP_3 activated I_{\max} channels (Fig. 5B). When PPKKFR or the recombinant protein Homer 1a did not activate I_{\min} channels, IP_3 also did not show any activation effect either. Therefore, we had demonstrated that the effects of IP_3 and the uncoupling of native Homer proteins with their protein targets on I_{\min} channels are nonadditive. These data led us to suggest that the regulation of I_{\min} channels by IP_3 and Homer proteins exploits the same signal pathway.

IP_3 Impairs Homer- $\text{IP}_3\text{R1}$ Interaction

A431 cells basically express type-1 IP_3R [Glushankova, unpublished data]. It has been reported that, in these cells, I_{\min} channels are apparently regulated by means of conformational coupling with type-1 IP_3R [15, 17]. Here we show that Homer proteins participate in the regulation of I_{\min} channels. X-ray diffraction data reveal that, in the IP_3R molecule, the Homer-recognizing motif and the IP_3 -binding domain are positioned closely

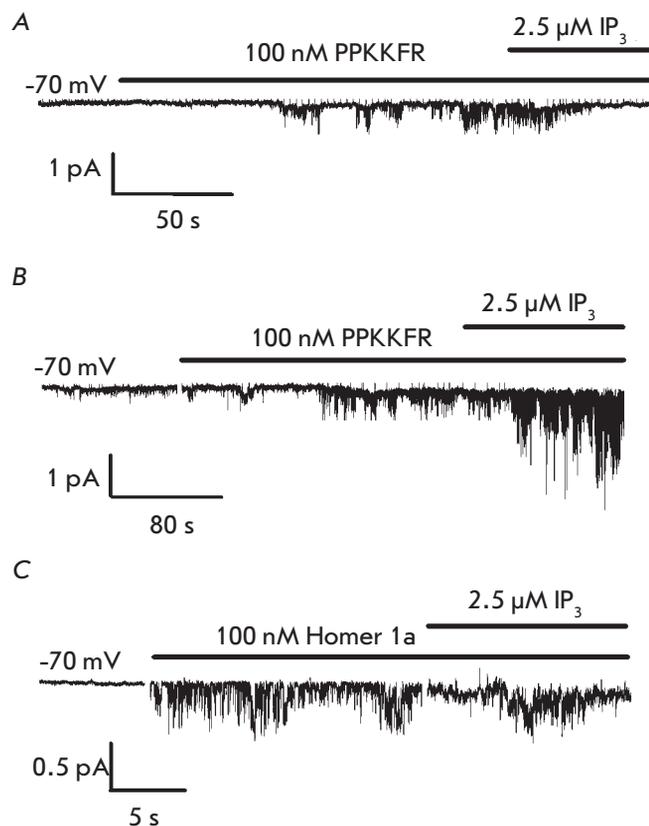


Fig. 5. Detachment of native Homer proteins and the IP_3 effect are not additive for I_{\min} -channel activation. A – The subsequent application of $2.5 \mu\text{M}$ IP_3 to patches did not increase I_{\min} -channel activity induced by 100 nM PPKKFR peptide. B – The application of 100 nM PPKKFR to the cytosolic surface activated I_{\min} channels. The subsequent application of $2.5 \mu\text{M}$ IP_3 activated I_{\max} . C – I_{\min} -channel activity, induced by 100 nM Homer 1a, did not change after the addition of $2.5 \mu\text{M}$ IP_3 .

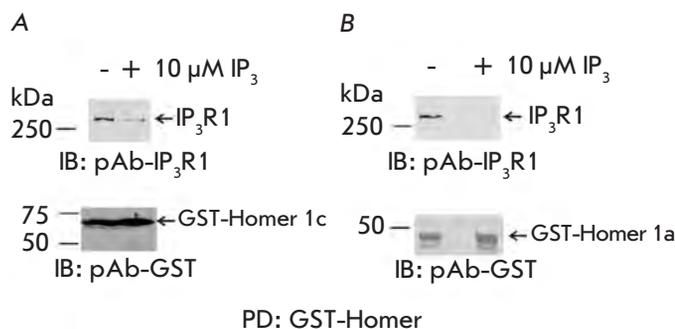


Fig. 6. IP_3 decreased the Homer interaction with $\text{IP}_3\text{R1}$. A, B – Results of pull-down experiments. IP_3 was incubated with GST-Homer 1c (A) or GST-Homer 1a (B) in the presence or absence of $10 \mu\text{M}$ IP_3 .

[22]. It was also shown that IP_3 causes the dissociation of Homer from its complex with type-3 IP_3R [4]. Our pull-down experiments show that IP_3 can impair interactions between type-1 IP_3R and Homer 1a or Homer 1c proteins (Figs. 6A, 6B). These results are in good agreement with the data obtained in inside-out experiments, which show that the effects of IP_3 , Homer 1a, and PP-KKFR peptides are nonadditive.

Thus, the Homer-mediated regulation of native store-operated channels in A431 cells has been investigated in our work for the first time. These results

deepen our knowledge of the organization of the components that mediate store-operated influx. ●

This work was supported by the Programs of the Education and Science Ministry (Government Contracts 02.740.11.5007 and P332), the Program “Molecular and Cell Biology” of the Russian Academy of Sciences, the Russian Foundation for Basic Research (grants № 09-04-12035, №10-04-01002, and №10-04-00956), and by the grant “Leading Scientific Schools” (NSh-3796.2010.4).

REFERENCES

1. Parekh A.B., Putney J.W., Jr. // *Physiol. Rev.* 2005. V. 85. № 2. P. 757–810.
2. Tu J.C., Xiao B., Yuan J. P., et al. // *Neuron.* 1998. V. 21. P. 717–726.
3. Yuan J.P., Kiselyov K., Shin D. M., et al. // *Cell.* 2003. V. 114. P. 777–789.
4. Kim J.Y., Zeng W., Kiselyov K., et al. // *J. Biol. Chem.* 2006. V. 281. № 43. P. 32540–32549.
5. Worley P.F., Zeng W., Huang G., et al. // *Cell Calcium.* 2007. V. 42. № 4–5. P. 363–371.
6. Hayashi M.K., Tang C., Verpelli C., et al. // *Cell.* 2009. V. 137. № 1. P. 159–171.
7. Hayashi M.K., Ames H.M., Hayashi Y. // *J. Neurosci.* 2006. V. 26. № 33. P. 8492–8501.
8. Ango F., Prézeau L., Muller T., et al. // *Nature.* 2001. V. 411. № 6840. P. 962–965.
9. Feng W., Tu J., Yang T., et al. // *J. Biol. Chem.* 2002. V. 277. № 47. P. 44722–44730.
10. Kaznacheyeva E., Lupu V.D., Bezprozvanny I. // *J. Gen. Physiol.* 1998. V. 111. P. 847–856.
11. Kaznacheyeva E., Glushankova L., Bugaj V., et al. // *J. Biol. Chem.* 2007. V. 282. № 32. P. 23655–23662.
12. Nikolaev A.V., Skopin A.Yu., Kaznacheyeva E.V. // *Biologicheskie Membrany (Biological Membranes).* 2004. V. 21. № 6. P. 451–457.
13. Kiselyov K.I., Mamin A.G., Semyonova S.B., Mozhayeva G.N. // *FEBS Lett.* 1997. V. 407. P. 309–312.
14. Kiselyov K.I., Semyonova S.B., Mamin A.G., Mozhayeva G.N. // *Pflügers Arch.* 1999. V. 437. P. 305–314.
15. Zubov A.I., Kaznacheyeva E.V., Nikolaev A.V., et al. // *J. Biol. Chem.* 1999. V. 274. P. 25983–25985.
16. Kaznacheyeva E., Zubov A., Gusev K., et al. // *Proc. Natl. Acad. Sci. USA.* 2001. V. 98. P. 148–153.
17. Gusev K., Glouchankova L., Zubov A., et al. // *J. Gen. Physiol.* 2003. V. 122. P. 81–94.
18. Bugaj V., Alexeenko V., Zubov A., et al. // *J. Biol. Chem.* 2005. V. 280. P. 16790–16797.
19. Feng W., Tu J., Pouliquin P., et al. // *Cell Calcium.* 2008. V. 43. № 3. P. 307–314.
20. Pouliquin P., Pace S.M., Dulhunty A.F. // *Pflügers Arch.* 2009. V. 458. P. 723–732.
21. Shin D.M., Dehoff M.D., Luo X., et al. // *J. Cell Biol.* 2003. V. 162. P. 293–303.
22. Bosanac I., Yamazaki H., Matsu-Ura T., et al. // *Mol. Cell.* 2005. V. 17. P. 193–203.

Oligonucleotide Microarray for the Identification of Carbapenemase Genes of Molecular Classes A, B, and D

M. M. Ulyashova¹, Yu. I. Khalilova¹, M. Yu. Rubtsova^{1,2*}, M. V. Edelstein³, I. A. Alexandrova⁴, A. M. Egorov¹

¹ Chemistry Faculty, Lomonosov Moscow State University

² ZAO 'NPP IMMUNOTEK', (NPP IMMUNOTECH, JSC)

³ Institute of Antimicrobial Chemotherapy, Smolensk State Medical Academy

⁴ Burdenko Institute of Neurosurgery, Russian Academy of Medical Sciences

* E-mail: mrubtsova@gmail.com

Received 23.08.2010

ABSTRACT This work is a report on the development of a method of hybridization analysis on DNA microarrays for the simultaneous identification and typing of carbapenemase-encoding genes. These enzymes are produced by the microorganisms that are responsible for causing infectious diseases. The method involves several steps, including DNA extraction from clinical samples and amplification of carbapenemase genes by multiplex PCR with simultaneous labelling by biotin. Following that, hybridization of the labeled PCR products with oligonucleotide probes immobilized on the surface of a nitrocellulose-based DNA microarray occurs. The biotin molecules attached to the DNA duplexes are detected by using conjugates of streptavidin-horseradish peroxidase, which is then quantified by colorimetric detection of the enzyme. We have designed the required oligonucleotide probes and optimized the conditions of the membrane microarray-based hybridization analysis. Our method allows to identify 7 types of carbapenemase genes belonging to the molecular classes A, B, and D, and it also allows additional typing into genetic subgroups. The microarrays have been tested with the control strains producing the carbapenemase genes which have been characterized by sequencing. The developed method of hybridization analysis was employed to investigate clinical strains of *Pseudomonas* spp. and *Acinetobacter* spp., which produce carbapenemases of different classes based on phenotypic testing. All strains of *Acinetobacter baumannii* resistant to carbapenems were producers of two carbapenemase OXA-type genes (OXA-51, in combination with OXA-23 (1 strain), OXA-40 (5 strains), or OXA-58 (4 strains)). The metallo- β -lactamase VIM-2 type gene was detected in all *Pseudomonas aeruginosa* strains resistant to carbapenems. Testing of carbapenem-sensitive strains did not detect any carbapenemase genes. The microarray method for the identification of carbapenemase genes is very accurate and highly productive. It can be employed in clinical microbiological laboratories for the identification and study of carbapenemase epidemiology.

KEYWORDS DNA microarrays, horseradish peroxidase, colorimetric detection, antibiotic resistance, carbapenemases

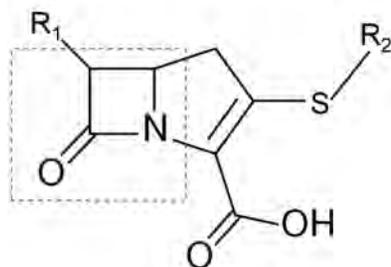
INTRODUCTION

During the last 20 years β -lactam antibiotics have been one of the drugs of choice for the treatment of a wide range of severe infectious diseases caused by gram-negative microorganisms. Among these antibiotics, carbapenems are one of the most successful drug groups (chemical structure shown in Fig. 1). This group of antibiotics is characterized by a wide spectrum of sensitive organisms, low toxicity, and good pharmacokinetics [1]. However, their efficiency has recently begun to show limitations with the emergence of drug-resistant strains. Carbapenem-resistant strains are most often found among nosocomial (in-hospital) infections belonging to the *Pseudomonas* spp. and *Acinetobacter* spp. genera. Healthcare facilities in the Russian Federation are, for example, noting a considerable increase in the number of pathogens resistant to carbapenem treat-

ment. 38% of the observed *Pseudomonas aeruginosa* strains are now resistant to these drugs [2].

The emergence of β -lactam resistance in gram-negative bacteria can happen through several mechanisms, which include alteration of membrane permeability due to defective porine channels [3, 4], or activation of efflux systems [5]. However, the most clinically and epidemiologically important mechanism is the production of bacterial enzymes β -lactamases which hydrolyze the β -lactam ring of the antibiotic drug - β -lactamases [6, 7]. The β -lactamases are currently divided into 4 molecular classes - A, B, C, and D, based on their primary structure. The A-, C- and D-class enzymes are serine-type hydrolases, while the B-class enzymes are metallo-hydrolases, which bear one or two zinc atoms in their active site [8]. Several molecular classes of β -lactamase possess carbapenemase activity: however, the most of-

Fig. 1. Chemical structure of carbapenemes. β -Lactam ring in dashed lines.



ten found and clinically important carbapenemases are KPC-type A-class enzymes [9], 5 groups of metallo- β -lactamases (VIM, IMP, SPM, GIM, SIM) [10], and a number of OXA-type D-class enzymes (subgroups OXA-23, OXA-40, OXA-51, OXA-58) [11].

Of all the numerous β -lactamases, carbapenemases are the most dangerous; they display high catalytic activity and wide substrate specificity, which includes practically every class of β -lactam antibiotics. Since carbapenemase-encoding genes are located on a plasmid, they can spread among pathogenic microorganisms at a rapid pace. Because of the variety of carbapenemases and the danger of their spreading, there is a need for robust methods for the detection of enzyme production, which can then be used for choosing the best suited treatment and for epidemiological control over the spreading of specific drug-resistant types. This is currently accomplished by using microbiological tests [6, 12, 13]. However, these tests take time and are ineffective for the identification of carbapenemase types. Identification of OXA-type enzymes by phenotype-based tests is virtually impossible [14].

Several PCR-based methods have been suggested for the identification of metallo- β -lactamase genes of the most spread VIM and IMP types [15, 16], and the main OXA-type carbapenemase subgroups [17]. A method for the identification of 5 groups of metallo- β -lactamases, involving multiplex real-time PCR with subsequent analysis of the melting curves of the obtained amplicons, has recently been developed [18]. However, the multiplex capacity of PCR is usually limited, which makes simultaneous detection of a large number of genes impossible.

Hybridization analysis based on a microarray-technology is a promising method of identification that yields quick results of high informative value. This technique has considerable advantages over traditional methods, since it allows a multi-parametric analysis and also uses a miniscule sample, which reduces cost and the time needed to obtain results [19, 20].

The goal of this work was to develop a method for the identification of A-, B-, and D-class carbapenemase genes involving hybridization analysis on membrane-based DNA microarrays which could be visualized by colorimetric detection.

EXPERIMENTAL PROCEDURES

The collection of primers for the amplification of carbapenemase genes and the amino-modified oligonucleotide probes were synthesized by Synthol (Moscow, Russia). Samples of the bacterial DNA extracted from control strains of *A. baumannii*, *Ps. aeruginosa*, *Escherichia coli*, and *Klebsiella pneumonia* were provided by the Institute of Antimicrobial Chemotherapy of the Smolensk State Medical Academy. These samples produced carbapenemases VIM-1, VIM-2, VIM-4, VIM-7, IMP-1, IMP-2, SPM-1, OXA-23, OXA-40, OXA-51, OXA-58, and KPC-3. Cell cultures of microorganisms from the *Enterobacteriaceae* family and *A. baumannii* and *Ps. aeruginosa* strains, either sensitive or resistant to carbapenems, according to phenotypic tests performed on a VITEK automatic analyzer (BioMerieux, France), were provided by the Burdenko Institute of Neurosurgery.

Bacterial DNA extraction

Extraction of bacterial DNA from a cell suspension with no less than 10^5 CFU/ml was performed using temperature lysis in a buffer. 500 μ l of suspension was placed in centrifuge tubes, and the cells were pelleted by centrifugation at 10,000 g for 1 minute. After removing the supernatant, 100 μ l of buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) was added to the pellet, which was then re-suspended with a shaker. The tubes were then incubated on a solid-medium thermostat for 20 minutes at a temperature of 99°C. After heating, the samples were centrifuged for 1 minute at 10,000 g . A PCR sample used 1 μ l of the obtained supernatant.

Amplification of carbapenemase genes using multiplex PCR with simultaneous biotin labeling

Amplification of carbapenemase (A-, B-, and D-class) gene-fragments, with simultaneous biotin labeling, was performed in two multiplex PCR reactions (one reaction amplified the genes of all the metallo- β -lactamases and the other amplified the OXA- and KPC-type β -lactamase genes). Each multiplex PCR sample was 25 μ l in volume and contained the following: 10 mM Tris-HCl-buffer with 2.5 mM of magnesium acetate, 50 mM KCl pH 8.3, 2.5 units of *Taq*-DNA-polymerase, 100 μ M dATP, dGTP, dCTP, 60 μ M dTTP, 40 μ M dUTP-11-biotin (Fermentas, Germany), 0.4 μ M each of the direct and reverse primer for each group of carbapenemase, and 1 μ l of the template DNA solution. Amplification was performed in a Mastercycler gradient amplifier (Eppendorf, Germany) according to the following protocol: initial denaturation at 94°C (2 min), 25 cycles of amplification (20 sec – denaturation at 94°C, 30 sec – annealing of the primers at 65°C, 1 minute – elongation at 72°C), and a final elongation step at 72°C

(6 min). Horizontal electrophoresis of the PCR products was performed in a 1% agarose gel with TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5) with etidium bromide added to a final concentration of 1.6 µg/ml. Visualization was performed on a UV-transilluminator at a wavelength of 260 nm.

Fragmentation of the PCR-products

The DNA was fragmented at room temperature for 5 minutes. Amplified DNA was diluted to a concentration of 30 ng/microliter with the reaction buffer (40 mM Tris-HCl, 10 mM MgSO₄, 1 mM CaCl₂, pH 8.0), and the mixture was then supplemented with DNAase I (Promega, Germany). The reaction was stopped by the addition of 3 mM EDTA and 10-minute incubation at 65°C.

Immobilization of the oligonucleotide probes on a membrane-based DNA-microarray

BioTrace NT nitrocellulose (Pall Corporation, USA) was used as a support for DNA-microarray. Modification of the membranes was performed according to [21], using 1-ethyl-3-(3-dimethylaminopropyl)carboimide (Sigma, USA). The oligonucleotides were diluted in a buffer (160 mM Na₂SO₄, 130 mM Na₂HPO₄) to a final concentration of 20 µM and then applied onto the membranes by an XactII™ Arrayer robot (LabNEXT Inc., USA) using 300 µm pins. After the procedure, the membranes were incubated at 60°C for 30 minutes.

Hybridization on the DNA microarray

Prior to hybridization, the microarrays were washed with PBST buffer (0.01 M K₂HPO₄, 0.15 M NaCl, 0.05% Tween-20, pH 7.0) twice, 10 minutes each time at room temperature, and then blocked in a solution of 1% bovine serum albumine (BSA) and 1% casein (Sigma, USA) in PBS buffer (0.01 M K₂HPO₄, 0.15 M NaCl, pH 7.0) at 37°C for 30 minutes. 500 ng of fragmented and labeled DNA was then diluted in the hybridization buffer - 2x SSPE (0.3 M NaCl, 0.02 M NaH₂PO₄, 2 mM EDTA, pH 7.4), which also included 1.6 pmol/ml of the control biotin-labeled oligonucleotide (positive hybridization control). The microarray was then placed into the hybridization mixture (300 µl per 1 array) and incubated at 45°C for 1 hour in a Thermomixer comfort apparatus (Eppendorf, Germany). After hybridization, the membranes were washed with PBST twice for 15 minutes at room temperature.

Detection and hybridization data analysis

The microarrays were incubated in a solution of streptavidine-peroxidase conjugate (Imtek, Russia) (0.2 µg/ml) in PBST for 30 minutes at 37°C. Then they were washed in PBST for 10 minutes and placed into a substrate solution containing 3,3',5,5'-tetramethylben-

zydene (TMB), H₂O₂ (NVO Immunotech, Russia) and sodium dextran sulphate (M_r = 8000, Pharmacia, Sweden) (final concentration - 0.5% (by mass)) for 10 minutes, after which the arrays were washed in distilled water and air-dried. The membrane microarrays were scanned on a Perfection V750 Pro (Epson, Germany) scanner at a resolution of 4,800 dpi. The obtained images (in TIFF format) were analyzed using Scan Array Express (PerkinElmer, version 3.0, Germany) software, and the intensity values of the analytic signals at various spots of the microarray were determined. The absolute values of the signals were then recalculated into relative signals, using the mean intensity of the positive hybridization control used in each array.

RESULTS AND DISCUSSION

Molecular design of the oligonucleotide probes

The Genbank database currently has information on 10 KPC-type enzymes, 52 metallo-β-lactamases (23 from the IMP group, 26 from the VIM group and one member of each of the following groups - SPM, SIM and GIM), and also 70 carbapenemases from the OXA group. Alignment of the amino acid and encoding sequences of these enzymes shows that only enzymes from the KPC group display a high degree of similarity within their group (differ by 1-2 amino acid substitutions), while the numerous members of the IMP, VIM, and OXA groups differ considerably from their group members. Because of this, each group was split into separate subgroups, which included enzymes whose genes were highly similar. Thus, the VIM group was divided into 3 subgroups (VIM-1, VIM-2, and VIM-7); the IMP group, into 6 subgroups (IMP-1, IMP-2, IMP-5, IMP-11, IMP-12, and IMP-14); while carbapenemases from the OXA groups were divided into 4 subgroups (OXA-23, OXA-40, OXA-51 and OXA-58). Group and subgroup classification of carbapenemases based on their amino acid sequence alignments and their β-lactamase molecular classification are shown in Fig. 2.

One of the main stages of DNA microarray development was the design of the oligonucleotide probe sequences required to detect various groups of carbapenemase genes. The selection of an oligonucleotide probe for the identification of a group of genes is based on the alignment of the coding sequences of all the carbapenemase genes in this group. What is needed is a sufficiently long fragment of the gene that is conserved in all of the members of this group and is no less than 18 nucleotides long. These regions were then analyzed in terms of melting temperature, G/C content and secondary structure formation. In order to perform microarray-based hybridization analysis, we selected the oligonucleotides that were unlikely to form secondary structures and

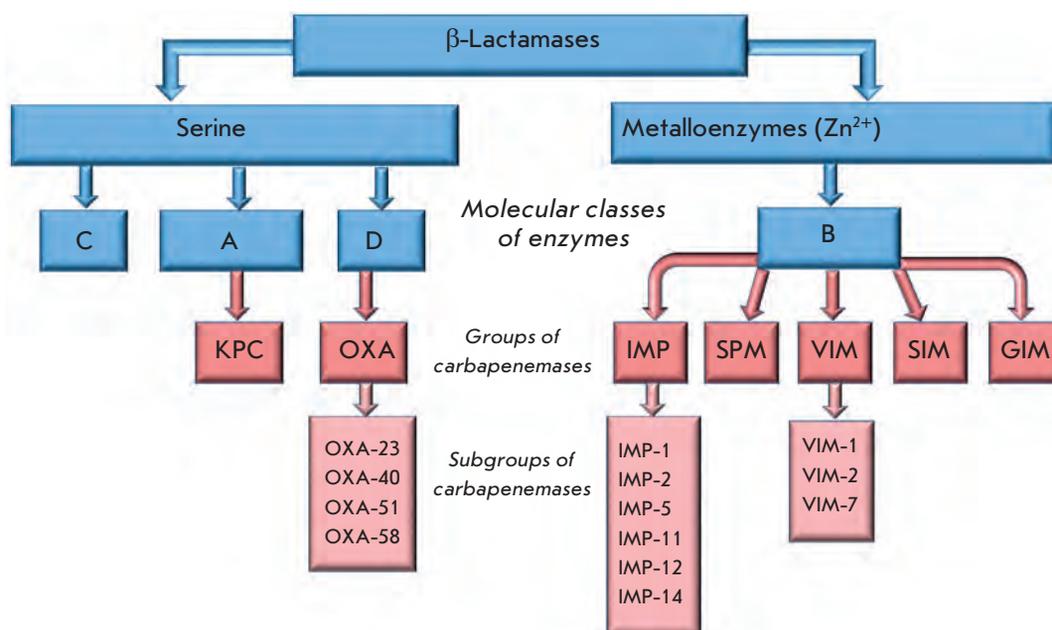


Fig. 2. Classification of carbapenemases into groups and subgroups based on the alignment of amino acid sequences and their correspondence to molecular classes of β -lactamases.

whose melting temperature differed by no more than 10°C . Based on the selected sequences, we synthesized two oligonucleotide probes which were complementary to the direct and reverse strands of the gene. For additional typing of IMP, VIM, and OXA carbapenemase gene-subgroups, we chose regions with high similarity which bore no mutations inside a given subgroup and had low similarity with genes from other subgroups. In order to increase the specificity of the analytical procedure, we chose two oligonucleotide probe variants which corresponded to different regions of the gene. In this case, we synthesized the probes which were complementary to the reverse strand of the gene.

The sequence of the chosen oligonucleotide probes and their characteristics are presented in Table 1. The length of the probes varied from 18 to 27 nucleotides, the G/C-content was 30–60%, and the melting temperature was $63\text{--}72^{\circ}\text{C}$. Each probe was supplemented by additional spacers at the 5'-terminus, which helped to distance the probe from the array surface, thus removing steric barriers for hybridization and increasing the availability of the probe for the DNA-target. Optimization procedures showed that the best spacer contained 13 thymidine residues. The additional thymidines did not have any significant effects on the intensity of the hybridization signals, or on the specificity of hybridization (data not shown).

Amplification of carbapenemase genes of various molecular classes with simultaneous labeling

Amplification of A-, B-, and D-class carbapenemase genes utilized multiplex PCR with simultaneous biotin labeling. Biotin-labeled deoxyribouridine triphos-

phate (dUTP) was used as a labeling reagent and was incorporated into the DNA, along with unlabeled deoxyribothymidine triphosphate (dTTP). The templates for the PCR reaction were bacterial DNA samples extracted from the control strains of microorganisms producing β -lactamases VIM-1, VIM-2, VIM-7, IMP-1, IMP-2, SPM-1, OXA-23, OXA-40, OXA-51, OXA-58, and KPC-3.

Design of primers for the amplification of various carbapenemase genes was based on an alignment of the coding regions of these genes. Primers for amplification of the full-size carbapenemase genes from the KPC group and metallo- β -lactamase groups SPM, SIM, and GIM were chosen from regions which were conserved in this group, namely the gene termini. We could not find conserved regions longer than 20 bp for metallo- β -lactamases from the IMP and VIM groups, nor could we find any for OXA-type carbapenemases, since these groups showed a low degree of in-group similarity, which is why separate primers had to be selected for each subgroup.

The length of the primers (20 – 28 nucleotides) was chosen so as to push their melting temperature to $62\text{--}68^{\circ}\text{C}$, which would allow simultaneous amplification of all the types of genes simultaneously and with equal efficiency. Primer selection also factored in G/C-content, and we chose structures that had a G/C content of 30 – 60%. We also estimated the possibility of primer-dimer and secondary structure formation and favored the sequences which were least likely to do so. As a result, each group was fitted with several direct and reverse primers with various parameters. Various combinations of these primers were tested in PCR reactions in order

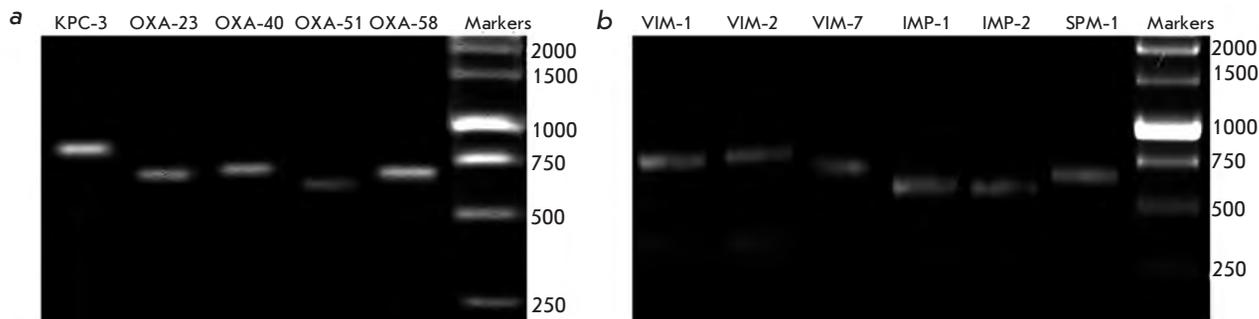


Fig. 3. Electrophoretic separation of multiplex PCR products after amplification of OXA-, KPC-type carbapenemases (a) and metallo- β -lactamases (b).

to test the specificity of gene amplification. The primers which showed the highest specificity with a good output were used for the multiplex procedure. The primer sequences are presented in Table 2.

In order to determine the optimal primer-annealing temperature for the multiplex amplification of all the carbapenemase gene types, we first calculated the optimal annealing temperature for each pair of primers in a specific reaction and then selected the lowest temperature value for the multiplex PCR reaction. The range of the studied primer T_a 's was 52 – 68°C. The optimal T_a turned out to be 60°C. Reactions at this temperature resulted in the efficient synthesis of specific PCR-products.

After determining the optimal carbapenemase gene amplification conditions, we evaluated the possibility of performing multiplex PCR with 16 pairs of primers in order to amplify the genes of all of the studied groups simultaneously. However, under these conditions the specific product for most of the carbapenemase groups was either absent or the yield was very low, which made further microarray hybridization analysis impossible. Because of this, we decided to amplify the carbapenemase genes using a two step multiplex PCR procedure: the first used a mixture of primers specific to metallo- β -lactamases (11 pairs of primers), and the second used a mixture of primers specific to OXA and KPC carbapenemases (5 pairs of primers). The results of electrophoretic analysis of the PCR-products obtained during the amplification of various carbapenemase genes from control microorganism strains are presented in Fig. 3.

The yield of labeled specific products from the multiplex PCR procedure was approximately 40-50 ng/ μ l for each type of carbapenemases gene, which was sufficient for further microarray-based hybridization analysis. Amplification of nonspecific products was detected only for genes from the VIM-1 and VIM-2 subgroups: however, these products had low yields and, as it follows from later experiments, their presence did not affect the specificity of the hybridization analysis.

Oligonucleotide microarray for detecting A-, B-, and D-class carbapenemase genes

The DNA microarray for the identification of the major types of carbapenemases is made on a support of nitrocellulose and has a size of 6.0 x 9.5 mm. On its surface there are 40 immobilized oligonucleotide probes (14 probes for the identification of 7 distinct groups of carbapenemases and 26 probes for additional typing of these genes into subgroups). Each microarray also includes 3 types of control oligonucleotides: an immobilization control (a biotin labeled oligonucleotide), a positive hybridization control (an oligonucleotide whose sequence is complementary to a biotin-labeled oligonucleotide which is added to the hybridization mixture), a negative hybridization control (an oligonucleotide with a random base sequence). In order to increase the reproducibility of this procedure, each oligonucleotide probe is present on the microarray in three copies. A schematic of the layout of the specific and control oligonucleotide probes on the surface of the DNA microarray is presented in Fig. 4.

The carbapenemase gene identification procedure involved hybridization analysis on a DNA microarray and included the following stages: 1) amplification of the β -lactamase gene from DNA isolated from the clinical strain (by two multiplex PCR reactions); 2) hybridization of the biotin-labeled DNA with oligonucleotide probes on the surface of the microarray; 3) visualization of the hybridization results using a streptavidin-peroxidase conjugate followed by colorimetric detection of the enzyme.

We also optimized hybridization conditions. We analyzed hybridization efficiency at temperatures ranging from 40 to 50°C; the temperature could not be higher than the melting temperature (T_m) of any oligonucleotide probe and was limited by the high level of nonspecific hybridization at lower temperatures. Hybridization at 40°C displayed strong signals: however, most of the probes showed cross-hybridization with the genes of various carbapenemases. Hybridiza-

RESEARCH ARTICLES

Table 1. Sequences of the specific and control oligonucleotide probes

Name	Nucleotide sequence, 5'→3'	Length, nucleotides	G/C, %	T _m , °C
Control oligonucleotides				
Immobilization control	TCTAGACAGCCACTCATA-Biotin	18	44.4	60.4
Positive hybridization control	GATTGGACGAGTCAGGAGC	19	57.9	66.1
Negative hybridization control	TCTAGACAGCCACTCATA	18	44.4	60.4
Oligonucleotide probes for determining the carbapenemase group				
KPC_direct	GCTTCCCCTGTGCAGCTCATTC	23	56.5	72.0
KPC_reverse	GAATGAGCTGCACAGTGGGAAGC	23	56.5	72.0
VIM_direct	GGAGATTGAAAAGCAAATTGGACT	24	37.5	66.6
VIM_reverse	AGTCCAATTTGCTTTTCAATCTCC	24	37.5	66.6
IMP_direct	GGAATAGAGTGGCTTAATTCTCG/A	23	41.3	64.7
IMP_reverse	C/TGAGAATTAAGCCACTCTATTCC	23	41.3	64.7
SPM_direct	GATGGGACCGTTGTCATTG	19	52.6	64.9
SPM_reverse	CAATGACAACGGTCCCATC	19	52.6	64.9
SIM_direct	CCTTGGCAATCTAAGTGACGCAA	23	47.8	69.7
SIM_reverse	TTGCGTCACTTAGATTGCCAAGG	23	47.8	69.7
GIM_direct	CACACTGGGAAATGGGCTTATA	22	45.5	66.7
GIM_reverse	TATAAGCCCATTTCCCAGTGTG	22	45.5	66.7
OXA_direct	CCACAA/GGTG/AGGC/TTGGTTG/AAC	20	55.0	67.0
OXA_reverse	GTC/AAACCAG/ACCC/TACT/CTGTGG	20	55.0	67.0
Oligonucleotide probes for determining the carbapenemase sub-group				
VIM-1_568	TCAGCGAACGTGCTATACGG	20	55.0	68.3
VIM-1_590	GTTGTGCCGTTTCATGAGTTGT	21	47.6	67.9
VIM-2_568	TCTGCGAGTGTGCTCTATGG	20	55.0	67.9
VIM-2_590	GTTGTGCCGATTTATGAGTTGT	21	38.1	63.7
VIM-7_127	GTTCCGCTGTACAAGATTGGCG	22	54.5	70.0
VIM-7_181	CTCGGTGACACGGTGTAC	18	61.1	65.8
IMP-1_135	GTGGGGCGTTGTTCTAAACATG	23	52.2	70.2
IMP-1_387	GGTTCAAGCCACAAATTCATTTAGC	25	40.0	67.8
IMP-2_264	TCAAAGGCACTATTTCTCACATTTTC	26	38.5	68.2
IMP-2_497	TACCTGAAAAGAAAATTTTATTCTGGTG	27	29.6	65.7
IMP-5_506	AATAGAGTTTTGTTCCGGTGGTT	22	36.4	65.0
IMP-5_459	TGGTCCAGGGCACACTCC	18	66.7	70.4
IMP-11_570	TGTTGAAGCATGGCCACATT	20	45.0	67.6
IMP-11_621	TGCAAAACTGGTTGTTCCAAGCC	23	47.8	70.9
IMP-12_226	AAATTAGTTGCTTGGTTTGTAGGG	24	37.5	66.4
IMP-12_495	GCTACCTGAAAACAAAATTTTATTTCG	26	30.8	64.8
IMP-14_292	GGTGACAGTACGGCTGGAATAG	22	54.5	68.4
IMP-14_374	AAAAAGACAATAAGGTACAAGCTA	24	29.2	63.4
OXA-23_225	AAATACAGAATATGTGCCAGCCTCT	25	40.0	68.8
OXA-23_309	GAAGGGCGAGAAAAGGTCATTTAC	24	45.8	68.0
OXA-40_225	AAATAAAGAATATGTCCCTGCATCA	25	32.0	65.6
OXA-40_329	GAACCTATCCTATGTGGGAGAAAAG	24	41.7	64.8
OXA-51_225	TTGACCGAGTATGTACCTGCTTCG	25	52.0	71.7
OXA-51_578	GCCCAAAAGTCCAAGATGAAG	21	47.6	65.8
OXA-58_225	AAAAACAGCTTATATTCCTGCATCT	25	32.0	66.0
OXA-58_206	GCACGCATTTAGACCGAGC	19	57.9	67.7

RESEARCH ARTICLES

Table 2. Primer sequences for the multiplex PCR-amplification of carbapenemase genes

Type		Sequence 5'→3'	Length, nucleotide	G/C, %	T _m , °C	Length of PCR-product, bp
KPC	direct	TTCTGCTGTCTTGTCTCTCATGG	23	47.8	64.7	801
	reverse	CCTCGCTGTGCTTGTTCATCC	20	60.0	65.7	
IMP-1	direct	GGCGTTTATGTTTCATACTTCGTTTG	25	40.0	64.4	584
	reverse	GTAAGTTTCAAGAGTGATGCGTCTCC	26	46.2	65.6	
IMP-2	direct	GGTGTTTATGTTTCATACATCGTTCG	25	40.0	63.8	584
	reverse	GTACGTTTCAAGAGTGATGCGTCCCC	26	53.8	67.8	
IMP-5	direct	GGTGTTTATGTTTCATACTTCGTTTG	25	36.0	62.5	584
	reverse	GTACGTTTCAAGAGTGATACATCTCC	26	42.3	63.4	
IMP-11	direct	GGTGTTTATGTTTCATACATCGTTTG	25	36.0	62.6	584
	reverse	GTAAGCTTCAAGAGCGACGCATCTCC	26	53.8	67.8	
IMP-12	direct	GGTGTTTATCTTCATACATCTTTTG	25	32.0	60.5	584
	reverse	GTAAGTTTCAAGAGTGATGCGTTCCC	26	46.2	66.0	
VIM-1	direct	GTAGTTTATTGGTCTACATGACCGCGTC	28	46.4	66.9	743
	reverse	CGCTGTGTGCTGGAGCAAGTC	21	61.9	68.1	
VIM-2	direct	GTAAGTTATTGGTCTATTTGACCGCGTC	28	42.9	65.9	743
	reverse	CGTTGTGTGCTTGAGCAAGTC	21	52.4	64.7	
VIM-7	direct	AGCATATTCCGCACAGCCTGG	21	57.1	67.5	685
	reverse	CCGGGCGGTCTGGAATTGCTC	21	66.7	67.7	
SPM	direct	CGTTTTGTTTGTGCTCGTTGCGGG	25	52.0	67.4	648
	reverse	CCTTCACATTGGCATCTCCCAGATAAC	27	48.1	67.2	
SIM	direct	GTTTGCGGAAGAAGCCCAGCC	21	61.9	68.6	613
	reverse	CTCCGATTTCACTGTGGCTTGGG	23	56.5	67.6	
GIM	direct	CTTGTAGCGTTGCCAGCTTTAGCTC	25	52.0	67.8	638
	reverse	CTGAACCTCCAACTTTGCCATGCC	24	50.0	66.9	
OXA-23	direct	GAAACCCCGAGTCAGATTGTTCAAG	25	48.0	65.8	686
	reverse	GGCATTCTGACCGCATTTCC	21	52.4	64.8	
OXA-40	direct	GTTTCTCTCAGTGCATGTTTCATC	23	43.5	62.3	714
	reverse	CATTTCTAAGTTGAGCGAAAAGGGG	25	44.0	64.6	
OXA-51	direct	CGAAGCACACACTACGGGTG	20	60.0	65.4	649
	reverse	CTCTTTTCGAACAGAGCTAGGTATTC	26	42.3	63.4	
OXA-58	direct	CTTGTGCTGAGCATAGTATGAGTC	24	45.8	63.3	684
	reverse	CCACTTGCCCATCTGCCTTTTC	22	54.5	66.5	

Table 3. Results on clinical sample testing on DNA microarrays

Type of microorganism	Carbapenemase sensitivity as determined by phenotypical tests	Number of samples	Detected carbapenemase types				
			OXA-23	OXA-40	OXA-51	OXA-58	VIM-2
<i>A. baumannii</i>	Resistant	10	1	5	10	4	-
<i>Ps.aeruginosa</i>		11	-	-	-	-	11
<i>A. baumannii</i>	Sensitive	2	-	-	-	-	-
<i>K. pneumonia</i>		3	-	-	-	-	-
<i>E. coli</i>		2	-	-	-	-	-

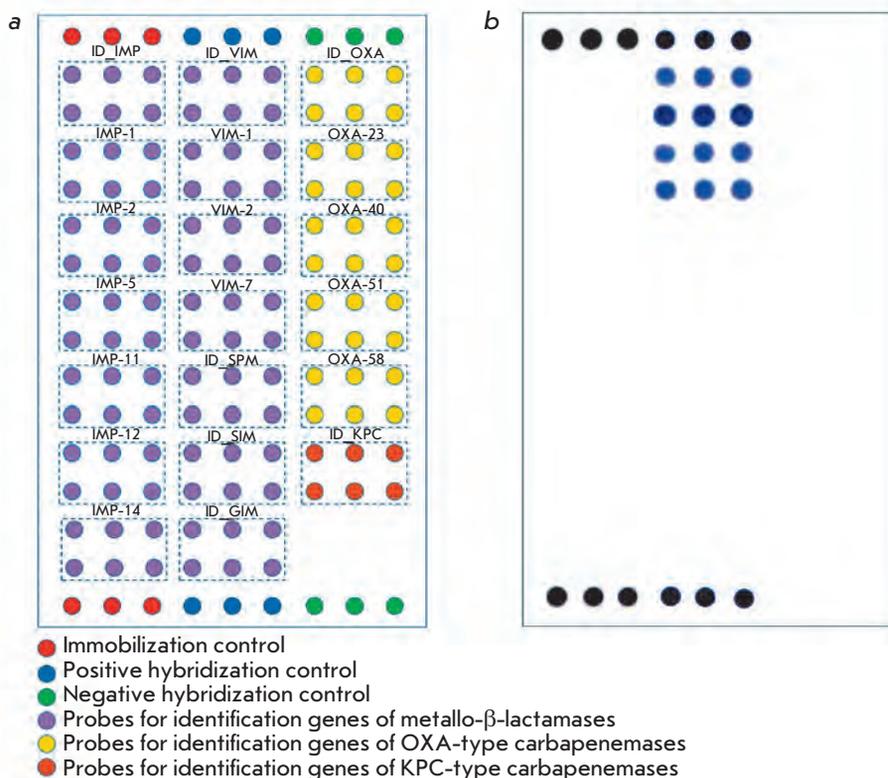


Fig. 4. Layout of specific and control oligonucleotide probes on the surface of the DNA microarray for the identification of the A-, B-, D-class carbapenemase genes.

tion at 50°C displayed weak signal intensities for some of the probes. For these reasons, we chose 45°C as the optimal temperature. The hybridization buffer consisted of 2x SSPE supplemented by 0.2 % sodi dodecylsulphate (SDS) in order to improve the membrane's wetting properties.

The size of the labeled DNA-target proved to be a critical parameter for hybridization. Hybridization of labeled PCR-products whose size was 580–800 nucleotides with the appropriate oligonucleotide probes proved to yield weak signals. Additional fragmentation by DNAase (yielding fragments of 50–150 nucleotides) proved to increase hybridization signal intensity for most of the probes.

Hybridization duration was assayed within a range of 0.5 to 4.0 hours. It was observed that the hybridization of biotin-labeled DNA onto immobilized probes reaches equilibrium after two hours of incubation with active mixing. We also noticed that conducting the reaction in kinetic conditions (1 hour) does not dramatically weaken the signals as compared to the equilibrium state (about 10 – 20% depending on the probe). It also did not lower the specificity of the analytic procedure, which allowed the positive identification of all types of carbapenemase genes in the hybridization mixture.

Figure 5 shows the results of an experiment in which control microorganism strains producing VIM-1 and IMP-1 metallo-β-lactamases and carbapenemases

OXA-51 and OXA-40 were tested on our DNA microarray. Identification of the β-lactamase group was assayed by the hybridization intensity with a group-specific probe, while additional typing was assayed by the hybridization intensity with subgroup-specific probes. The advantage of the microarray-based hybridization analysis is the possibility of simultaneously detecting several genes, which is demonstrated by testing the control *A. baumannii* strain for OXA-type carbapenemases.

Testing clinical strains of microorganisms resistant to carbapenems

The DNA microarray developed was tested on clinical strains of gram-negative microorganisms, either resistant or sensitive to carbapenems, as assayed by a phenotypical test. Table 3 shows the results for 28 clinical strains of *Ps. aeruginosa*, *A. baumannii* and *Enterobacteriaceae* spp., which display various levels of carbapenem sensitivity (strains were provided by the N.N. Burdenko Institute of Neurosurgery and Institute of Antimicrobial Chemotherapy of Smolensk State Medical Academy). All the *A. baumannii* strains which proved resistant to carbapenems (as assayed by phenotyping) expressed two carbapenemase genes (OXA-51 and OXA-23 (1 strain), OXA-40 (5 strains), OXA-58 (4 strains)). All the carbapenem-resistant *Ps. aeruginosa* strains happened to possess a VIM-2-type metallo-β-lactamase gene. Testing of carbapenem-sensitive

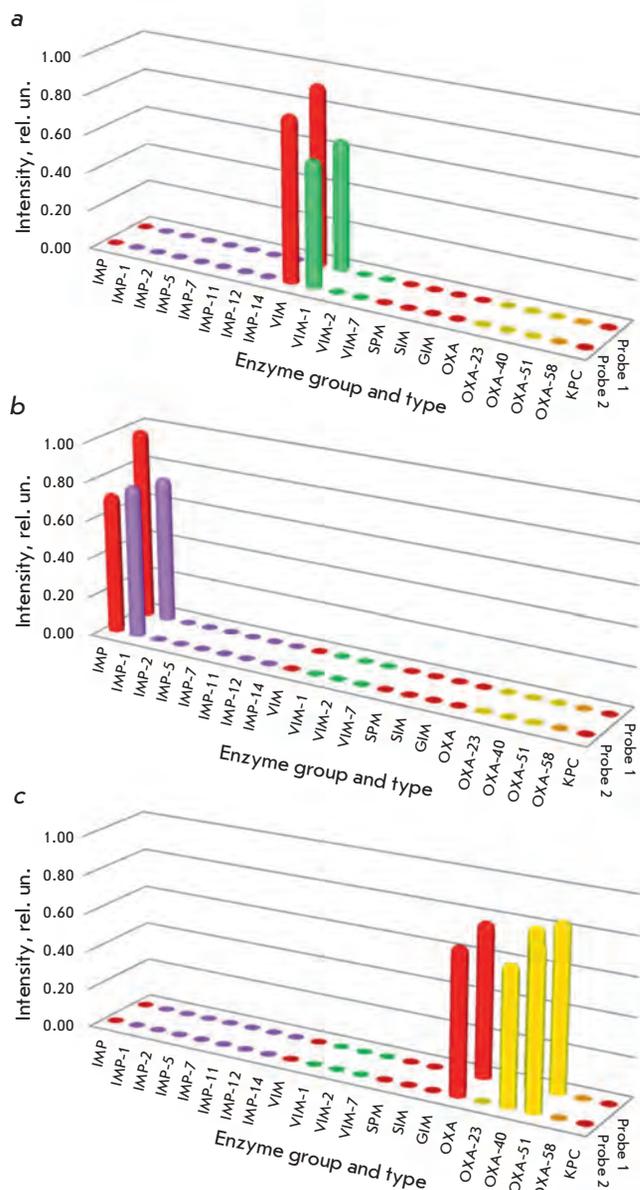


Fig. 5. Results of a DNA microarray-based hybridization analysis for control *Ps. aeruginosa* (a, b) and *A. baumannii* (c) strains, producing carbapenemases VIM-1 (a), IMP-1(b), OXA-51 and OXA-40 (c).

strains revealed no carbapenemase genes. Thus, the results of the microarray-based hybridization analysis are in accordance with the phenotyping tests. Moreover, the structure of the observed carbapenemases was confirmed by gene sequencing for two carbapenem-resistant samples - one strain of *A. baumannii* and one strain of *Ps. aeruginosa*, which expressed OXA-40, OXA-51 and VIM-2 genes.

Thus, our method of hybridization analysis based on DNA microarray for the identification and typing of

carbapenemase genes is highly accurate, productive, and can be used in clinical microbiological laboratories for the identification of carbapenemases and for studying their epidemiology. The phenotyping tests currently being used take time (from 24 to 48 hours) and are not always effective for determining carbapenemase types, such as OXA-type carbapenemases. Identification of carbapenemase genes on DNA microarrays allows rapid diagnostics, with the whole procedure taking only 4.5 hours, including 0.5 hours for bacterial DNA extraction, 1.5 hours for amplifying the carbapenemase genes and fragmenting the PCR-products, 1.5 hours for the hybridization and washing steps, and 1 hour for colorimetric detection of the hybridization results. An important feature of this method is the possibility of simultaneously identifying several genes in one sample. ●

REFERENCES

- Galkin D. V. // Klin. microbiol. antimicrob. khimioter. (Clin. Microbiol. Antimicrob. Chemother). 2007. V.9. P. 133–152.
- Skleenova E., Shevchenko O., Edelstein M., et al. // Abstracts of 20th European Congress of Clinical Microbiology and Infectious Diseases. Vienna, Austria, 2010, 10–13 April. O 562.
- Livermore D. // J. Antimicrob. Chemother. 2001. V.47. P. 247–250.
- Jacoby G., Mills D.M. // Antimicrob. Agents Chemother. 2004. V.48. P. 3203–3206.
- Ziha-Zarifi I., Llanes C. // Antimicrob. Agents Chemother. 1999. V.43. P. 287–291.
- Queenan A.M., Bush K. // Clin. Microbiol. Rev. 2007. V.20. P. 440–458.
- Miriagou V., Cornaglia G., Edelstein M. // Clin. Microbiol. Infect. 2010. V.16. P. 112–122.
- Bush K., Jacoby G. // Antimicrob. Agents Chemother. 2010. V.54. P. 969–976.
- Rasmussen J.W., Hoiby N. // J. Antimicrob. Chemother. 2007. V.60. P. 470–482.
- Walsh T.R., Toleman M.A. // Clin. Microbiol. Rev. 2005. V.18. P. 306–325.
- Rasmussen J.W., Hoiby N. // J. Antimicrob. Chemother. 2006. V.57. P. 373–383.
- Moland E.S., Hong S.G. // Clinical Microbiology Newsletter. 2008. V.30. P. 79–85.
- Noval M.J., Menezes G.A., Harish B.N., et al. // Indian. J. Med. Res. 2009. V. 129. P. 707–712.
- Poirel L., Naas T., Nordmann P. // Antimicrob. Agents Chemother. 2010. V.54. P. 24–38.
- Shevchenko O.V., Edelstein M.V., Stepanova M.N. // Klin. microbiol. antimicrob. khimioter. (Clin. Microbiol. Antimicrob. Chemother). 2007. V.9. P. 211–218.
- Pitout J.D., Gregson D.B., Poirel L. // J. Clin. Microbiol. 2005. V. 43. P. 3129–3135.
- Woodford N., Ellington M. // Int. Journal of Antimicrobial Agents. 2006. V. 27. P. 351–353.
- Mendes R.E., Kiyota K.A. // J. Clin. Microbiol. 2007. V. 45. P. 544–547.
- Bilitewski U. // Methods Mol. Biol. 2009. V. 509 P. 1–14.
- Ehrenreich A. // Appl. Microbiol. Biotechnol. 2006. V. 73. P. 255–273.
- Zhang Y., Coyne M.Y., Will S.G. et al. // Nucleic Acids Res. 1991. V. 19. P. 3929–3935.

New Test System for Serine/Threonine Protein Kinase Inhibitors Screening: *E. coli* APHVIII/Pk25 design.

O.B. Bekker¹, M.G. Alekseeva¹, D.I. Osolodkin², V.A. Palyulin², S.M. Elizarov³, N.S. Zefirov², V.N. Danilenko^{1*}

¹Vavilov Institute of General Genetics, Russian Academy of Sciences

²Lomonosov Moscow State University, Department of Chemistry

³Bach Institute of Biochemistry, Russian Academy of Sciences

*E-mail: valerid@rutenia.ru

Received 28.06.2010

ABSTRACT An efficient test system for serine/threonine protein kinase inhibitors screening has been developed based on the *E. coli* protein system APHVIII/Pk25. Phosphorylation of aminoglycoside phosphotransferase VIII (APHVIII) by protein kinases enhances resistance of the bacterial cell to aminoglycoside antibiotics, e.g. kanamycin. Addition of protein kinase inhibitors prevents phosphorylation and increases cell sensitivity to kanamycin. We have obtained modifications of APHVIII in which phosphorylatable Ser146 was encompassed into the canonical autophosphorylation sequence of *Streptomyces coelicolor* Pk25 protein kinase. Mutant and wild-type *aphVIII* were cloned into *E. coli* with the catalytic domain of *pk25*. As a result of the expression of these genes, accumulation of corresponding proteins was clearly observed. Extracted from bacterial lysates, Pk25 demonstrated its ability to autophosphorylate. It was shown that variants of *E. coli* containing both *aphVIII* and *pk25* were more resistant to kanamycin than those carrying only *aphVIII*. Protein kinase inhibitors of the indolylmaleimide class actively inhibited Pk25 and reduced cell resistance to kanamycin. Modeling of APHVIII and Pk25 3D structures showed that pSer146 is an analog of phosphoserine in the ribose pocket of protein kinase A. Pk25 conformation was similar to that of PknB of *Mycobacterium tuberculosis*. Potential indolylmaleimide inhibitors were docked into the ATP-binding pocket of Pk25. The designed test system can be used for the primary selection of ATP-competitive small molecule protein kinase inhibitors.

KEYWORDS serine/threonine protein kinases, indolylmaleimides, protein kinase inhibitors screening, bacterial test system, *Streptomyces*

Serine/threonine protein kinases (STPK) are a group of universal regulators of the cellular metabolism in eukaryotes [1-3]. They play leading roles in the regulation of apoptosis, proliferation and differentiation, cellular transport, etc. It has been shown that kinase malfunction can be associated with various human diseases, such as diabetes [4], schizophrenia [5], cardiovascular disorders [6,7], and immune dysfunctions [8]. Over the past decades, the search for new targeted modulators (inhibitors) of protein kinases, that are considered as potential drugs, has been intensive [9-11].

Eukaryotic type serine/threonine protein kinases were found in bacteria, including human pathogens [12]. It was shown that STPKs participate in virulence, bacterial biofilm formation, tolerance, and persistence of pathogenic microorganisms. STPKs were shown to play the key role in the virulence of *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and certain other pathogens [12-14]. It is established that STPKs par-

ticipate in the modulation of antibiotic resistance in *M. tuberculosis* [15]. Intensive STPK inhibitors screening is currently under way [16-19].

We have developed a test system [20] for prescreening of STPK inhibitors based on the newly constructed strain *Streptomyces lividans* TK24 (66) APHVIII+. The key part of this test system is type VIII aminoglycoside phosphotransferase (APHVIII), an enzyme that inactivates aminoglycoside antibiotics. Gene *aphVIII*, which is isolated from *Streptomyces rimosus*, was cloned into *S. lividans* TK24 (66) and expressed. Importantly, the activity of APHVIII *S. rimosus* is dependent on phosphorylation by endogenous STPKs [21]. Phosphorylation of APHVIII confers *Streptomyces* cells resistance to aminoglycoside antibiotics, while inhibitors of STPK render cells more sensitive to aminoglycosides [20]. Alteration of the cellular sensitivity to aminoglycoside antibiotics in the presence of STPK inhibitors allows to perform primary screening of these inhibitors. Upon annotation of the genome of the *S. coelicolor* A3(2) strain (NC_003888), which is close to *S. lividans* TK24

(66) (ACEY01000000), 34 STPKs were identified. At least one of them – Pk25 (NCBI Reference Sequence: Protein NP_628936.1) – is able to phosphorylate APH-VIII [22]. In order to rule out a nonspecific action of STPK inhibitors on the other STPKs of *S. lividans* TK24 (66) that are presumably able to phosphorylate APHVIII, the catalytic domain of Pk25 kinase and APHVIII were hosted in *Escherichia coli*. The genome of *E. coli* does not contain its own eukaryotic STPKs, which makes the test system more sensitive and allows the screening of inhibitors that are specific to Pk25 and its homologous enzymes [23].

EXPERIMENTAL PROCEDURES

Strains: *S. coelicolor* A3(2) (Russian Collection of Pathogenic Microorganisms, Moscow), *S. lividans* TK24 (66) APHVIII+ (GenBank ACEY01000000), *E. coli* DH5 α : F⁻, Φ 80 Δ lacZ Δ M15, Δ (lacZYA-argF), U169 (Promega); BL21(DE3): F⁻, dcm, ompT, hsdS(r_B⁻m_B⁻), gal λ (DE3) (Novagen).

Plasmids: pET16b, pET22b and pET32a (Novagen).

Media: *S. coelicolor* A3(2) and *S. lividans* TK24 (66) APHVIII+ strains were grown on YSP and YEME media [24]. *E. coli* strains were grown on Luria broth (L-broth), NZCYM, M9 supplemented with 1.5 % glycerol (1 L): 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl, 1 g NH₄Cl, pH 7.4, 2 mL 1 M MgSO₄, 15 mL glycerol. Ampicillin (100 μ g/ml) was added to the media to ensure the survival of plasmid-containing cells. Protein expression was induced by IPTG (1 mM).

Molecular cloning: Total DNA was isolated from *S. coelicolor* A3(2) and *S. lividans* TK24 (66) APHVIII+ strains according to [24]. Isolation of plasmid DNA, preparation of competent *E. coli* cells, transformation and analysis of recombinant plasmids was performed according to standard protocols [25]. DNA amplification by PCR was carried out by the “Amplification” kit (Dialat Ltd.) on the Tertsik TP4-PCR01 amplifier (DNA-tekhnologiya). The temperature cycle was designed according to primer length and composition. Oligonucleotides were purchased from Syntol (see Table 1). The DNA sequence was determined according to Sanger. Nucleotide sequences were compared with BLAST (<http://www.ncbi.nlm.nih.gov/blast>).

Protein electrophoresis was carried out in 12% polyacrylamide gels under denaturing conditions, as described earlier [21]. Bacterial cells containing constructed plasmids were grown on a NZCYM liquid medium supplemented with ampicillin (150 μ g/ml) at 34°C up to an optical density equal to 0.6 (~1,5 h). Expression of the Pk25 catalytic domain was induced by the addition of IPTG (final concentration 1 mM). Cells were grown at 28°C for 4 h, harvested and re-suspended in a buffer containing 62.5 mM TrisHCl, pH 6.8; 5% glycerol; 2%

β -mercaptoethanol; 0.1% SDS; and Bromophenol Blue. Cells were lysed by boiling for 10 min in the above-mentioned buffer and subjected to electrophoresis in polyacrylamide gels. Approximately 25 μ g of the protein was loaded into each well in the gel. Electrophoregrams were scanned by laser densitometer Ultrosan 2205 LKB. A protein fraction of *E. coli* BL21(DE3) cells transfected with the empty vector was used as control.

Isolation of the catalytic domain of Pk25 cloned into *E. coli* cells. Cells were destroyed by sonication in a buffer containing 20 mM TrisHCl, pH 7.8, 10 mM 2-mercaptoethanol, 300 mM NaCl, and 1 mM PMSF, or in the same buffer supplemented with 8 M urea. Cell debris and other undissolved material was removed by centrifugation at 20,000g in 20 min. Fractions of soluble proteins were loaded onto a Ni-NTA agarose column (Qiagen) which was washed by the above-described buffer containing 50 mM imidazole pH 6.0. The protein was eluted by the buffer with imidazole concentration increasing from 0.05 M up to 0.5 M [20]. Protein fractions were analyzed by SDS-PAGE.

Analysis of autophosphorylation of isolated protein by catalytic domain of Pk25 *in situ* was performed after separation of the protein under denaturing conditions. Re-naturation of the kinase in gel was carried out according to Kashemita and Fujisawa [26]. Gels containing the protein were intensively washed in 50 mM TrisHCl, pH 7.8, with 25% 2-propanol and 8 M urea in order to remove SDS. Following that, protein re-naturation was carried out by washing gels in buffer A: 50 mM TrisHCl, pH 7.8 and B: 50 mM TrisHCl, pH 7.8, 100 mM NaCl, 6 mM β -mercaptoethanol, 5 mM MgCl₂, and 1 mM CaCl₂. After re-naturation, the gels were incubated in the presence of 50 μ Ci/ml [γ -³²P]ATP (7000 Ci/mM, Phosphor, Russian Federation) in the buffer for the analysis of kinase activity [21]. The gels were fixed and stained in 40% TCA, washed in 5% acetic acid, dried and autoradiographed by exposure to a Kodak X-Omat AR film.

Cloning into expressional vectors pET32a, pET22b, and pET16b. Gene *pk25* of the *S. coelicolor* A3(2) strain and the gene of *S. lividans* TK24 (66) were cloned into *E. coli* in pET32a plasmid at EcoRI and HindIII (primers Pk25EN and Pk25C) (Table 1). The gene of the catalytic domain of *pk25* of the *S. coelicolor* A3(2) strain was cloned into *E. coli* in pET22b plasmid at NdeI and HindIII (primers Pk25CN and Pk25CC). The modified gene *aphVII* was cloned into *E. coli* in pET16b plasmid at NdeI and XhoI (primers AphN and AphC). The gene of the catalytic domain of *pk25* of the *S. coelicolor* A3(2) strain was cloned into *E. coli* in pET16b + *aphVIII146-S* with the non-modified phosphorylation site of APHVIII, pET16b + *aphVIII146-1*, pET16b + *aphVIII146-2*, and pET16b + *aphVIII146-3* with modified phospho-

rylation sites at the BamHI (primers Pk25NBgl and Pk25CBgl).

Cloning of the nucleotide sequence of the *pk25* catalytic domain was performed in the pET22b vector at NdeI–HindIII restriction sites (primer Pk25CN homologous to the N-terminal region of the catalytic domain, and primer Pk25CC homologous to the C-terminal region of the catalytic domain). The DNA sequence of the catalytic domain was amplified using total DNA of *S. coelicolor* as a template. The PCR product was purified from the agarose gel, then sequenced and cloned into the pET22b vector at the NdeI and HindIII. *E. coli* DH5a cells were transformed by the resulting ligase mix, and screening of recombinant clones was carried out by PCR using standard primers T7prom and T7term. Plasmid DNA was isolated from selected transformants, and the obtained recombinant plasmids were sequenced and subjected to restriction analysis to verify the presence of the insert. Then, the plasmids were used for the transformation of *E. coli* BL21 (DE3) cells.

Site-directed mutagenesis of Ser146 in aminoglycoside phosphotransferase APHVIII was carried out according to Nelson [27]. In order to obtain the mutant variant 1 (amino acid substitutions Ser146Thr, Glu144Thr, Asp148Ser), the primers APH 146-1(+) and APH 146-1(-) (Table 1) were used. The primers

APH146-2(+) and APH146-2(-) were used to obtain the mutant variant 2 (Glu144Thr, Asp148Ser, Glu150Ser). The mutant variant 3 represents substitution Ser146Thr, which was introduced using the APH146-T(+) and APH146-T(-) primers.

AphN and AphC corresponding to the 5'- and 3'-terminal fragments of the *aphVIII* gene were used as flanking primers.

The obtained mutant PCR fragments were sequenced to verify the nucleotide substitutions and cloned at the NdeI and BamHI restriction sites into the high-copy number plasmid pET16b containing T7 phage transcriptional and translational regulatory elements in the same reading frame with the ATG codone of the gene of interest. This plasmid was used for the transformation of *E. coli* DH5a cells, and screening of the recombinant clones was carried out by PCR using T7prom and T7term primers. The selected transformants were used for purification of plasmid DNA, which was re-sequenced to ensure the substitutions.

Determination of resistance to kanamycin in selected transformants of *E. coli* BL21(DE3). *E. coli* BL21(DE3) clones containing genes of either native or modified *aphVIII* or *aphVIII* and *pk25* in the pET16b vector were used for the analysis. The clones resistant to ampicillin (100 µg/ml) were transferred on the plates with the LB-medium containing various concentrations

Table 1. Primers which were used in the present work*

Primer	Restriction site	Primer structure, 5'–3'
Pk25EN	EcoRI	ATCCGAATTATGGCACGGAAGATCGGCAG
Pk25C	HindIII	CCGCAAGCTTGGTGCCGTTGCCGGAACCG
Pk25CN	NdeI	TCGTCATATGCGTTACCGGCTCCATGAGCGGC
Pk25CC	HindIII	CCGCAAGCTTCATCCGCTGGGCCGACGCCG
Pk25NBgl	Bgl II	TTTTAGATCTAATAAAGGAGATATACATGTACCGGCTCCATGAGCGGCT RBS beginning of <i>pk25</i> catalytic domain
Pk25CBgl	Bgl II	CCG CAG ATC TAT CCG CTG GGC CGA CGC CGC
T7prom	—	TTAATACGACTCACTATAGG
T7term	—	CTAGTTATTGCTCAGCGG
APH 146-1(+)	—	GCTGTCGCTACAGGGACGGTCAGCTTGGAGGATCTGGAC
APH 146-1(-)	—	GTCCAGATCCTCCAAGCTGACCGTCCCTGTAGCGACAGC
APH 146-2(+)	—	GCTGTCGCTACAGGGAGCGTCACCTTGTCCGATCTGGACGAG
APH 146-2(-)	—	CTCGTCCAGATCCGACAAGGTGACGCTCCCTGTAGCGACAGC
APH 146-T(+)	—	GTCGCTGAAGGGACCGTCCGACTTGGAG
APH 146-T(-)	—	CTCCAAGTCGACGGTCCCTTCAGCGAC
AphN	NdeI	TTTTCATATGGACGATGCGTTGCGTGC
AphC	BamHI	TTTTGGATCCTCAGAAGAACTCGTCCAAC

Restriction sites are given in **bold, nucleotide substitutions are underlined.

of aminoglycoside antibiotic kanamycin and IPTG as the inducer. The growth of colonies was monitored after 25 hours of incubation at 37°C.

Determination of activity of protein kinase inhibitors in the bacterial test system. Inhibitor activity was determined using paper discs. The paper discs soaked with antibiotic or an antibiotic/inhibitor mixture were placed on the surface of agar plates, and the size of the zone of bacterial growth suppression was measured. Test system: *E. coli* BL21(DE3)APHVIII/Pk25. Bacteria grown on agar plates supplemented with ampicillin were washed off into a liquid LB-medium and grown overnight at 37°C. Then the cells were pelleted (4,000 rpm, 10 min) and re-suspended in a liquid M9 medium. Bacterial suspension was mixed up in a 1:1 ratio with a melted M9-agar medium containing ampicillin and IPTG as the inducer. The mixture was poured onto Petri dishes with a M9-agar medium containing ampicillin and IPTG. Ampicillin is necessary for maintaining the plasmid in the *E. coli* cells. Paper discs containing kanamycin or kanamycin plus the kinase inhibitor were placed on the surface of the agar plates and incubated for 16 hours at 37°C.

Modeling of Pk25 catalytic domain structure. X-ray structures of kinase from *M. tuberculosis PknB* (PDB entries PDB [28]: 1MRU [29], 1O6Y [30], 2FUM [31], 3F61 [32], 3F69 [32]) were used as templates for the model building. Amino acid sequences of the template protein were extracted directly from structural files; the sequence of *S. coelicolor* A3(2) was taken from GenBank (access code 21223157 [33]). The catalytic domain was annotated according to homology. Alignment of amino acid sequences was carried out using the ClustalX 2.0.11 [34] software. The modeling was performed with the Modeller 9v5 [35] program. Thirty-five models of the catalytic domain were generated, and each of them was optimized by simulated annealing. The best model was selected according to the DOPE scoring function (Modeller software) and PROCHECK validation score [36]. Further optimization was performed in SYBYL 8.0 [37]; all hydrogen atoms were added to the models, and the energy was minimized in the Tripos force field [38] by the Powell method.

Modeling of the APHVIII catalytic domain structure was based on the X-ray structure of its closest homologue APH(3')-IIa (PDB entry 1ND4 [39], 36% identity of amino acid sequence) in complex with kanamycin. The modeling method is similar to the one described above: the only difference was that 50 models were generated in each case.

Docking of inhibitors into the Pk25 model was performed using the Autodock 4.1 [40] software. The structures of the inhibitors were drawn by SYBYL 8.0 and optimized with the MMFF94 force field [41]. Docking

preparation was carried out in MGLTools 1.5.4 [42] according to standard recommendations. Generation of the grids and docking were performed using default parameters, while the position of the docking grid was chosen to include all essential amino acid residues of the ATP-binding site. During docking of each ligand, 100 runs of the genetic algorithm were performed. Docking results were grouped into clusters using a threshold value of RMSD equal to 2.0 Å. The results were analyzed in MGLTools 1.5.4.

RESULTS

Cloning and comparison of full-size genes of serine/threonine protein kinase *pK25* from *S. coelicolor* and *pK25 S. lividans*.

According to genome sequencing data, the gene of the *pK25* kinase from *S. lividans* TK24 (ACEY01000000) and the gene of *pK25* kinase from *S. coelicolor* share 99.8% homology and differ only in 6 base pairs, including an insertion of C at position 664. The presence of this insertion leads to the shift of the reading frame in the catalytic domain sequence. In order to compare the kinases of interest from our collection of *S. coelicolor* A3(2) and *S. lividans* TK24 (66) strains, we cloned and sequenced the genes of these kinases. To isolate *pK25* kinase genes from the genomes of *S. coelicolor* and *S. lividans*, two oligonucleotides containing HindIII and EcoRI were synthesized (*pK25EN*, homologous to the N-terminal part of the gene; and *Pk25C*, complementary to the C-terminal part of the gene, Table 1). Amplification from total DNA of *S. coelicolor* and *S. lividans* yielded the needed DNA fragments, which were cloned into the pET32a vector at the HindIII and EcoRI and introduced into *E. coli* DH5a cells. Sequencing verified that the obtained DNA fragments contained the *pK25* gene. Comparison of the genes from *S. coelicolor* and *S. lividans* revealed nucleotide substitutions at positions 123, 237, 279, 435, and 963 starting from the first ATG of the structural region of the *S. lividans pK25* gene. These nucleotide substitutions do not lead to amino acid substitutions in the corresponding proteins. Therefore, the amino acid sequences of the full-size *pK25* genes from *S. coelicolor* and *S. lividans* were shown to be identical.

Cloning and expression of the *S. coelicolor* Pk25 catalytic domain nucleotide sequence into pET22b vector.

Two oligonucleotides (*Pk25CN* and *Pk25CC*) containing NdeI and HindIII were synthesized for the cloning (Table 1). The fragment obtained during amplification was digested with the relevant restriction endonucleases and cloned into the pET22b vector that was introduced into *E. coli* DH5a cells. After resequencing, the obtained vectors were introduced into the *E. coli* BL21(DE3) strain used for protein overexpression.

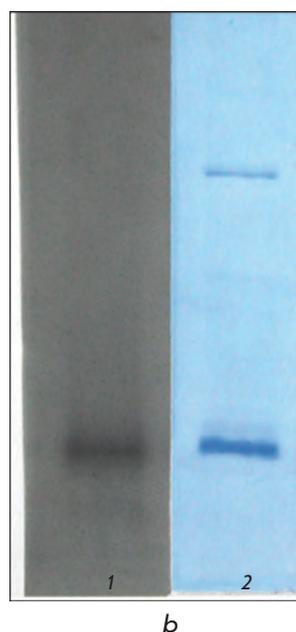
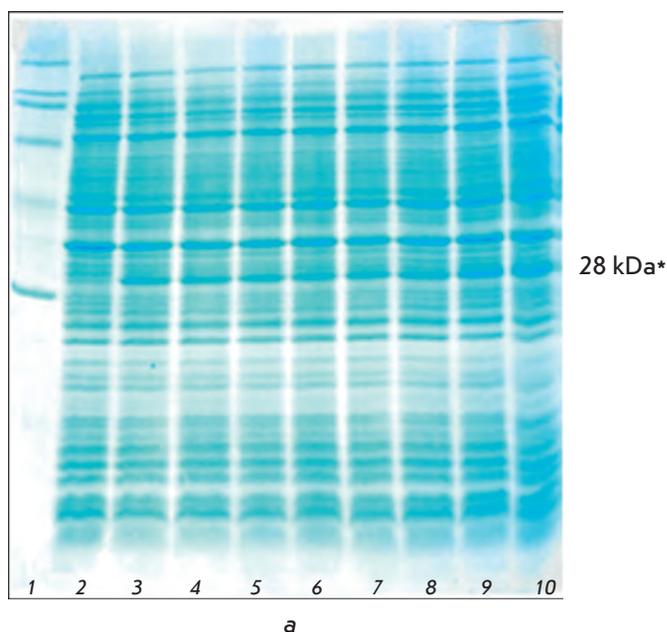


Fig 1. a – Electrophoresis of soluble protein fraction of *E. coli* BL21(DE3). 1 – marker, 2 – control protein fraction of *E. coli* BL21(DE3) pET22b, 3-10 – *E. coli* clones containing pET22b-pk25. (* – protein fraction of Pk25 catalytic domain). b – Electrophoresis of isolated protein fraction of *S. coelicolor* Pk25 catalytic domain in PAAG: 1 – autoradiogram autophosphorylate Pk25 catalytic domain, 2 – coloring coomassie brilliant blue.

In order to study the expression of the catalytic domain of Pk25 kinase in *E. coli* cells, we set up gel electrophoresis of a fraction of soluble cellular proteins under denaturing conditions. A protein fraction from *E. coli* BL21(DE3) transformed with an empty pET22b vector were used as a control. Electrophoresis showed that *E. coli* cells transformed with the plasmid carrying the gene of the pk25 catalytic domain containing an additional protein fraction of 28 kDa (Fig 1a) as compared with the control cells. This value coincides with the calculated molecular weight (27.8 kDa) of the catalytic domain of Pk25 kinase from the *S. coelicolor* A3(2) strain. Scanning revealed that the additional fraction represents about 3.5% of the total cellular protein.

Modeling of the Pk25 catalytic domain 3D structure.

The key structural features of Pk25 are revealed by the model, which is very close to the structure of the template protein PknB *M. tuberculosis* due to the close similarity of the sequences (Fig. 2a). The net charge of the catalytic domain is equal to -3. The most obvious differences between the model and the template are observed in the region of helix C (an insertion of four amino acid residues in Pk25), in the loop between fragments β 4 and β 5 (deletion of four amino acid residues in Pk25), and in the region of the η 3 helix (deletion of five amino acid residues in Pk25). Conformation of the activation loop differs from that in the template due to its flexibility. However, the mentioned differences do not affect the orientation of amino acid residues in the ATP-binding site. There are five amino acid residues in the ATP-binding pocket of Pk25 that differ from the corresponding residues of PknB (Fig. 2b): Val72Ile, Ile90Met, Tyr94Leu,

Met145Leu, and Met155Thr (numeration of the PknB sequence is used). The former three substitutions are relatively conserved and should not seriously affect the interaction with the inhibitor. Substitution Tyr94Leu is located in the hinge region, so ligands interact with the backbone of this residue but not with its sidechain. The latter two substitutions are not conservative and, therefore, should influence the interaction with the inhibitor. In particular, these substitutions increase the accessible volume of the binding pocket. Finally, an additional hydroxy group appears in this region due to the introduction of a threonine residue at position 155.

Analysis of autophosphorylation of the Pk25 activation loop.

A protein fraction of the Pk25 catalytic domain was isolated from lysed *E. coli* cells by means of chromatography on a His-binding resin under either native conditions or in the presence of 8M urea. After electrophoresis, we were able to analyze its intracellular localization and autophosphorylation *in situ*. The catalytic domain of Pk25 was absent in the fraction of salt-soluble cellular proteins. This recombinant protein is present in the fraction of insoluble cellular proteins and can be dissolved in the presence of urea, as was done. After separation by electrophoresis, the molecular mass of the studied fragment was determined as equal to 28 kDa, which is very close to the calculated value. In the presence of [γ - 32 P] ATP, the protein accumulates labeled phosphate (Fig. 1b). Separation of the protein components by gel electrophoresis excludes any possible interfering influences of other proteins. Over all, the obtained data indicate that the isolated catalytic domain of Pk25 kinase is an enzy-

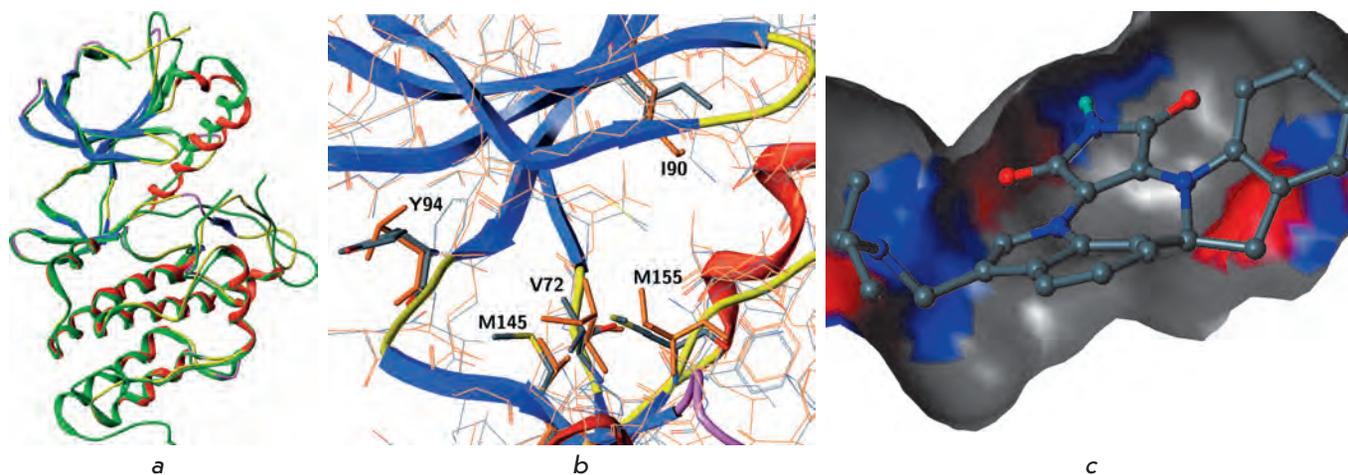
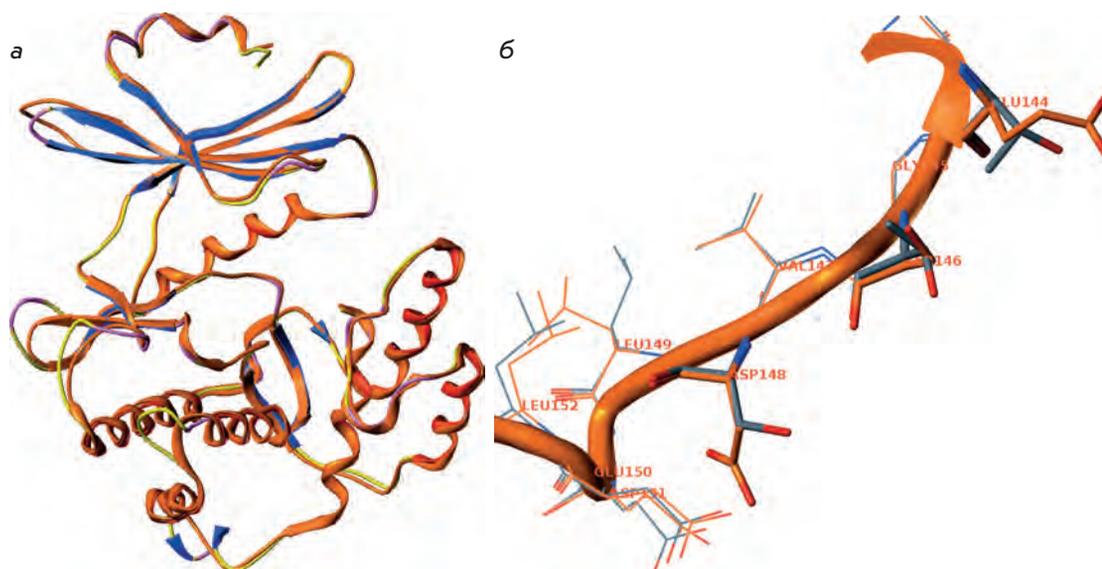


Fig 2. The model of Pk25 3D structure. *a* — Superimposition of the template structure PknB *M. tuberculosis* (green) and Pk25 model (colored by secondary structure type). *b* — ATP-binding pockets of PknB (colored by atom type) and Pk25 (orange); nonconservative amino acid residues are shown as sticks. *c* — binding mode of LCTA-1425 with Pk25. The kinase VdW surface is colored according to hydrogen bond donor/acceptor properties (red — donor, blue — acceptor).

Fig 3. The model of APHVIII 3D structure. *a* — Superimposition of the template structure APH(3')-IIa (PDB ID 1ND4, orange) and APHVIII model (colored by secondary structure type). *b* — Amino acid residues in the activation loop of wild-type APH-VIII (carbons are colored orange) and 146-1 mutant (grey carbons).



matically active protein and undergoes autophosphorylation. Localization of the Pk25 catalytic domain in the fraction of insoluble cellular proteins does not contradict the possibility of its autophosphorylation during expression and, likewise, does not exclude its activity towards both soluble and insoluble proteins. These results are in agreement with the data obtained on the catalytic domains of *Streptomyces* and *Mycobacterium* protein kinases [43–45]. It was shown that phosphorylation in the activation loop of STPK occurs at a Ser residue. Analysis of the substrate specificity of STPK from *M. tuberculosis* showed that the highest phosphorylation efficiency was observed at the regions that are similar to the autophosphorylation sites of the kinase [46].

Modeling of the aminoglycoside phosphotransferase VIII 3D structure

The model of the APHVIII catalytic domain structure closely resembles the template protein APH(3')-IIa (Fig. 3a). The two structures are most similar in the sites of the ATP and kanamycin binding, as well as in the activation loop. The small insertions in the APHVIII sequence are located in the structurally nonconserved regions of the loops, and between various elements of the secondary structure they are unlikely to seriously affect the 3D pattern of the polypeptide chain. The physico-chemical properties and conformations of the crucial amino acid residues that form the ATP and kanamycin binding sites also correspond within the template and

Table 2. Modifications of phosphorylation site Ser-146 in APHVIII

Native enzymes and modified species of APHVIII	Phosphorylation sites and their modifications*
Pk25	ATTLTESGSFVG
APHVIII	AVAEGS ₁₄₆ VDLED
APHVIII146-1	AVAT <u>GT</u> ₁₄₆ V <u>S</u> LED
APHVIII146-2	AVAT <u>GS</u> ₁₄₆ V <u>S</u> LSD
APHVIII146-3	AVAEG <u>T</u> ₁₄₆ VDLED
APHVIII146-4	AVAEG <u>A</u> ₁₄₆ VDLED

*Amino acid substitutions are underlined.

the model. The structure of APHVIII146-1 almost fully matches that of wild-type protein. Differences are observed in the activation loop and are unlikely to affect the global conformation of the protein (Fig. 3b).

The phosphorylation site of APHVII (Ser146) is an homologue of phosphoserine in the ribose pocket of PKA-type serine/threonine kinases [47]. A molecular dynamics simulation of unphosphorylated APHVIII in complex with kanamycin (with bound ATP and two Mg²⁺ ions) revealed pronounced alterations of the enzyme's structure, such as weakening of the contact between the N- and C-terminal domains. [48, 49].

Modification of Ser146 region of aminoglycoside phosphotransferase APHVII – the site of phosphorylation by Pk25 kinase

The autophosphorylation site in the Pk25 activation loop was determined through a comparison with the corresponding region of *M. tuberculosis* PknB [50]:

PknB DFGI ARAIAD SGNSVTQTAAVGTAQYLSPE
 Pk25 DFGV AQVAGA TTLTESGSFVGSPEYTAPE

To optimize the test system *E. coli* APHVIII/Pk25, we modified the potential phosphorylation site AVAEGS₁₄₆VDLED in the APHVIII activation loop. The objective was to make site Ser146 APHVIII more structurally similar to the Pk25 autophosphorylation site TTLTESGSFVG. To achieve that, we modified the amino acid residues in the vicinity of APHVIII Ser146 (Table 2), introducing the underlined amino acid substitutions. The obtained mutant variants of the gene *aphVIII146-1*, *aphVIII146-2*, *aphVIII146-3*, and *aphVIII146-4* were ligated into the pET16b vector at the NdeI-BamHI and introduced into *E. coli* DH5a. After resequencing, the plasmids pET16baphVIII m1 (Fig. 4a) were used for the transformation of *E. coli* BL21(DE3).

Expression of all variants of APHVIII in *E. coli* was checked by gel electrophoresis of soluble cellular proteins under denaturing conditions.

E. coli cells containing pET16baphVIII plasmid expressed a 31 kDa protein, which corresponded to the calculated molecular mass of APHVIII equal to 31.5 kDa.

Creation of a construct containing genes of aminoglycoside phosphotransferase *aphVII* and protein kinase *pk25*

Amplification of *pk25* was performed with the primers Pk25NBgl and Pk25CBgl (Table 1) from the DNA of the plasmid vector pET22b-*pk25*. The Pk25NBgl primer contained a ribosome-binding site (RBS) and an ATG codone for the catalytic domain of protein kinase Pk25. The nucleotide sequence of *pk25* was amplified and digested by BglII and then ligated into the pET16baphVIII m1 containing the previously described variants of *aphVIII* at the BamHI. At the first stage, the presence of the insert was checked by amplification with T7 primers. At the second stage, the clones were selected by amplification with AphN and Pk25CC primers (table 1) according to the presence and the size of the insert. After resequencing, pET16APC plasmids (Fig. 4b) were used for the transformation of *E. coli* BL21(DE3). Expression of genes *aphVIII* and *pk25* after induction was checked by gel electrophoresis. In the first four variants (Fig. 5) (lanes 2-5), additional fractions of APHVIII are observed. Lane 6 contains an additional fraction of *pk25*, while lanes 7-10 contain both a 31.5 kDa fraction and a 28 kDa fraction, which are APHVIII and Pk25, respectively. Mass-spectrometry was used to verify the protein fraction on lane 7. It was shown that the heavier fraction contains APHVIII (AA Y27879.1 aminoglycoside-O-phosphotransferase VIII). The other fraction contains the catalytic domain of Pk25 (1100219

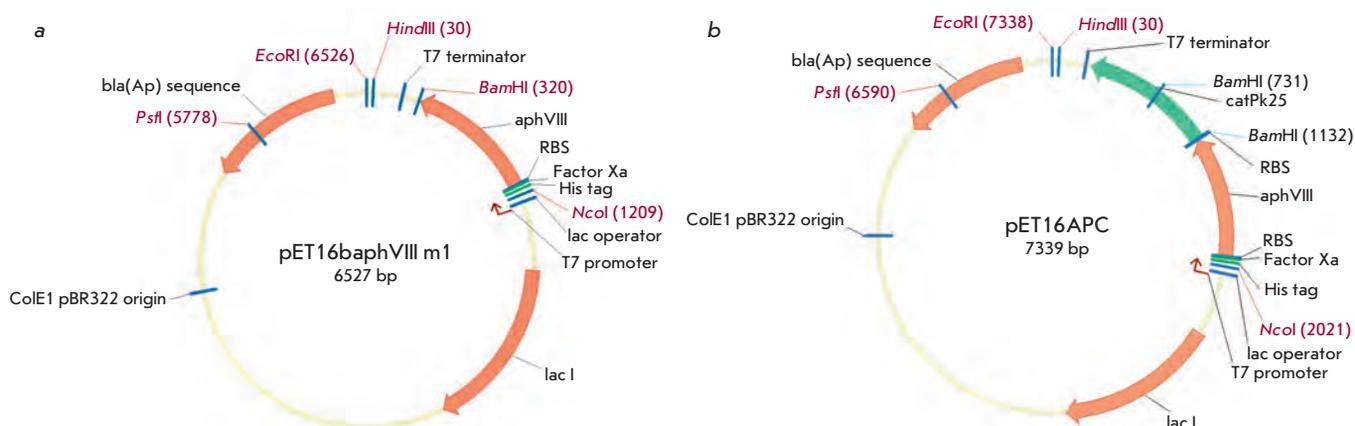


Fig 4. Vectors: a - pET16baphVIII m1 containing *aphVIII* mutant variants: *aphVIII146-1*, *aphVIII146-2*, *aphVIII 146-3*, *aphVIII146-4* and primary *aphVIII146-S*. b - pET16APC containing *pk25* and *aphVIII* mutant variants *aphVIII146-1*, *aphVIII146-2*, *aphVIII146-3*, and primary *aphVIII146-S*.

NP_628936.1 serine/threonine protein kinase *S. coelicolor* A3(2)).

Analysis of kanamycin resistance in *E. coli* BL21(DE3) variants containing different modifications of *aphVIII* and their combinations with *pk25*.

Resistance of all constructs to aminoglycoside antibiotic kanamycin was studied (Table 3). The BL21(DE3) strain containing the plasmid pET16b encoding gene *aphVIII* was resistant to kanamycin (325 µg/ml).

The substitutions contained in the APHVIII146-1 variant led to a 48% decrease in resistance. In APH-

VIII146-2, resistance underwent a 54% decrease. The level of resistance in APHVIII146-3 containing the Ser(146)Thr substitution remained unchanged. In the case of the Ser(146)Ala substitution (APHVIII146-4), which causes the full inactivation of phosphorylation at Ser146, a 70% decrease in kanamycin-resistance was observed. The level of activity of APHVIII146-4 *in vitro* corresponds to the obtained data - the mutant variants exhibited a level of activity equal to 30% of that in wild type [51]. All constructs containing APHVIII with Pk25 showed a higher level of activity. We observed a 91% increase in kanamycin resistance in the case of APHVIII146-1/Pk25, an 83% increase in the case of APHVIII146-2/Pk25, and a 23% increase in the case of the initial construct APHVIII/Pk25, as well as in that with the Ser(146)Thr substitution.

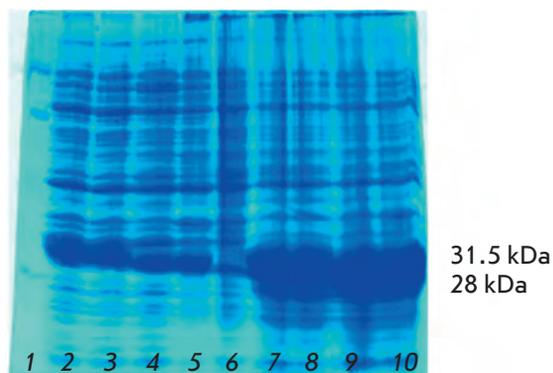


Fig 5. Electrophoresis of *E. coli* BL21(DE3) proteins including Pk25 catalytic domain, APHVIII: 1 – marker, 2 – primary APHVIII146-S fraction, 3 - APHVIII146-1 fraction, 4 – APHVIII146-2, 5 – APHVIII146-3, 6 – Pk25 catalytic domain, 7 – APHVIII146-S /Pk25 fractions, 8 – APHVIII146-1 /Pk25, 9 – APHVIII146-2 /Pk25, 10 – APHVIII146-3 /Pk25. *E. coli* BL21 (DE3) pET22b/pk25 and *E. coli* BL21 (DE3) pET16baphVIII protein fractions analyzed as a negative control.

Docking of indolylmaleimide inhibitors into the Pk25 model

The constructed test system *S. lividans*APHVIII+ has been used earlier for the screening of various chemical substances such as benzodiazines, benzophthalazines, cyclopentendions, indolylmaleimides, pyrazoles, thiazoles, thiazoltetrazines, etc. (unpublished data). A number of indolylmaleimide compounds that exhibit inhibitory activity towards protein kinases were identified. In order to propose the binding mechanisms of these substances, we carried out a molecular docking study. The results suggest that the inhibitors selected in the *S. lividans* TK24 (66) APHVIII/STPK test system (LCTA-1385, LCTA-1398, LCTA-1425 [20], bis-indolylmaleimide-1 [52]) can potentially interact with the ATP binding site of Pk25.

Docking of these inhibitors into the Pk25 model reveals conservative interactions between the maleimide

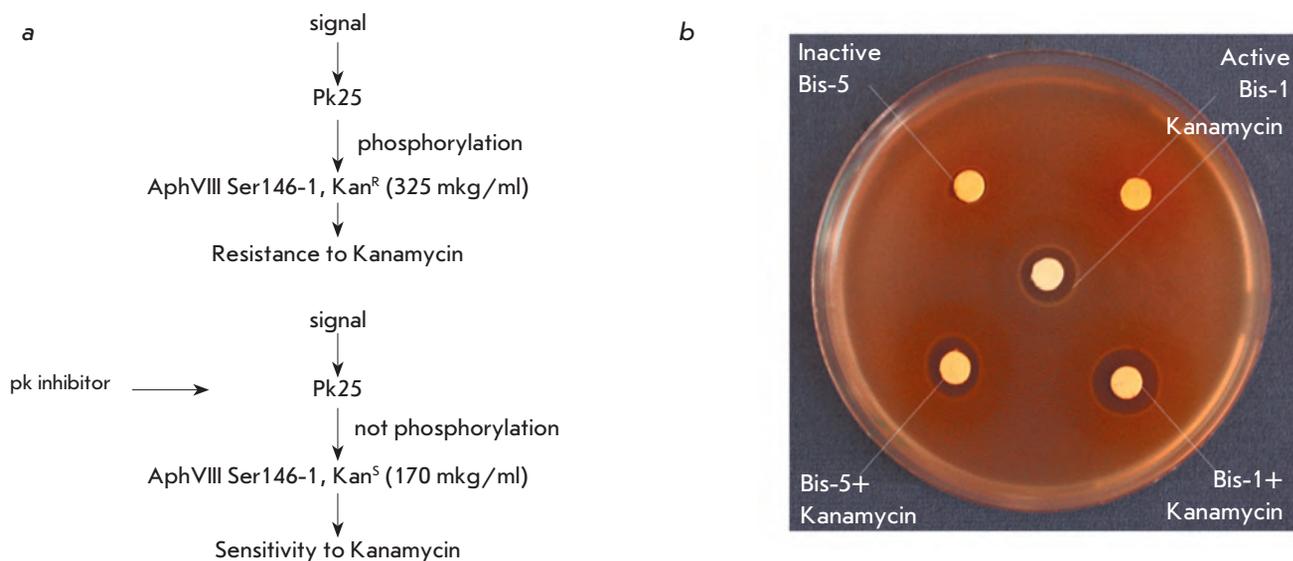


Fig. 6. Bacterial test-system *E. coli aphVIII /Pk25* for screening of inhibitors of serine-threonine protein kinases. **a** – The principle of the test-system: phosphorylation of Ser-146 Pk25 in APHVIII leads to kanamycin-resistance in *E.coli*; addition of inhibitor prevents phosphorylation and reduces kanamycin resistance. **b** – Validation of the test-system with application of Bis-1 and Bis-5 as classical inhibitors [52]: addition of Bis-1 leads to the increase in the zone's size.

moiety and the backbone of the kinase (Fig. 2c). The inhibitors bis-indolylmaleimide-1 (Bis-I) and LCTA-1425 form two hydrogen bonds – one between the carbonyl oxygen atom of the maleimide moiety and the amide hydrogen of Val96, and another one between the imide hydrogen atom of the maleimide moiety and the carbonyl oxygen of Glu94. Inhibitors LCTA-1385 and LCTA-1398 were shown to form only the former hydrogen bond, since the hydrogen atom of the maleimide moiety is substituted in their molecules. Therefore, the estimated energy of interaction between Pk25 and the latter two inhibitors is 1 kcal/mol less than that for the first two compounds. However, in both cases the interaction between the kinase and inhibitor is considered favorable.

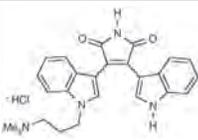
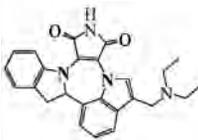
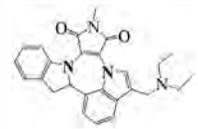
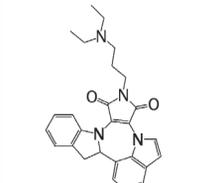
Choice and validation of *E.coli* APHVIII/Pk25-based test system

Earlier we constructed and validated [20, 53, 54] a test system based on *S. lividans* TK24 (66) APHVIII/STPK [20]. The effect registered in this system was based on the cumulative action of the antibiotic kanamycin and a STPK modulator in a sub-inhibiting concentration [20] that resulted in the appearance of or increase in a zone of growth inhibition of the indicator culture. The size of the no-growth zone allowed to roughly estimate the efficiency of the STPK inhibitor [20]. That is why the range of changes of the resistance level to kanamycin, which is determined by various constructions of APHVIII, is of crucial importance. Based on this consideration, APHVIII146-1 is more favorable; so, in fur-

Table 3. Kanamycin resistance of *E. coli* BL21(DE3) strain containing various APHVIII species.

Name	Modified constructs APHVIII	Kanamycin resistance of <i>E. coli</i> , , µg/ml	
		APHVIII	APHVIII+Pk25
146-S	AVAEGS ₁₄₆ VDLED	325±5	400±10
146-1	AVATGT ₁₄₆ VSLED	170±10	325±5
146-2	AVATGS ₁₄₆ VSLSD	150±10	275±10
146-3	AVAEGT ₁₄₆ VDLED	325±5	400±10
146-4	AVAEGA ₁₄₆ VDLED	100±5	-

Table 4. Dependence of *E. coli* APHVIII146-1/Pk25 kanamycin resistance on various STPK inhibitors.

Known indolymaleimide inhibitors of STPKs.	Structure	Subinhibiting concentration, nmol/disc	Inhibition zone in <i>E. coli</i> APHVIII146-1/Pk25 test-system In the presence of kanamycin and inhibitor, mm
Bis-1		700	13.0
LCTA-1425		125	12.0
LCTA-1398		250	13.0
LCTA-1385		125	12.0

Note. Kanamycin-caused inhibition (5 mg/disc) zone size is 10 mm. Kanamycin resistance of *E. coli* strain containing APHVIII146-1 and Pk25 was determined by means of the paper disc method as described in "Experimental procedures". Indolymaleimides of LCTA series, provided by Prof. M.N. Preobrazhenskaya, were described earlier [20].

ther studies we used *E. coli* APHVIII146-1/Pk25 cells. For test system validation, we employed the previously described indolymaleimide STPK inhibitors LCTA-1385, LCTA-1398, LCTA-1425 [20], and Bis-1 [52] (Table 4). The standard concentration of kanamycin was 5 mg/disc, causing the appearance of a 10-mm zone of growth inhibition. All investigated compounds lowered kanamycin resistance. Substances from the indolymaleimide library (LCTA-1033, LCTA-1196, Bis-5) that did not exhibit inhibitory activity in *S. lividans* TK24 (66) APHVIII/STPK [20] showed no effect in the *E. coli* APHVIII146-1/Pk25-based test system that confirms the relevance of the latter.

DISCUSSION

Perspectives of application of *E. coli* APHVIII146-1/Pk25-based test system in screening for STPK inhibitors

The proposed test system can be used in prescreening for ATP-competitive low molecular weight inhibitors [2] that are able to diffuse in an agar-based medium, permeate through the *E. coli* cell wall, and interact with the adenine binding site of Pk25. The selectivity of the inhibitors depends on their affinity to adenine binding

pocket of the kinase. If functional similarity of the amino acid sequences leads to the similarity of the three-dimensional structures. This is also applicable in the case of the ligand binding sites that are responsible for the selectivity of inhibitors [55, 56].

The alignment of the amino acid sequence of the Pk25 catalytic domain with the catalytic domains of other bacterial STPKs, including those from pathogenic microorganisms performed in Genomic BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), revealed 13 proteins that share more than 35% identity. Among them are STPKs from *Mycobacterium*, *Staphylococcus*, *Streptococcus*, and *Pseudomonas*. Comparison of Pk25 with human STPKs revealed 19 proteins with more than 30% identity, including kinases from the SAD, BR, NUAk (SNF), DAPK3, PNCK, CAMKII, CAMKI, Zip, PKA, HUNK, PAK2, Mark-PAR1, SIK2, and OPK NimA families.

STPK inhibitors compete with ATP for the binding site and interact with the adenine binding pocket of the ATP binding site, which contains conservative and variable amino acid residues. We classified STPKs of gram-positive bacteria [23, 57] based on the physico-chemical properties of the sidechains of 9 variable amino acids in

Table 5. Ligand-binding amino acid residues of adenine-binding pocket of bacterial and human serine-threonine protein kinases.

Protein kinase	Functions of kinases in pathogenesis and physiology	Amino acids of the binding pocket	Effect of substitution
Bacterial STPKs			
Pk25 <i>S. coelicolor</i>	Modulation of resistance of aph-gene	LVA <u>V</u> MLVL T	Original
PknA <i>M. tuberculosis</i>	Synthesis of cell wall	<u>I</u> VA <u>I</u> MLVL T	Non-essential
PknJ <i>M. tuberculosis</i>	Persistence in the host	LVA <u>I</u> MFVL <u>S</u>	»
Stk1 <i>Streptococcus agalactiae</i>	Gerulation of intracellular segregation of GBS and virulence	<u>I</u> VA <u>I</u> MYVL T	»
StkP <i>S. pneumonia</i>	A fragment of signal pathway involved in lung and blood invasion	<u>I</u> VA <u>I</u> MYVL T	»
SP-STK <i>S. pyogenes</i>	Cell division, colony morphology, virulence	<u>I</u> VA <u>I</u> MYVL T	»
PpkA <i>P. aeruginosa</i>	Regulation of expression of virulence factors	LVA <u>T</u> MYLL <u>S</u>	Essential
PknB/Stk1 <i>Staphylococcus aureus</i> subsp. <i>aureus</i>	Regulation of purine biosynthesis, autolysis, core metabolism pathways	LVA <u>I</u> MY <u>I</u> L <u>F</u>	»
PknB <i>M. tuberculosis</i>	Cell division, inhibition of lysosome fusion	LVA <u>I</u> MY <u>V</u> M <u>M</u>	»
<i>Homo sapiens</i> STPK			
PKA <i>H. sapiens</i>	Allergy, myocardial disorders	LVA <u>L</u> MYVL T	Non-essential
CaMK ID <i>H. sapiens</i>	Type II diabetes	LVA <u>I</u> MLVL <u>S</u>	»
Pac2 <i>H. sapiens</i>	UV-caused epidermis diseases	<u>I</u> VA <u>I</u> MYLL T	»
BR kin1 <i>H. sapiens</i>	Regulation of cellular homeostasis	LVA <u>V</u> <u>L</u> HVL <u>A</u>	Essential
NUAK SNF1-11 <i>H. sapiens</i>	Modulation of TNF-alpha in cancer cells	LVA <u>I</u> MY <u>A</u> L <u>A</u>	»
CaMKII <i>H. sapiens</i>	Induction of long-termed synaptic memory	LVA <u>I</u> MY <u>A</u> L <u>A</u>	»

Note. Table represents serinen-threonine kinases containing no more than four amino acid substitutions in ligand-binding sequence LVAVMLVLT Pk25. Non-essential amino-acid substitutions are marked (–), while essential substitutions are marked (=), substitutions in the gatekeeper region are marked (–). Selectivity of inhibitors is determined by their affinity to the 9 amino acid motifs of the adenin-binding pocket of STPKs.

the adenine binding pocket of the catalytic domain. In this work, we used this classification to select STPKs of pathogenic bacteria that can be inhibited by the compounds selected in the proposed *E. coli* APHVIII/Pk25 test system. Table 5 contains 9 of the 13 potential ligand binding motifs of STPKs from pathogenic bacteria and 6 of the 19 motifs of human STPKs. The selection was based on the presence of no more than 4 out of 9 amino acid substitutions in variable positions of the adenine-binding pocket of Pk25 *S. lividans* (*S. coelicolor*) LVAVMLVLT. Substitutions of nonpolar amino acids by polar ones at the first four positions, as well as in position

8 (double underlined), were considered essential. All substitutions (except for the introduction of a similar amino acid) in position 9 (double underlined), and any substitutions in position 5 gatekeeper (underlined), were also considered essential. Substitutions in the hinge region (position 7 and in position 6) are less essential. Nonessential substitutions were underlined. The presence of two out of four nonessential substitutions does not alter the interaction mode of inhibitors with the adenine binding site of the protein kinase. Therefore, Pk25 can serve as a tool for the selection of inhibitors for 5 of the 13 STPKs from pathogenic bacteria and 3 of the 19 human STPKs.

The principle of structural similarity of the adenine binding pockets is a more reliable criterion than homology of the sequences of the whole catalytic domain.

In conclusion, the proposed test system can be used in prescreening for inhibitors of STPKs from some pathogenic microorganisms such as PknA, PknJ from

M. tuberculosis, StkP from *S. pneumonia*, SP-STK from *S. pyogenes*, as well as some human STPKs, e.g. PKA, CaMkinase1, and Pac2. ●

This work was supported by the Russian Foundation for Basic Research (grant № 09-04-12025).

REFERENCES

- Manning G., Whyte D.B., Martinez R., et al. // *Science*. 2002. V. 298. P. 1912–1934.
- Cohen P. // *Nat. Rev. Drug Discov.* 2002. V. 1. P. 309–315.
- Levitzi A. // *Acc. Chem. Res.* 2003. V. 36. P. 462–469.
- D'Alessandris C., Lauro R., Presta I., Sesti G. // *Diabetologia*. 2007. V. 50(4). P. 840–849.
- Barbier E., Zapata A., Oh E., et al. // *Neuropsychopharmacology*. 2007. V. 32(8). P. 1774–1782.
- Shirai H., Autieri M., Eguchi S. // *Curr. Opin Nephrol. Hypertens.* 2007. V. 16(2). P. 111–115.
- Collins I., Workman P. // *Nat. Chem. Biol.* 2006. V. 2. P. 689–700.
- Ishida A., Kameshita I., Sueyoshi N., et al. // *J. Pharmacol. Sci.* 2007. V. 103(1). P. 5–11.
- Goldstein D.M., Gray N.S., Zarrinkar P.P. // *Nat. Rev. Drug Discov.* 2008. V. 7. P. 391–397.
- Bain J., Plater L., Elliott M., et al. // *Biochem. J.* 2007. V. 408. P. 297–315.
- Liu Y., Gray N.S. // *Nat. Chem. Biol.* 2006. V. 2. P. 358–364.
- Shi L., Potts M., Kennelly P.J. // *FEMS Microbiol. Rev.* 1998. V. 22. P. 229–253.
- Echenique J., Kadioglu A., Romao S., et al. // *Infect. Immun.* 2004. V. 72(4). P. 2434–2437.
- Cozzone A.J. // *J. Mol. Microbiol. Biotechnol.* 2005. V. 9. P. 198–213.
- Pérez J., Garcia R., Bach H., et al. // *Biochem. Biophys. Res. Commun.* 2006. V. 348(1). P. 6–12.
- Drews S.J., Hung F., Av-Gay Y. // *FEMS Microbiol. Lett.* 2001. V. 205(2). P. 369–374.
- Saini D.K., Tyagi J.S. // *J. Biomol. Screen.* 2005. V. 10(3). P. 215–224.
- Wehenkel A., Bellinzoni M., Graña M., et al. // *Biochim. Biophys. Acta*. 2008. V. 1784(1). P. 193–202.
- Wázquez F., Szabadkai I., Németh G., et al. // *Immunol. Lett.* 2008. V. 116(2). P. 225–231.
- Danilenko V.N., Simonov A.Y., Lakatosh S.A., et al. // *J. Med. Chem.* 2008. V. 51. P. 7731–7736.
- Elizarov, S. M., Sergienko, O. V., Sizova, I. A., et al. // *Mol. Biologia (Molecular Biology)*, 2005. V. 35. P. 226–233.
- Bekker, O. B., Elizarov, S. M., Alekseeva, M. T., et al. // *Mikrobiologia (Microbiology)*. 2008. V. 75 (5). P. 559–567.
- Danilenko V.N., Osolodkin D.I., Lakatosh S.A., et al. // *Current Top. Med. Chemistry*. 2010. *In press*.
- Kieser T., Bibb M.J., Buttner M.J., et al. *Practical Streptomyces Genetics*. Norwich, John Innes Foundation. England, 2000. P. 613.
- Sambrook J., Fritsch E.F., Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor, N.Y.; Cold Spring Harbor Lab. Press, 1989.
- Kameshita I., Fujisawa H. // *Anal. Biochem.* 1989. V. 183. P. 139–143.
- Nelson R.M., Long G.L. // *Anal. Biochem.* 1989. V. 180. P. 147–151.
- Berman H.M., Westbrook J., Feng Z., et al. // *Nucleic Acids Research*. 2000. V. 28. P. 235–242.
- Young T.A., Delagoutte B., Endrizzi J.A., et al. // *Nat. Struct. Biol.* 2003. V. 10. P. 168–174.
- Ortiz-Lombardía M., Pompeo F., Boitel B., Alzari P.M. // *J. Biol. Chem.* 2003. V. 278. P. 13094–13100.
- Wehenkel A., Fernandez P., Bellinzoni M., et al. // *FEBS Lett.* 2006. V. 580. P. 3018–3022.
- Mieczkowski C., Iavarone A.T., Alber T. // *EMBO J.* 2008. V. 27. P. 3186–3197.
- Bentley S.D., Chater K.F., Cerdeno-Tarraga, et al. // *Nature*. 2002. V. 417. P. 141–147.
- Larkin M.A., Blackshields G., Brown N.P. // *Bioinformatics*. 2007. V. 23. P. 2947–2948.
- Šali A., Blundell T.L. // *J. Mol. Biol.* 1993. V. 234. P. 779–815.
- Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. // *J. Appl. Cryst.* 1993. V. 26. P. 283–291.
- SYBYL 8.0, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- Fong D.H., Berghuis A.M. // *EMBO J.* 2002. V. 21. P. 2323–2331.
- Nurizzo D., Shewry S.C., Perlin M.H., et al. // *J. Mol. Biol.* 2003. V. 327. P. 491–506.
- Huey R., Morris G.M., Olson A.J., Goodsell D.S. // *J. Comput. Chem.* 2007. V. 28. P. 1145–1152.
- Halgren T.A. // *J. Am. Chem. Soc.* 1990. V. 112. P. 4710–4723.
- Sanner M.F. // *J. Mol. Graphics Mod.* 1999. V. 17. P. 57–61.
- Umeyama T., Horinouchi S. // *Bacteriol.* 2001. V. 183. P. 5506–5512.
- Petrickova K., Tichy P., Petricek M. // *Biochem. Biophys. Res. Commun.* 2000. V. 279(3). P. 942–948.
- Scherr N., Müller P., Perisa D., et al. // *J. Bacteriol.* 2009. V. 191(14). P. 4546–4554.
- Villarino A., Duran R., Wehenkel A. // *J. Mol. Biol.* 2005. V. 350. P. 953–963.
- Taylor S.S., Yang J., Wu J., et al. // *Biochim. Biophys. Acta*. 2004. V. 1697(1–2). P. 259–269.
- Scheeff E.D., Bourne P.E. // *PloS Comput. Biol.* 2005. V. 1. P. 0359–0381.
- Nurizzo D., Shewry S.C., Perlin M.H., et al. // *J. Mol. Biol.* 2003. V. 327. P. 491–506.
- Durán R., Villarino A., Bellinzoni M. // *Biochem. Biophys. Res. Commun.* 2005. V. 333(3). P. 858–867.
- Alekseeva M., Elizarov S. // *Congress EurasiaBio-2010*, 13–15 apr. 2010, Moscow.
- Brehmer D., Godl K., Zech B., et al. // *Mol. Cell Proteomics*. 2004. V. 3(5). P. 490–500.
- Konings E.J.M., Roomans H.H.S. // *Food Chem.* 1997. V. 59. P. 599–603.
- Popov A.Y. Validation – what, where, when? // *Chistie pomeschenia i technologicheskie sredi (Clean placement and technological mediums)*, 2003. №3.
- Scherr N., Honnappa S., Kunz G., et al. // *Proc. Natl. Acad. Sci. USA*. 2007. V. 104. P. 12151–12156.
- Zakharevich N.V., Osolodkin D.I., Artamonova I.I., et al. // *Congress EurasiaBio-2010*, 13–15 apr. 2010, Moscow.

Individual Genome of the Russian Male: SNP Calling and a *de novo* Assembly of Unmapped Reads

N. N. Chekanov², E. S. Boulygina¹, A. V. Beletskiy², E. B. Prokhortchouk^{1,2##}, K. G. Skryabin^{1,2##}

¹ Russian Scientific Centre Kurchatov Institute

² Bioengineering Center of the Russian Academy of Sciences

Authors contributed equally to this work (in alphabetical order)

* E-mail: prokhortchouk@biengi.ac.ru

Received 03.09. 2010

ABSTRACT A somatic cell genome was recently resequenced for a patient with renal cancer. The data were submitted to the NCBI Sequence Read Archive under the accession number SRA012240. Here, we have performed SNP calling for the genome and compared it with several published genomes. We have found 2,921,724 SNPs, including 1,472,679 newly described ones. Among them, 63,462 SNPs have been mapped to the Y chromosome and, based on 18 markers, the genome has been ascribed to the R1a1a haplogroup predominant in Russian males. The mitochondrial haplogroup has been determined as U5a, which is also common in the European part of Russia. Short reads unmapped to the human genome were used for the *de novo* assembly of DNA sequences. This resulted in genome-specific contigs (more than 100 bp in length) with an overall length of 154 kbp (for GAI) and 4.7 kbp (for SOLiD).

KEYWORDS human genome, sequencing platform, single-nucleotide polymorphism, bioinformatics

ABBREVIATIONS SNP – single-nucleotide polymorphism, RCS – reconstructed consensus sequence

INTRODUCTION

The implementation of modern sequencing platforms has allowed widely accessible sequencing of individual genomes. In August 2010, the 1000 Genomes project [1] published (at <http://www.1000genomes.org/>) preliminary data on the resequencing of 2,500 individual genomes from various ethnic groups. A detailed report is expected. The general purpose of these studies is to identify frequent (with a frequency of more than 1% of the population) genome variations in human populations. Apart from fundamental problems of population genetics, the medical aspect of these studies is obvious. For example, at the end of 2009, the International Cancer Genome Consortium (ICGC) was established to investigate tumor-cell genomes [2]. Russia is affiliated with this consortium through the Russian Research Centre Kurchatov Institute, the Bioengineering Center of the Russian Academy of Sciences, and the Blokhin Cancer Research Center of the Russian Academy of Medical Sciences, which are involved in studies on renal cancer-cell genomes. The first successful resequencing of the human genome in Russia was done at the end of 2009 [3]. Libraries of short DNA reads were obtained from the genome of patient N, a Russian man suffering from renal cancer, using two sequencing platforms (SOLiD and GAI). Thus, the first genome from the Slavic population, which was never been present in the

population sampling of the 1000 Genomes project, was resequenced. On the other hand, it was the first step within the framework of the renal cancer-cell genome sequencing project.

In this study we have performed a bioinformatics analysis of the data on patient N's genome resequencing directed at SNP calling. In addition, we have assembled long DNA contigs specific to patient N.

MATERIALS AND METHODS

SNP calling

Short DNA sequences that had been read on a GAI sequencer were mapped using a SOAPaligner/soap2 v.2.20 alignment program [4] with default parameters; except for the paired-end reads' insert size. The acceptable insert size range was specified as 100–700 nucleotides, based on previous data [3]. Then, SNPs were identified using the SOAPsnp v.1.02 resequencing utility [5] with default parameters. The short DNA sequences that were read on a SOLiD sequencer were mapped using a Bowtie build 0.12.5 short-read aligner [6] in a quality-aware colorspace, specifying the max mismatches in the seed as two. The acceptable insert size range was specified as 600–1,400 nucleotides, which is also in accordance with the previous data [3]. SNP calling was carried out with a SAMtools 0.1.7 package [7] using only the uniquely mapped reads.

Determination of mitochondrial and Y-chromosomal haplogroups

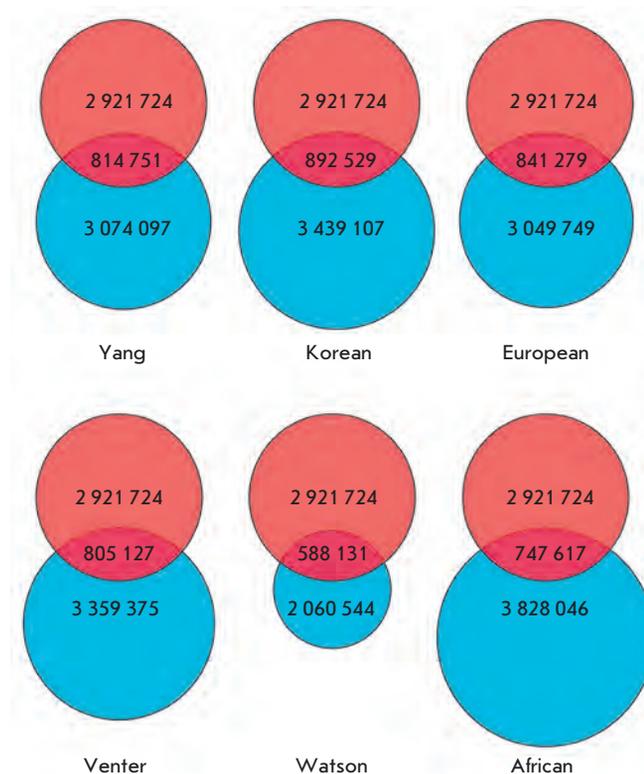
To determine the mitochondrial haplogroup, we used reads that were obtained using the SOLiD sequencer and processed with a Corona Lite package [3]. The list of mitochondrial genome SNPs, with coordinates and allele values, was acquired from the PhyloTree database (updated in August, 2010; <http://www.phylotree.org/>). In ascent to the mitochondrial haplogroup phylogenetic tree taken from it, we determined the allele of each distinct SNP as follows: (1) we found an allele by the specified coordinates in the RCS of mitochondrial genome, and (2) we verified these coordinates by comparing flanking sequences (no less than 10 bp from each end).

The haplogroup of the Y chromosome was determined from the reads obtained on both the GAI and SOLiD platforms and processed using the Illumina Genome Analyzer Pipeline and Corona Lite program packages, respectively [3]. The SNP list for the Y chromosome was acquired from the website <http://isogg.org/> (updated in August 2010), excluding the markers that were absent in dbSNP. In ascent to the Y chromosome haplogroup tree, which was also taken from the site mentioned above, we determined the allele of each distinct SNP as follows: (1) We identified the allele in mapped nucleotide sequences from the GAI library by the coordinates of that SNP in the hg18 reference genome specified in dbSNP and verified these coordinates by comparing flanking sequences (no less than 10 bp from each end or no less than 20 bp from one end). (2) For the data from SOLiD, the allele in the Y chromosome RCS was identified by a comparison with the SNP flanking sequences acquired from dbSNP, if their size was no less than 100 bp, and RCS coverage by reads had no more than 50% gaps. The ancestral status of alleles was determined by SNP description in dbSNP.

De novo reconstruction of genome texts

We chose those reads primarily from both platforms which were not mapped to the human genome (hg18, excluding unmapped sites). The number of these sequences was 291.57 and 628.86 million for GAI and SOLiD, respectively. They were used as input data for the ABySS v.1.1.0 short read assembler [8], which offers a distributed implementation of the de Bruijn graph for the search for overlaps between k-mers (sequences whose length is k). ABySS was started several times for the optimization of the k-mer length. The optimum length of k-mer providing the longest contigs (≥ 200 bp) was 23 for the data from GAI and 16 from SOLiD.

Then, the sequences obtained *de novo* were mapped to the reference human genomes GRCh37 (hg19), Celera, and HuRef using the NCBI BLAST v.2.2.23 [9] with the megablast search algorithm and with enabled



Unique and common SNPs in different individual genomes (blue circles) compared with those of patient N (red circles).

filtering of repeats (simple and human-specific). Sequences that were not found in any of these three reference genomes were mapped again, both to the same reference genomes and to the genomes of primates, using the discontinuous megablast search algorithm.

RESULTS AND DISCUSSION

Identification of SNPs in patient N's genome

The data of patient N's genome resequencing, which was obtained using SOLiD and GAI sequencing platforms, are presented as a set of reads at the site of the National Center for Biotechnology Information (NCBI), Acc. No. SRA012240. The data had been statistically processed earlier [3]. Another immediate task of this study was to identify SNP coordinates by comparing all readings mapped to a distinct genome region (SNP calling). SNP calling was carried out separately for GAI and SOLiD data. The allele number was 1,824,006 and 410,383 SNPs, respectively. The data from SOLiD were converted from the colorspace to FASTQ and combined with those from GAI, followed by the repetition of SNP calling. The total number of SNPs (2,921,724) exceeds the sum of SNPs identified in separate analyses

Table 1. Comparative numbers of SNPs found in different individual human genomes and the genome of patient N.

	Venter	Watson	Yang	Korean	European	African
Total SNP number	3359375	2060544	3074097	3439107	3049749	3828046
SNPs in Russian genome	1824006					
Common SNPs	510444	365955	518294	570937	532194	479420
One allele is the same	427096	285913	425024	457469	431977	384934
Both alleles are the same	81957	79797	92752	113042	99967	89402
SNPs in Russian genome, SOLiD	410383					
Common SNPs	179948	141703	187675	204235	192773	178744
One allele is the same	116376	73735	119837	130518	125589	111031
Both alleles are the same	27202	57292	30423	34023	33756	32133
SNPs in Russian genome, SOLiD+GAII	2921724					
Common SNPs	805127	588131	814751	892529	841279	747617
One allele is the same	508066	411521	486809	513621	481542	424153
Both alleles are the same	276881	171052	307802	357562	341765	301925

Note: The data were obtained using two sequencing platforms separately and in combination

Table 2. Allele values of patient N’s mitochondrial DNA known polymorphisms characterizing his affiliation with the haplogroup U5a

Haplogroup	Position	Reference allele (H2)	Diagnostic allele	SOLiD allele
L3	3594	C	C	C
N	10398	A	A	A
N	10400	C	C	C
N	10873	T	T	T
R	12705	C	C	C
UK	12308	A	G	G
U	11467	A	G	G
U5	9477	A	A	A
U5	16270	C	T	T
U5-sub	16399	A	G	G
U5a	14793	A	G	G
U5a	16256	C	T	T

of the data from each platform. This is indicative of the mutual supplementation of these two datasets in the coverage of genome regions. A comparison of allele coordinates and values was performed with the following genomes: Craig Venter [10], James Watson [11], and Huanming Yang [12], as well as genomes of a Korean [13], an African [14], and a European (CEU Trio Father NA12891 from the 1000 genomes project). The data are shown in Table 1. A comprehensive datasheet of coordinates and allele values of SNPs is shown on the site <http://www.russiangenome.ru/>. The figure summarizes the number of common and unique SNPs found in patient N’s genome and the genomes of other individuals. We found no correlation between the resemblance of one or two equal SNP alleles (see Table. 1, rows “one

allele is the same” and “both alleles are the same”) and the distance between the nominal habitat of the corresponding person and Moscow, which is taken as the nominal habitat of Russians (Venter and Watson are considered Western Europeans). However, the Principal component analysis arranged individuals in accordance with the distance between their birthplaces (data not shown). The correlation is 0.89 at $p\text{-value} = 10^{-5}$.

Determination of mitochondrial and Y-chromosomal haplogroups of patient N

The identified coordinates and allele values of SNPs have made it possible to determine the mitochondrial and Y-chromosomal haplogroups of patient N’s genome. Initially, we collected all reads obtained from

Table 3. Allele values of patient N's Y-chromosomal SNPs characterizing his affiliation with the haplogroup R1a1a

Haplogroup	SNP	GA allele	SOLiD allele	Ancestral allele
R	rs2032658	N/A	G	A
R	rs17307398	T	T	C
R	rs4481791	C	N/A	G
R	rs9786261	N/A	A	G
R	rs891407	G	G	C
R1	rs17307070	N/A	T	G
R1	rs9786232	G	G	T
R1	rs9785959	G	N/A	C
R1	rs9786197	N/A	C	T
R1	rs7067478	A	N/A	G
R1a	rs17222573	N/A	G	A
R1a	rs17307677	N/A	C	T
R1a	rs17306692	A	N/A	C
R1a1	rs17222202	N/A	A	T
R1a1	rs17316227	N/A	G	A
R1a1	rs2534636	N/A	T	T*
R1a1a	rs17222146	N/A	T	C
R1a1a	rs17315926	T	T	C
R1a1a	rs17221601	N/A	A	T

Note: The markers found using both sequencing platforms are drawn in bold. *rs2534636 is the back mutation for the haplogroup R1a1.

Table 4. Summary of the *de novo* reconstructed contigs that were unequivocally attributed to one of three human reference genomes.

	Not found		Found in unplaced genomic contig		Found in unlocalized genomic contig on known chromosome		Found	
	GA	SOLiD	GA	SOLiD	GA	SOLiD	GA	SOLiD
hg19	292	3	31	6	0	15	154	1
Celera	147	10	47	4	0	3	307	0
HuRef	125	9	69	8	0	0	300	0

Table 5. General statistics on *de novo* assembled contigs specific for patient N. The length of the contigs in kilobases is given in parentheses.

	GA	SOLiD
Univocally found in hg19	146 (44.7)	1 (0.3)
Simultaneously found in less than three human reference genomes	93 (27.4)	3 (0.7)
Not found in any human genome	72 (21.3)	0 (0)
Found in genomes of primates	51 (15.4)	2 (0.5)
Of them with homology > 95%	22 (6)	1 (0.2)
Total number of contigs	495 (154)	17 (4.7)

SOLiD and mapped them to the reference mitochondrial DNA (revised Cambridge Reference Sequence (rCRS); Acc. No. in GenBank: NC_012920) [15]. On the basis of these reads, an RCS was constructed and published at <http://www.russiangenome.ru/>. The mean coverage of the mitochondrial genome was 291. A comparison of this RCS with the reference one has shown that the mitochondrial genome of patient N belongs to the U5a haplogroup (Table. 2), one of the most common in European Russia.

The Y-chromosomal haplogroup was determined as R1a1a by four markers identified using both SOLiD and GAI and 19 markers coinciding with the data of one of two sequencing platforms (Table. 3). The coincidence of the SNP allele rs2534636 of patient N with the ancestral allele confirms the haplogroup R1a1, because this polymorphism is considered to be a result of back mutation. Since the Y chromosome is not recombinant, we can expect a high nonequilibrium coupling degree of its genetic markers. Therefore, all 63 462 SNPs identified in this work as belonging to the Y chromosome can implicitly characterize the haplotype of most men born in European Russia because of the prevalence of the R1a1a haplogroup in this region. The datasheet of all Y-chromosomal SNPs is also available at the site of the project.

De novo reconstruction of genome texts specific to patient N

The certain possibility of reconstructing a complete individual genome makes it possible to identify specific sites for a given individual. Despite the current inaccessibility of these data in the framework of the 1000 Genomes project, studies conducted by a group led by Prof. Huanming Yang at the Beijing Genomics Institute have shown that his own genome contains about 7,200 unique contigs covering about 5 million bp [16]. We have reconstructed *de novo* the unique texts of patient N's genome. All collected contigs exceeding 100

nucleotides were divided into two groups: those giving an unequivocal search result in the BLAST program (Table. 4) and those requiring additional analysis (see general statistics in Table. 5). The nucleotide sequences obtained using the SOLiD platform were insignificant both in amount and summary length. In all likelihood, this is because of the impropriety of short 25-nucleotide sequences for the reconstruction of complex genomic texts. Among the contigs collected using the GAI sequencer, the most interesting are the regions with no homology with reference human genomes, as well as those strikingly similar to genomes of primates (which have a slight difference). We can (with some degree of probability) attribute the first group of sequences to possible errors in assembling *de novo* by ABySS; however, the second group of sequences apparently cannot be the assembling errors and are characteristic of patient N. The search for open reading frames in these contigs has not revealed long (more than 30 aminoacids) coding sequences. All contigs assembled *de novo* are available at the website of the project. The difference in the number and length of contigs in the genomes of patient N and Huanming Yang can be explained by the different genome coverage (7 and 30, respectively).

Here we characterize patient N's genome compared with the reported data on other human genomes. To estimate the significance of the polymorphous and unique differences in (1) the formation of ethnic diversity and (2) the predisposition of patient N to various diseases, we need additional data on individual genomes from various ethnic groups, as well as the data obtained in associative studies using both high-density DNA chips and pangenomic sequencing. ●

This study was supported by the Federal Program Development of Russian Nanoindustry Infrastructure for 2008-2012. The authors are grateful to Prof. M.V. Kovalchuk for general assistance and for paying close attention to this work.

REFERENCES

- Siva N. // Nat. Biotechnol. 2008. V. 26(3). P. 256.
- Hudson T.J., Anderson W., Artz A., et al. // Nature. 2010. V. 464(7291). P. 993–998.
- Skryabin K.G., Prokhorchuk E.B., Mazur A.M., et al. // Acta Naturae. 2009. V. 1. №3. P. 102–107.
- Li R., Yu C., Li Y., et al. // Bioinformatics. 2009. V. 25(15). P. 1966–1967.
- Li R., Li Y., Fang X., et al. // Genome Res. 2009. V. 19(6). P. 1124–1132.
- Langmead B., Trapnell C., Pop M., Salzberg S.L. // Genome Biol. 2009. V. 10(3). P. R25.
- Li H., Handsaker B., Wysoker A., et al. // Bioinformatics. 2009. V. 25(16). P. 2078–2079.
- Simpson J.T., Wong K., Jackman S.D., et al. // Genome Res. 2009. V. 19(6). P. 1117–1123.
- Altschul S.F., Gish W., Miller W., et al. // J. Mol. Biol. 1990. V. 215(3). P. 403–410.
- Levy S., Sutton G., Ng P.C., et al. // PLoS Biol. 2007. V. 5(10). P. e254.
- Wheeler D.A., Srinivasan M., Egholm M., et al. // Nature. 2008. V. 452(7189). P. 872–876.
- Wang J., Wang W., Li R., et al. // Nature. 2008. V. 456(7218). P. 60–65.
- Kim J.I., Ju Y.S., Park H., et al. // Nature. 2009. V. 460(7258). P. 1011–1015.
- Bentley D.R., Balasubramanian S., Swerdlow H.P., et al. // Nature. 2008. V. 456(7218). P. 53–59.
- Andrews R.M., Kubacka I., Chinnery P.F., et al. // Nat. Genet. 1999. V. 23(2). P. 147.
- Li R., Li Y., Zheng H., et al. // Nat. Biotechnol. 2010. V. 28(1). P. 57–63.

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.

GUIDELINES FOR AUTHORS

- 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