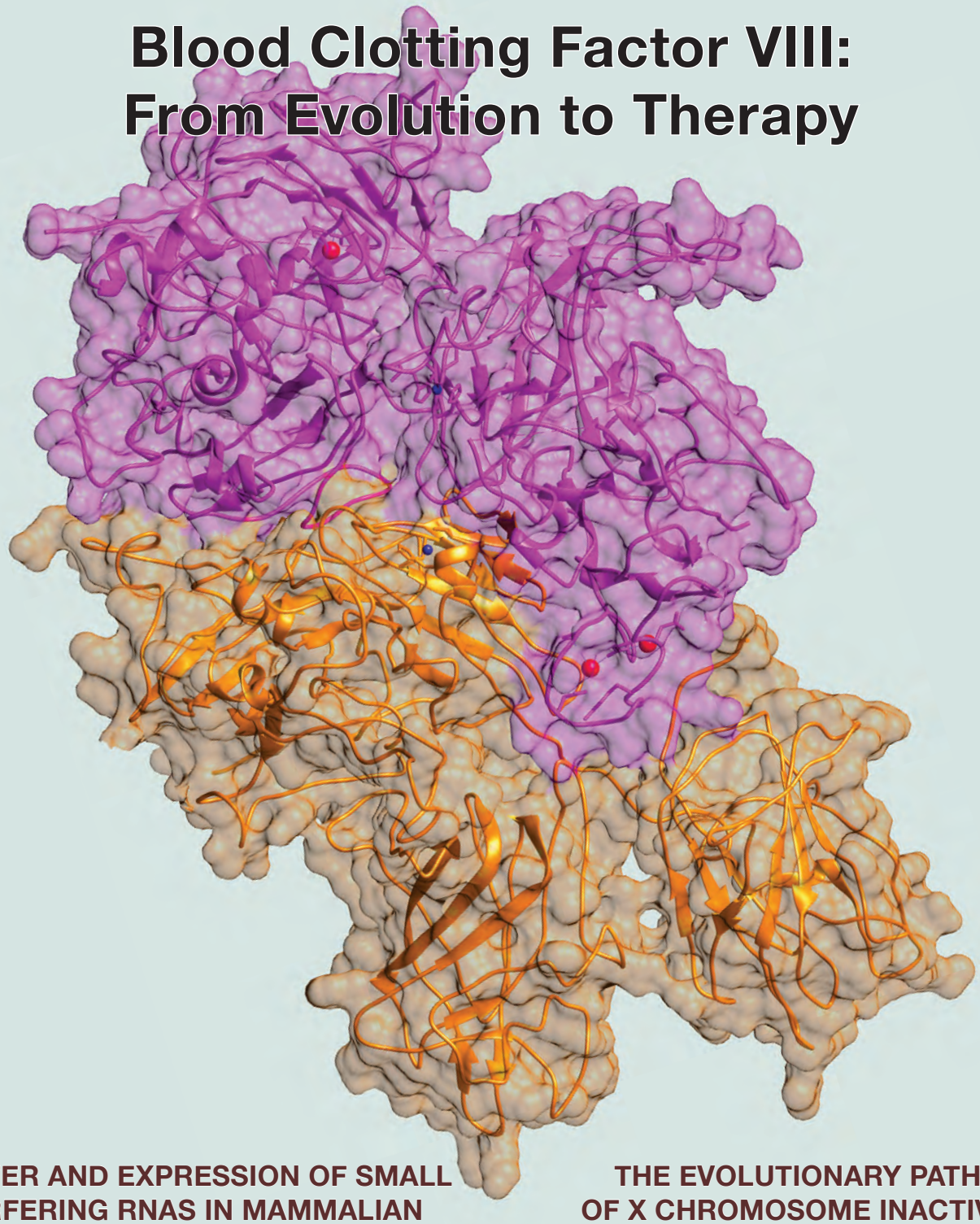
APRIL–JUNE 2013 VOL. 5 № 2 (17)

ISSN 2075-8251

ActaNaturae

Blood Clotting Factor VIII: From Evolution to Therapy



APRIL–JUNE 2013 VOL. 5 № 2 (17)

ActaNaturae

Letter from the Editors

Dear readers of *Acta Naturae*,

We are delighted to bring you the 17th issue of our journal.

The research section of the journal has a conventional structure: it consists of reviews and research articles that embrace a broad range of problems of modern medicine and biology. It is worth mentioning that two publications (A.I. Shevchenko *et al*. and A.V. Panova *et al*.) are devoted to different aspects of X chromosome inactivation. They can be regarded as our "response" to the masculinity crisis that is being widely debated in the mass media.

The review by N.A. Orlova *et al.* is devoted to one of the most successful examples of the application of fundamental science and biotechnology to practical medicine: hemophilia A, an almost incurable disease, can be controlled to a significant extent through the extensive use of recombinant factor VIII. The review by T.D. Lebedev *et al*. focuses on an approach that also has a potential to significantly contribute to medicine.

Three out of five research articles in this issue are also devoted to the scientific foundations of medicine. The range of approaches used is rather broad, which dovetails smoothly into existing trends.

The final article was penned by Ukrainian authors. We are glad to welcome these new authors and hope that our scientific ties, which used to be so strong, will grow stronger again. We are elated that the share of articles written by foreign authors in our journal consistently increases. We hope that this is the result of our efforts to a great extent.

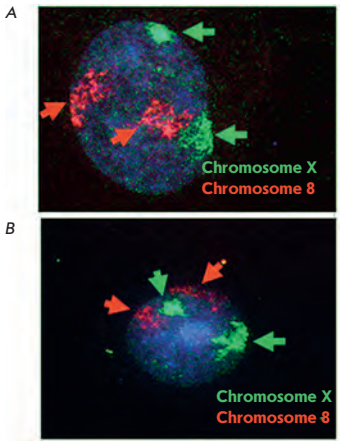
We will keep doing our best.

Sincerely, Editorial Board

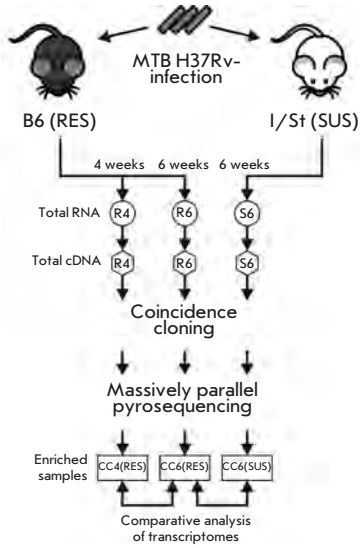
Late Replication of the Inactive X Chromosome Is Independent of the Compactness of Chromosome Territory in Human Pluripotent Stem Cells

A. V. Panova, E. D. Nekrasov, M. A. Lagarkova, S. L. Kiselev, A. N. Bogomazova

It is demonstrated that the inactivated X chromosome may occupy either a compact or dispersed territory in human pluripotent stem cells. Late replication of the inactivated X chromosome does not depend on the compactness of chromosome territory. However, the Xi reactivation and the synchronization in the replication timing of X chromosomes upon reprogramming are necessarily accompanied by the expansion of X chromosome territory.



X chromosome territories in human pluripotent stem cells



Scheme of transcriptome comparison

Mycobacterium tuberculosis Transcriptome Profiling in Mice with Genetically Different Susceptibility to Tuberculosis

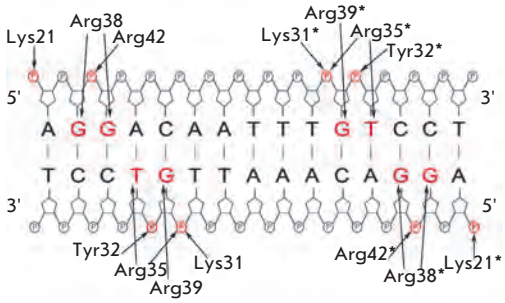
T.A. Skvortsov, D.V. Ignatov, K.B. Majorov, A.S. Apt, T.L.Azhikina

In this study, we describe the results of genome-wide transcriptional profiling of the major human bacterial pathogen *M. tuberculosis* during its persistence in lungs. Two mouse strains differing in their susceptibility to tuberculosis were used for experimental infection with *M. tuberculosis*. It was hypothesized that the changes in the *M. tuberculosis* transcriptome may attest to the activation of the metabolism of lipids and amino acids, transition to anaerobic respiration, and increased expression of the factors modulating the immune response. A total of 209 genes were determined whose expression increased with disease progression in both host strains.

Peculiarities of the Regulation of Gene Expression in the Ecl18kI Restriction–Modification System

O. Yu. Burenina, E. A. Fedotova, A. Yu. Ryazanova, A. S. Protsenko, M. V. Zakharova, A. S. Karyagina, A. S. Solonin, T. S. Oretskaya, E. A. Kubareva

The complex formation of (cytosine-5) DNA-methyltransferase Ecl18kI and RNA polymerase from *Escherichia coli* with the promoter regions of the DNA methyltransferase and restriction endonuclease genes is studied. The mechanism of gene expression in the Ecl18kI restriction–modification system is thoroughly investigated.



Scheme of the contacts between the amino acid residues in the N-terminal domain of M.SsoII dimer and the regulatory site in DNA

ActaNaturae
APRIL—JUNE 2013 VOL. 5 № 2 (17)

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

Address: 119234 Moscow, Russia, Leninskiye Gory, Nauchny
Park MGU, vlad.1, stroeniye 75G.
Phone/Fax: +7 (495) 930 88 50
E-mail: vera.knorre@gmail.com, mmorozova@strf.ru,
actanaturae@gmail.com

Номер подписан в печать 17 мая 2013 г.
Тираж 200 экз. Цена свободная.
Отпечатано в типографии «МЕДИА-ГРАНД»

CONTENTS

RESEARCH ARTICLES

IMAGE ON THE COVER PAGE
3D structure of the FVIII deletion variant

FEDERATION OF EUROPEAN BIOCHEMICAL SOCIETIES CONGRESS 2013
Mechanisms in Biology



Summer is coming, and FEBS-38, this year's biggest scientific forum in the field of life sciences, will be here soon. In the previous issue of *Acta Naturae*, we announced the Congress program, the main plenary lecturers (among whom there are 11 Nobel Prize winners in chemistry and biology), and the themes of 37 symposia. Abstract submission ends on May 1, 2013, so some preliminary results about participant registrations can already be summarized. A total of about 2,800 participants have registered thus far. Delegates from many countries are going to attend the Congress.

Although this will be the European Congress, many representatives of the USA are going to attend (96 participants). There are also researchers from Japan, China, Egypt, and India. Of course, most participants will come from European countries. The United Kingdom, France, and Germany are the leading countries. About 1,000 delegates from Russia are expected to attend the Congress. The program of Congress bursaries (a total of 170 winners) was not supposed to cover the participation expenses of Russian young scientists. However, the Organizing Committee has recently managed to agree with the Ministry of Education and Science of the Russian Federation upon supporting 150–170 young Russian scientists. Registration of this group of delegates is almost complete. As we mentioned previously, the Congress will be immediately preceded by the FEBS Young Scientists' Forum that will be held in the historic building of the Russian Academy of Sciences. The main activities of the Congress will be held in the Lenexpo exhibition complex, the Opening Ceremony will be held in Grand Concert Hall "Oktyabrsky." It is obvious now that the Congress will be accompanied by a rather extensive exhibition of reagents, materials and equipment for research, where scientific publishing companies will also be represented. The Ministry of Industry and Trade of the Russian Federation recently decided to ensure the presence of representatives of Russian institutes and companies that work at implementing the Pharma 2020 program and are ready to share the results of the research conducted within the framework of the program. Thus, there will be a unique opportunity to evaluate the stages of implementation of the program, together with world leaders in the field of biopharmaceuticals and the leading scientists in novel approaches to the therapy of cancer, autoimmune, and neurodegenerative diseases, as well as infections.

The initiative of the Ministry of Education and Science of the Russian Federation for presenting the achievements of the leading higher education institutions at the Congress in order to attract foreign experts and to provide participation of our PhD and undergraduate students in international projects is currently being discussed. But the most stimulating activity will be the participation of Russian young scientists in the Congress. There is a clear trend of inviting young scientists who submitted interesting abstracts to present short (15 min) lectures at the end of the symposium sessions. Public discussions of these lectures will be vivid and memorable events that will be interesting both to the audience and the young scientists. It will be a powerful incentive for the career advancement of a young scientist.

The Organizing Committee has to solve a number of social problems facing the Congress, including the problem of budget accommodation for young scientists. Available dormitories have been found thanks to the effort of the students of St. Petersburg Academic University of the Russian Academy of Sciences. The Organizing Committee will provide this information on the Congress' website. The unique social program, which includes a classical ballet program at the Opening Ceremony after the Nobel Prize winner Jules Hoffmann delivers his lecture and excursion to the Hermitage, will be an incentive for the participants to work productively.

We look forward to welcoming you in Saint Petersburg on July 6–11. The program promises to be very interesting. ●

Transfer and Expression of Small Interfering RNAs in Mammalian Cells Using Lentiviral Vectors

T. D. Lebedev*, P. V. Spirin, V. S. Prassolov
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilova Str., 32, Moscow, Russia, 119991
*E-mail: saint_john@list.ru
Received 19.09.2012

ABSTRACT RNA interference is a convenient tool for modulating gene expression. The widespread application of RNA interference is made difficult because of the imperfections of the methods used for efficient target cell delivery of whatever genes are under study. One of the most convenient and efficient gene transfer and expression systems is based on the use of lentiviral vectors, which direct the synthesis of small hairpin RNAs (shRNAs), the precursors of siRNAs. The application of these systems enables one to achieve sustainable and long-term shRNA expression in cells. This review considers the adaptation of the processing of artificial shRNA to the mechanisms used by cellular microRNAs and simultaneous expression of several shRNAs as potential approaches for producing lentiviral vectors that direct shRNA synthesis. Approaches to using RNA interference for the treatment of cancer, as well as hereditary and viral diseases, are under active development today. The improvement made to the methods for constructing lentiviral vectors and the investigation into the mechanisms of processing of small interfering RNA allow one to now consider lentiviral vectors that direct shRNA synthesis as one of the most promising tools for delivering small interfering RNAs.
KEYWORDS lentiviral vectors; shRNA; RNA interference.
ABBREVIATIONS siRNA – small interfering RNA; miRNA – microRNA; RISC – RNA-induced silencing complex; shRNA – small hairpin RNA; dsRNA – double-stranded RNA; HIV-1 – human immunodeficiency virus type I; VSV-G – G protein of vesicular stomatitis virus; CMV – cytomegalovirus; H1, U6 – DNA polymerase III promoters.

INTRODUCTION

RNA interference is commonly used to inhibit gene expression. The advantages of this method include its simplicity, the possibility of quickly and significantly reducing the expression of any gene of interest, and the high specificity of the action. These properties render RNA interference a useful tool for investigating the role of specific genes in various cellular processes. For this purpose, entire libraries of siRNAs directed against a large number of genes have been created. Methods for applying RNA interference to the treatment of hereditary diseases, various neurodegenerative diseases, cancer, and as an antiviral therapy agent are currently under development. The search for new targets, the influence on which is efficient for treating a variety of diseases, is yet another application for RNA interference.

RNA INTERFERANCE

RNA interference is a sequence-specific mechanism of suppressing gene expression, which is induced by the presence of exogenous or endogenous double-stranded RNA (dsRNA) in a cell [1]. This evolutionarily conserved mechanism functions in virtually all eukaryotic organisms. The sources of exogenous dsRNA include viruses or artificially introduced dsRNA. Endogenous dsRNA is formed as a result of the transcription of a cell's own genes and often performs regulatory functions. The cleavage of long dsRNA by the Dicer protein, which belongs to the RNase type III family (*Fig. 1*), resulting in the formation of small 21- to 25-nucleotide-long siRNA duplexes is the shared stage of all types of RNA interference. The duplex contains a pair of unpaired nucleotides and a pair of hydroxyl groups at the 3'-ends and monophosphates at the 5'-ends (*Fig. 1*). This structure of RNA duplexes enables their normal processing by a protein belonging to the Ago family, which plays a key role in the formation of the RISC complex (RNA-induced silencing complex) [2]. The RNA fragments formed as a result of Dicer-mediated cleavage of dsRNA are included in the structure of a RLC complex (RISC-loading complex) containing Dicer and TR BP proteins. During the next phase, the formation of a

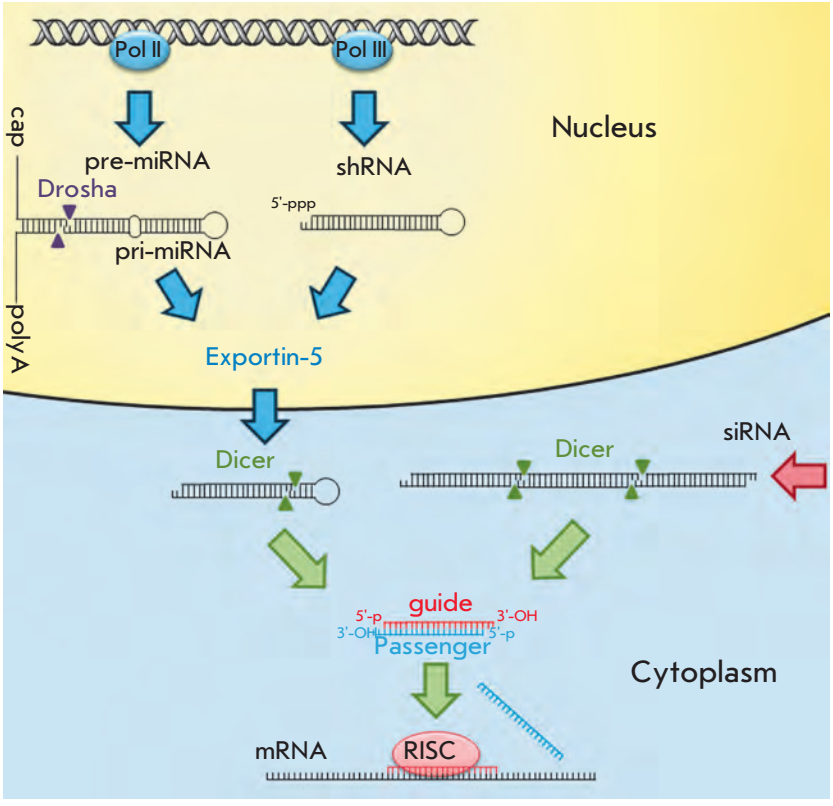


Fig. 1. Cellular processing of small interfering RNAs. pri-miRNAs are transcribed by RNA polymerase II; the resulting transcripts undergo capping and polyadenylation. After these processes, pre-miRNAs are spliced out under the influence of Drosha (RNase type III). shRNA is transcribed by RNA polymerase III to form shRNA with a triphosphate at its 5'-end. Both hairpin structures (pre-miRNA and shRNA) are transported from the nucleus to the cytoplasm by the Exportin-5 protein. In the cytoplasm, the Dicer protein splices out the sequences of future miRNAs and siRNAs from the hairpin structures and exogenous double-stranded RNAs. As a result of the processing by the Dicer protein, 21- to 25-bp-long RNA duplexes having a pair of unpaired nucleotides at their 3'-ends, OH-groups at their 3'-ends, and monophosphates at their 5'-ends are formed. The guide strand is loaded into the RISC protein complex, which binds to the mRNA that is complementary to the sequence of the guide strand. Meanwhile, the passenger strand is removed from the complex

pre-RISC complex (complex preceding RISC) occurs. The structure of this complex includes the Ago-2 protein, which cleaves the RNA duplex, so that only the guide strand is retained in the complex [3]. This strand determines the specificity of expression suppression, while the other strand (known as the passenger strand) is removed from the complex [4]. The selection of the guide strand is independent of the prospective target; the strand whose 5'-end is characterized by a lower thermodynamic stability becomes the guide strand [5]. During the next phase, the guide strand forms a part of the RISC complex and binds to the site of the target mRNA according to the principle of complementarity (*Fig. 1*). The process of mRNA destruction involves two stages. First, a primary gap appears in the mRNA molecule, which is attributed to the endonuclease activity of the PIWI-domain of the Ago protein. This is followed by the destruction (degradation) of the target mRNA by cellular exonucleases [6]. If the complementarity of siRNA and mRNA is incomplete, the primary gap is not formed; hence, mRNA is not subjected to degradation. It is important to mention that even if the complementarity between the guide strand and mRNA is incomplete, suppression of gene expression can occur at the translation stage in a similar fashion to miRNA [7]. An alternative mechanism of siRNA action is associated with the formation of the RITS (RNA-induced transcriptional silencing) complex, which includes the Ago-1 protein. The target mRNA is recognized by the RITS complex due to its interaction with RNA polymerase II during transcription [8]. During the next phase, the histone methyltransferases that methylate histones can become a part of the RITS complex, resulting in chromatin compaction and inhibition of the expression at the epigenetic level.

MicroRNAs differ from siRNAs by their mechanism of action and some features of their processing. Transcription of miRNAs is carried out by RNA polymerase II. The resulting RNAs undergo capping and polyadenylation [9]. Certain miRNAs are encoded by individual genes, while others are encoded by entire gene clusters. miRNAs can be transcribed together with mRNAs; the sequence encoding miRNA is located in the intron of a protein-coding gene [7]. As a result of the transcription, pri-miRNA (miRNA precursor) is formed. Its structure includes a sequence of the future miRNA, the terminating loop, and flanking sequences [10]. Processing of pri-miRNAs occurs in the nucleus with assistance from a complex consisting of two types of RNase III, Drosha and DGCR8 (in mammals). Approximately 65-bp-long hairpin-like miRNA precursors are formed as a result of the processing (*Fig. 1*) [11]. Transport of pre-miRNAs

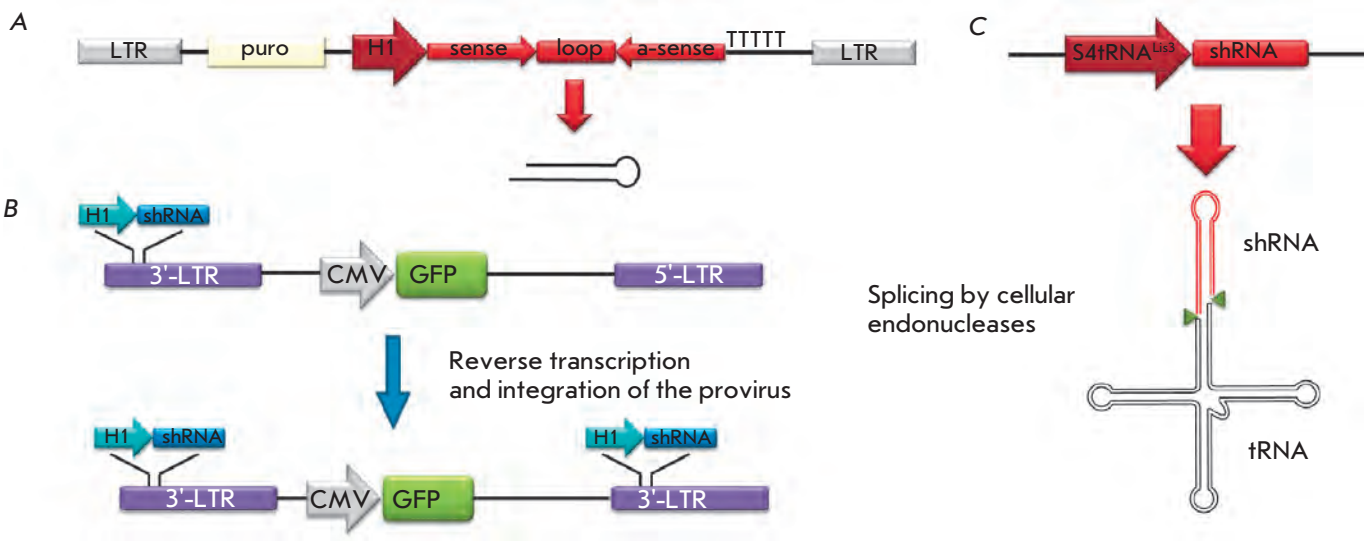


Fig. 2. Schematic representations of the vectors that direct shRNA synthesis. *A* – The expression cassette is inserted between two LTR-sequences of the lentiviral vector. The expression of shRNA is directed by the H1 promoter. Transcription is initiated from the sequence of future siRNA (sense), followed by the "loops," the inverted sequence, which is complementary to siRNA (α-sense), and termination sequence (thymines). The puromycin-resistance (puro) gene is also present in the vector; it enables the selection to be carried out [47]. *B* – The expression cassette is cloned into the 3'-LTR sequence. The cassette is doubled during reverse transcription; two copies of the expression cassette are formed following the integration of the provirus. A marker gene (e.g., green fluorescent protein (GFP) gene) under the control of the CMV promoter can also be cloned using this vector [45]. *C* – The expression cassette encoding shRNA fused with tRNA (S4tRNALys3-shRNA). After the transcription, the chimeric RNA undergoes processing in a similar fashion to normal tRNA, resulting in shRNA release [52]

to the cytoplasm is facilitated by the Exportin-5 protein (*Fig. 1*). In the cytoplasm, pre-miRNAs are cleaved by the Dicer protein, resulting in the formation of an approximately 22-bp-long duplex [12]. Unlike the siRNA duplex, the miRNA duplex typically contains unpaired nucleotides in the middle. The miRNA is subsequently included in the RISC complex in a similar fashion to siRNA [13].

In contrast to siRNA, miRNA is usually fully complementary only to a small fragment of the mRNA (several nucleotides long). The miRNA fragment, which is completely complementary to mRNA, most frequently comprises the nucleotides 2–8 from its 5'-end and is known as a "seed region." The "seed region" determines specific miRNA targets [14]. miRNA usually binds to the mRNA site, which is located in the 3'-untranslated region and is represented by multiple copies of the same mRNA. Since the length of the region that must be fully complementary is rather small, several different mRNAs can act as targets for a single miRNA. It is presumed that full complementarity of miRNA and target mRNA can lead to mRNA degradation, while partial complementary binding of miRNA to mRNA can disrupt translation [7, 15].

The introduction of long dsRNAs to mammalian cells induces interferon response; hence, short chemically synthesized siRNAs are used. Their structure is similar to that of natural siRNAs [16]. However, the effect of synthetic siRNAs is short-term (only a few days), which is attributed to their degradation by cellular nucleases. Moreover, the concentration of these siRNAs decreases during cell division. These drawbacks can be avoided if one uses vectors that direct the synthesis of siRNA precursors: small hairpin RNAs (shRNAa). shRNAs contain the sequence of the siRNA guide strand (21–29 bp long), followed by a loop consisting of approximately 9 nucleotides, and a sequence that is complementary to the siRNA guide strand (*Fig. 2A*). The use of this structure enables to achieve long-term suppression of gene expression [17].

Lentiviral vectors are optimal tools for the delivery of shRNAs into cells. An important feature of the life cycle of lentiviruses consists in their ability to integrate their genomes (in combination with proviral DNA) into cellular DNA. In addition, lentiviruses, as opposed to simple retroviruses, are capable of infecting nondividing cells. Despite the fact that such lentiviruses as the equine infectious anemia virus, feline immunodeficien-

cy virus and the bovine immunodeficiency virus are used as templates for lentiviral vectors, the human immunodeficiency virus type 1 (HIV-1) remains the most commonly used virus for vector production. This is associated with the fact that the life cycle of this virus is better understood than those of the other viruses [18, 19].

LENTIVIRAL VECTORS

Replication-incompetent systems based on lentiviruses are used for gene transfer and expression [20]. These systems enable the integration of the encoding target gene into the genome of the target cell's DNA (transgene). A typical lentiviral vector contains *cis*-elements of the viral genome, which are required for the assembly and integration of the viral particle and the sequence encoding the target gene. All trans-elements of the viral genome are removed from the vector. Co-transfection of a vector and the plasmids encoding viral proteins is the main approach used for obtaining lentiviral vectors [19]. In order to reduce the risk of occurrence of replication-competent particles due to recombination, the components of the viral genome required for the assembly of lentiviral vectors are typically divided into three or four plasmids: one or two packaging plasmids, the vector plasmid, and the plasmid encoding the viral envelope protein. Constructs with all cis-elements (except for RRE and the splice donor site that is required for post-transcriptional processing of mRNA) removed are used today in third generation packaging systems. A heterologous promoter (usually CMV) and the polyadenylation signal of the SV40 virus are used instead of a long terminal repeat (LTR). The *rev* and *gag/pol* genes are integrated into the cells using various expression cassettes. Humanization of the *gag/pol* genes is also employed, which enables their expression independently of *rev*. This also renders RRE removal from the packaging system possible [21]. It is important that such significant modifications of the packaging system do not affect the efficiency of the transduction by the lentiviral vector and significantly reduce the risk of occurrence of replication-competent particles due to homologous recombination. In order to reduce the risk of a nonhomologous recombination, a *trans*-lentiviral packaging system has been developed where the coding region of the *gag/pol* is divided into two parts and is incorporated into the structure of two different expression plasmids [22].

Pseudotyping of lentiviral vectors

In order to increase the tropism of lentiviral particles, the HIV-1 envelope protein is frequently replaced with the G protein of the vesicular stomatitis virus (VSV). These pseudotyped lentiviral particles enable the transduction of virtually all cell types. This modification not only expands the tropism of viral particles, but also increases their stability. Another important property of VSV-G is its ability to facilitate the penetration of the vector into the cell via endocytosis, thus reducing the need for auxiliary membrane proteins [23]. The main drawback of VSV-G pseudotyped lentiviral particles consists in their rapid elimination by the components of the immune system from the circulatory system [24].

One of the major problems encountered during the use of small interfering RNAs is the insufficient specificity of their delivery into the target cells. In addition to VSV-G, heterologous glycoproteins of lyssaviruses, the lymphocytic choriomeningitis virus, alphavirus and baculoviruses can also be used to carry out pseudotyping [25]. The transduction efficiency of liver cells is increased with the use of the hepatitis C virus or baculovirus envelope proteins [26]. Pseudotyping of lentiviral particles by the envelope proteins of the Rabies virus enables the lentiviruses to infect the cells of the central nervous system *in vivo* [27]. The envelope proteins of other viruses are frequently used to ensure more efficient tissue-specific transduction.

Methods that enable the presentation of various cellular receptors and their corresponding antibodies on the surface of viral particles are becoming more common [28–30]. The general principle in this approach is to create a fusion protein which can be successfully integrated into the envelope of the vector particles to ensure a relative stability of these particles, on the one hand. On the other hand, this protein carries a fragment of the ligand required for binding to the receptor. Most frequently, this chimeric protein is based on a glycoprotein of the amphotropic murine leukemia virus (A-MLV) and the hemagglutinins of the influenza and measles viruses. These viral envelope proteins are modified in such a way that they can no longer recognize their natural receptors, thus avoiding nonspecific infection. Lentiviral vectors containing the epidermal growth factor (EGF) or an anti-CD20 single-chain variable antibody fragment (scFv) on their surface, which are fused with the hemagglutinin of the measles virus and intended for infecting B cells, have been produced based on this scheme [31]. Another approach consists in producing lentiviral particles containing the glycoprotein A-MLV fused with anti-CD3 scFv or with interleukin-7 (IL-7), presented on their surface [32, 33]. This system enables the infection of T cells. Two ligands can be simultaneously used for pseudotyping: the stem cell factor (SCF) fused with the A-MLV glycoprotein and thrombopoietin (TPO) conjugated to the hemagglutinin of the influenza virus. Transduction of $CD34^+$ cells with lentiviral particles carrying either thrombopoietin, or SCF, or both ligands on their surface, has proved sig-

nificantly more efficient than the use of VSV-G as an envelope protein [34].

Utilization of viral surface envelope proteins is not the only way of presenting cell receptor ligands on the surface of viral particles. In this case, the utilized protein must contain a transmembrane domain; the surface of the viral vector must contain an envelope protein that can facilitate the fusion of the virus with the cell. Modified envelope proteins of the Sindbis virus or VSV-G, which have lost the ability to bind to their "native" receptor, are used for this purpose. The Sindbis virus has two surface envelope proteins, E1 and E2. The E1 protein is responsible for fusion with the cell, and E2 is responsible for binding to the receptor. The E1 protein functions independently of E2. A lentiviral vector containing the transmembrane form of SCF and a modified envelope protein of the Sindbis virus was produced according to this principle [35]. In the absence of the transmembrane domain, which is necessary for localization on the surface of lentiviral particles, the protein is attached to the transmembrane domain of VSV-G or human leukocyte antigen (HLA) [36]. For pseudotyping of lentiviral particles with antibodies, the packaging system must contain not only genes encoding light and heavy chains of antibodies, but also genes encoding Igα and Igβ proteins which are, required for antibody exposure on lentiviral particle surface.

This scheme was used to obtain lentiviral particles with surfaces containing anti-surface protein (CD20, DS-SIGN and CD3) antibodies [37–39].

Sindbis virus envelope proteins are also used for the pseudotyping of lentiviral particles by antibodies. For this purpose, the E2 protein is modified by incorporating the Fc-binding domain of protein A (ZZ-domain), which binds to immunoglobulin IgG, into its structure. Transduction using these lentiviral particles is only possible in the presence of monoclonal antibodies. The selection of antibodies determines the tropism of the lentivirus, enabling one to design viral particles that are specific with respect to cells of various origins without modifying the packaging system [40]. The disadvantages of Sindbis viral envelope proteins include the dependence of the protein E1 activity on pH (the pH value must lie within the range of 4.5–5.0). The reduced stability of these chimeric proteins during the pseudotyping of lentiviral particles can also be regarded as a drawback. The reduced efficiency of target cell infection using these lentiviral particles (which, however, can be compensated for by high specificity) should also be mentioned here.

Another approach that can provide specific infection of cells is the use of proteins as a component of the viral envelope, whose binding to a specific surface receptor results in a significant reduction in the efficiency of the transduction of those cells, the introduction of a transgene into which is undesirable [29]. This contamination-preventing protein can be bound to a viral glycoprotein using an amino acid sequence that is sensitive to certain proteases. The infection in this case involves two stages: first, the ligand on the viral surface binds to the cell receptor, and then cleavage of the peptide insertion occurs under the influence of certain proteases. After the insertion is cleaved, the glycoprotein can bind to its specific receptor on the cell surface. This approach enables to infect cells in the presence of specific proteases.

The use of tissue-specific promoters

Nonspecific cellular transduction, and therefore, transgene expression in these cells can cause a variety of adverse effects. In particular, transgene expression in antigen-presenting cells (APC) can result in the development of an immune response and T cell activation [41]. Tissue-specific promoters are used to reduce the effect of the nonspecific infection. Pseudotyping and the use of tissue-specific promoters enable to achieve transgene expression exclusively in the desired cells. However, tissue-specific promoters can be quite weak, and the level of expression of the target gene may be insufficient. The enhancers of stronger promoters can be used to strengthen these promoters. The enhancer of the CMV promoter used in combination with a variety of tissue-specific promoters provides a multifold increase in the expression of the target gene without decreasing the promoter specificity [42]. The site of transgene incorporation into the genome of the target cell is determined randomly; however, the incorporation takes place preferentially in the transcriptionally active regions. It is important to make an allowance for the fact that the incorporated transgene can accidentally come under the control of a strong promoter. In this case, its expression will be independent of the tissue specificity of the promoter. Insulators that block the effects of the neighboring enhancers are used to avoid this effect [28].

Transgene expression can be regulated at the post-transcriptional level. The mechanism underlying this regulation is based on RNA interference. Over 200 miRNAs exhibiting tissue-specific expression have been identified thus far. It has been demonstrated that the introduction of four sites recognized by miR-142 miRNA, which is expressed mainly in hematopoietic cells, into the gene encoding the green fluorescent protein (GFP) reduces the level of fluorescence exclusively in these cells [28, 43]. Taking into account the fact that new miRNA expression patterns are continuously identified in various cells, it can be assumed that this method is of significant interest for precise control of the expression of the introduced genes.

Small hairpin RNAs

shRNAs are siRNA precursors. They are typically expressed using U6 or H1 RNA polymerase III promoters (mouse or human) [43]. These promoters are small in size (about 400 bp long); transcription is initiated at the +1 position, and in the case of the U6 promoter it is desirable for the transcription to be initiated with guanine [44]. A sequence of 5–6 thymine residues acts as a transcription termination signal, resulting in the formation of double-stranded shRNA containing an unpaired 3'-end, which is essential for further processing by the Dicer protein. U6 and H1 promoters provide a stable and a relatively high level of shRNA expression in all cell types. shRNAs obtained as a result of RNA polymerase III-mediated transcription have neither 5'-caps nor 3'-poly (A) sequences; they are not processed by the Drosha protein. Their transport to the cytoplasm is carried out by the Exportin-5 protein [12]. The use of the RNA polymerase III promoter during the production of lentiviral vectors that direct the synthesis of shRNA allows one to attain a high level of shRNA expression in virtually all cell types. There are approaches in which the cassettes expressing shRNA are cloned into the 3'-LTR region of a lentiviral vector [45]. During the synthesis of a provirus, 3'-LTR is used as a template for 5'-LTR. As a result, two copies of the expression cassette are incorporated into the proviral insertion (*Fig. 2B*).

The lentiviral vector frequently includes marker genes. Genes encoding fluorescent proteins or antibiotic resistance genes are typically employed. The presence of marker genes in a vector enables the selection of transduced cells and evaluation of the transduction efficiency. Lentiviral vectors that direct the synthesis of shRNA were used in the production of cell lines characterized by a stable suppression of the expression of the activated oncogenes detected in acute myeloid leukemias. The puromycin resistance gene was introduced into the vector as a marker gene (*Fig. 2A*) [46, 47].

When constructing the vectors that direct the shRNA synthesis, it is important to take into account the fact that the increased level of shRNA expression in cells may have adverse consequences. It was demonstrated that transduction of mouse hepatocytes using an adeno-associated virus-based vector, the shRNA transcription in which is controlled by the U6 promoter, results in liver lesions in 50% of cases [48]. A total of 49 different vectors, each encoding a unique shRNA, were used in the study [48]. The toxic effect of these vectors is associated with the competition between shRNAs and cellular miRNAs for interaction with the Dicer and Exportin-5 proteins involved in the processing of both types of small RNAs. It is of significance that the resulting shRNA contains triphosphate at its 5'-end, which can cause an interferon response and stop the translation of cellular proteins. The presence of two unpaired nucleotides at the 3'-end of the shRNA stem is essential for efficient operation of the processing proteins (Exportin-5 and Dicer). An increase in the number of unpaired nucleotides significantly reduces the functional activity of these shRNAs [49–51]. The formation of triphosphate at the 5'-end can be avoided using an approach characterized by simultaneous transcription of shRNA and tRNA [52]. The chimeric RNA processed by cellular endonucleases is synthesized, resulting in the formation of shRNA containing monophosphate at its 5'-end (*Fig. 2C*). The use of tRNA promoters allows one to prevent the emergence of nonspecific responses; the expression level of shRNAs is considerably lower than when Polymerase III promoters are used.

If the first 2–8 nucleotides of the siRNA guide strand are complementary to the "seed region" of a particular miRNA molecule, then this siRNA molecule can function as a miRNA. This can trigger a nonspecific action from siRNA. The ability of siRNA to act as miRNA can be used to suppress the expression of certain genes (e.g., the CCR5 gene) [53]. siRNA, which is specific with respect to the CCR5 gene, is complementary to the "seed region" located in the 3'-UTR of mRNA. This siRNA caused the degradation of mRNA and resulted in disruption of translation in a fashion similar to the action of miRNA. When selecting shRNA sequences, one should bear in mind that the 5'-end of the guide strand of the duplex formed as a result of processing must be characterized by a lower thermodynamic stability. Inconsistency with these rules can result in the following: the passenger strand will become part of the RISC complex, instead of the guide strand, leading to a reduced specificity of the shRNA action. It is assumed that the H1 promoter is better suited for *in vivo* application as compared to the U6 promoter, since the H1 promoter is less toxic despite its lower efficiency [54]. Successful application of lentiviral vectors guiding the shRNA synthesis was demonstrated using animal models of various diseases [55–57]. In particular, the expression of shRNA persisted for 9 months following the injection of lentiviral particles, and suppression of the reporter gene expression in mouse brain cells was maintained [58].

The adverse effects associated with the use of shRNAs (interferon response, competition with cellular miRNAs, nonspecific action) can be avoided by employing various approaches. Several of them are based on the adaptation of the artificial shRNA processing to the mechanisms used by cellular miRNAs [59]. To achieve this objective, the sequence of the guide strand of the future miRNA can be replaced with an artificial sequence, while conserving the structure of miRNA pre-

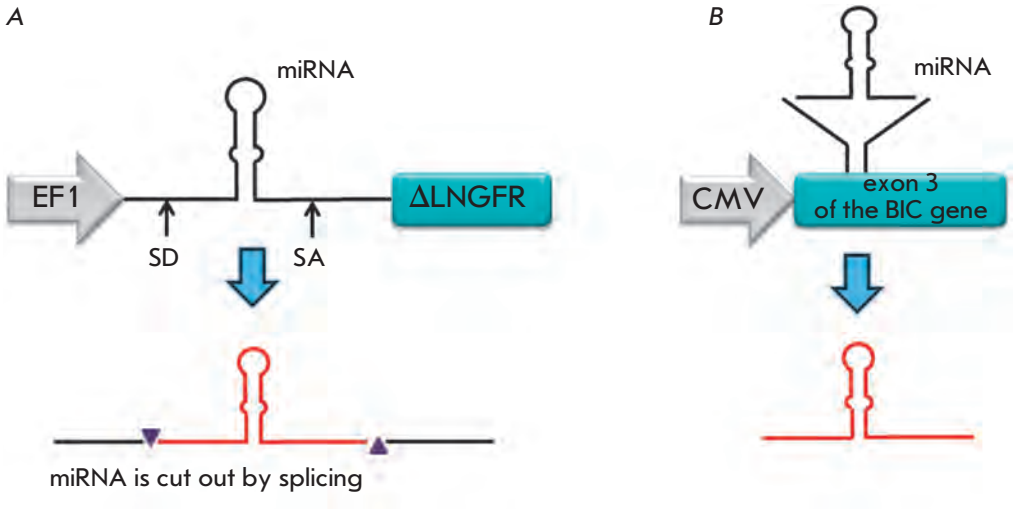


Fig. 3. Schematic representations of the vectors that direct the synthesis of the modified miRNAs: *A* – miRNA is cloned under the action of the EF1′ promoter in such a manner that it is expressed along with a fragment of the *NGFR(ΔLNGFR)* gene in the first intron and is cut out by splicing. SD – splice donor site; SA – splice acceptor site [62]. *B* – miRNA is expressed as a component of exon 3 of the *BIC* gene, where it was originally present [63]

cursors. miRNAs are transcribed by Polymerase II; thus, it is preferable to use the promoters of this enzyme when constructing the vector. It was demonstrated that expression of shRNA under the control of the U6 promoter in an adeno-associated virus-based vector is 10 times more efficient than the expression of miR-30 under the control of the same promoter. However, the suppression level of the reporter gene was approximately the same, while the toxic effect of the construct containing miR-30 was much lower [60, 61]. The fact that the Dicer protein can select both strands of the miRNA duplex is a drawback of miR-30-based systems. Cell transduction with lentiviral vectors carrying the gene for the nerve growth factor receptor (NGFR), whose first intron contains an integrated sequence encoding pri-miR-223 (200 bp) under the control of the integrated EF1α promoter, results in stable expression of the *NGFR* gene and miRNA (*Fig. 3A*). The sequence of the guide strand in the "stem" of miRNA can be replaced by other guide strand sequences from other miRNAs or siRNAs [62]. An approach enabling one to achieve stable expression of the mouse *BIC* gene and its miR-155 product characterized by an altered sequence of the guide strand has also been developed (*Fig. 3B*) [63]. The vector containing a fragment of the *BIC* gene (including the sequence of miR-155) directed the successful expression of both the source miRNA and miRNA with an altered sequence of the guide strand (*Fig. 3B*) [63]. General rules for constructing artificial miRNAs have yet to be developed, which is primarily due to insufficient knowledge with regard to their processing.

Simultaneous synthesis of several small interfering RNAs

In some cases, the simultaneous expression of multiple siRNAs is preferable (e.g., during antiviral therapy). This is attributed to the fact that some viruses mutate at a high rate, and the probability of developing resistance to specific siRNA among them is high. The use of multiplex constructs enabling the synthesis of several siRNAs significantly reduces the probability of emergence of resistant forms of the virus.

Therefore, a lentiviral vector that directs the synthesis of long hairpin RNAs (lhRNAs) containing the "loop-stem" structure was constructed. Processing of these lhRNAs occurs with assistance from the Dicer protein; several siRNAs are formed under the influence of the latter. The lhRNA (the precursor of siRNA) nucleotide sequence is selected according to the same principle as per shRNA selection process. Suppression of HIV-1 replication was achieved with assistance from the 50- to 80-bp-long lhRNA, which acts as a precursor for 2–3 siRNAs against various parts of the general region of *tat/rev* genes, (*Fig. 4A*) [64]. A similar approach was used to suppress the replication of the hepatitis B and C viruses [65, 66]. The efficiency of lhRNA processing by the Dicer protein decreases as the siRNA sequence approaches the "loop," resulting in the formation of various amounts of siRNA and a nonuniform suppression of target gene expression.

Since the promoters of RNA polymerase III are relatively small (200–400 bp), a single vector can incorporate several siRNA sequences, each controlled by its own promoter. Different RNA polymerase III promoters (U6, H1 and 7SK) are used in this case, since the utilization of identical promoters may induce recombination between their sequences and deletion of one or several expression cassettes in 80% of the cases [67]. A vector ensuring the synthesis of four shRNAs under the control of the mouse U6, H1 and human U6, 7SK promoters has been constructed. In this case, the mouse H1 and human U6 promoters were fused into a single

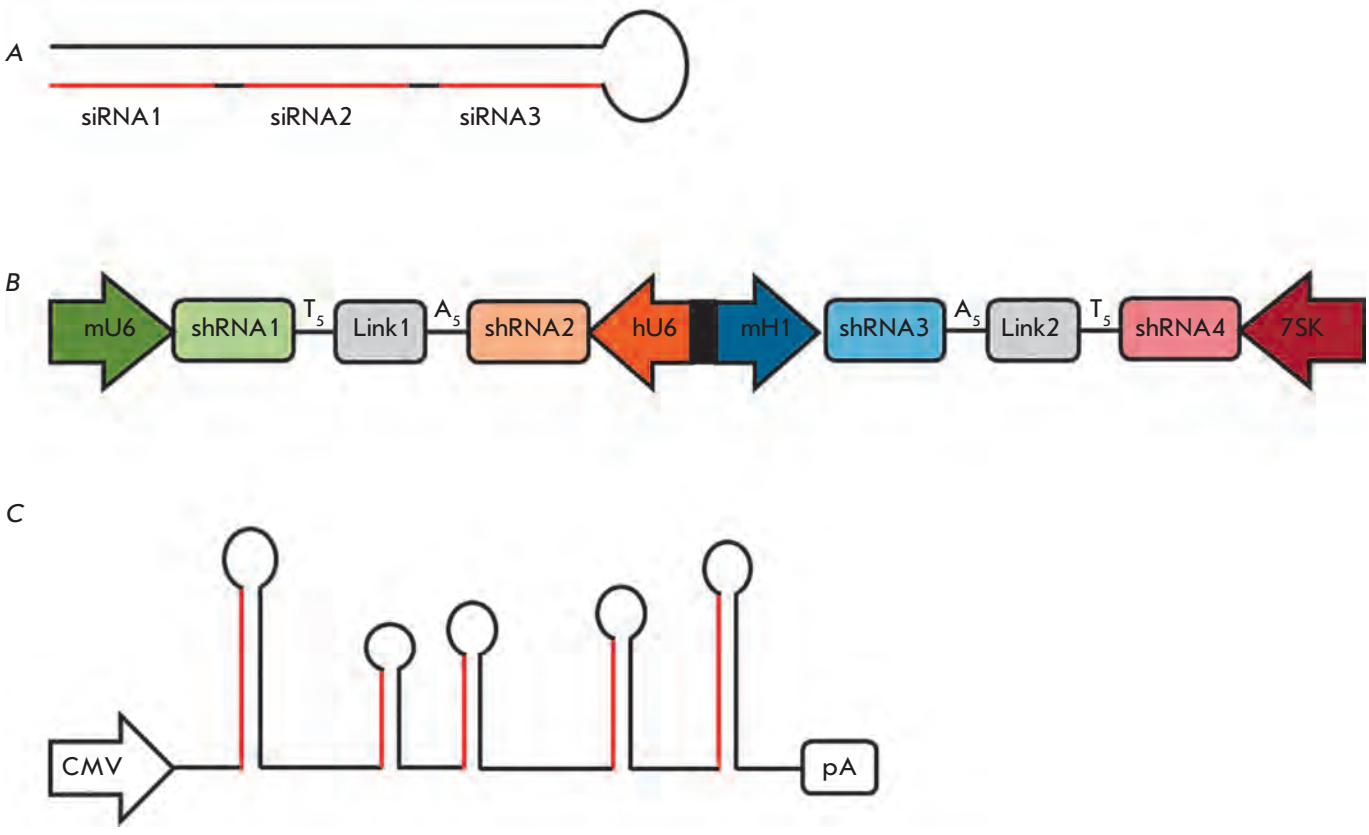


Fig. 4. Various approaches to the multiplex expression of small interfering RNAs. *A* – long-hairpin RNAs (lhRNAs) contain sequences of several siRNAs (highlighted in red), which are subsequently spliced out by the Dicer protein [64]. *B* – Expression of four shRNAs from a single vector under the influence of various promoters of RNA polymerase III [68]. *C* – Expression of several miRNAs using *mir-17-92* polycistron. miRNAs with altered sequences of the guide strands (highlighted in red) were integrated into the base of *mir-17-92* polycistron [69]

bidirectional promoter (*Fig. 4B*). Suppression of the expression of four different genes was achieved using this vector [68].

Clusters encoding polycistronic miRNAs, which form several pre-miRNAs, can be used for simultaneous expression of multiple siRNAs. Transcription of the *miR-17-92* gene cluster gives rise to double-stranded pri-miRNAs approximately 1 kbp in length. The latter are precursors of six different pre-miRNAs. The *mir-17-92* gene cluster was used to create lentiviral vectors that direct the synthesis of four HIV1-specific miRNAs. Sequences encoding pre-miRNAs and containing 40 nucleotides on each side of the "loop-stem" structure were obtained from the gene cluster and were incorporated into the vector. Sequences of the guide strands of the future miRNAs were replaced with segments specific with respect to HIV-1 (*Fig. 4C*) [69]. Due consideration was given to such features of the original structure of miRNAs as mismatches and thermodynamic stability during the replacement of the sequences of the guide strand.

The use of small interfering RNA

Approaches for the clinical application of small interfering RNAs are currently being developed. Dozens of siRNA-based medicinal agents designed to treat different kinds of diseases are currently undergoing clinical trials. Only one drug, which is based on the lentiviral delivery of shRNA, has been tested thus far. The use of lentiviral vectors directing the synthesis of shRNAs is constrained by the fact that they are relatively unsafe. This is attributed to possible nonspecific responses, which can be caused by shRNA expression in cells and the probable insertional mutagenesis. However, the use of lentiviral vectors for siRNA delivery has a number of significant advantages. They can be used to achieve stable and prolonged shRNA synthesis in dividing and nondividing cells, making their application rather

promising for the treatment of chronic diseases.

The mechanism of RNA interference is a component of the antiviral defense system of the organism; therefore, the use of RNA interference in chronic viral infections [59, 70, 71] (including the diseases caused by the hepatitis B and C and HIV-1 viruses) is of considerable interest. However, the use of RNA interference may result in the emergence of resistant forms of the virus, which limits the application of this method [72]. Contemporary methods enable the creation of lentiviral vectors that can simultaneously encode three or four shRNAs which are specific with respect to various viral genes. This can significantly reduce the probability of emergence of resistant forms of the virus. Existing methods of siRNA delivery to T cells and macrophages (HIV-1 targets) are inefficient. The use of lentiviral vectors can be an efficient approach to introducing siRNAs into cells targeted by HIV-1. However, in the case of lentiviral vectors directing the synthesis of shRNAs, which are specific with respect to viral genes, reduction in the efficiency of lentiviral particles and their titer is possible [73]. Thus, point mutations that do not affect the synthesis of the proteins required for the assembly of viral particles are introduced into the genes used in the packaging system. Selection of these mutations complicates the process of vector construction, especially if shRNAs are selected for the conserved HIV-1 sites. After the infection with HIV-1, the viral envelope protein binds to the $CD4^+$ receptor exposed on the surface of the target cells; the virus uses the CCR5 cell receptor as a co-receptor. It has been demonstrated that homozygous deletion of the human *CCR5* gene renders cells resistant to the HIV-1 infection, and the mutation apparently has almost no effect on the normal functioning of the cells [74]. It was demonstrated that shRNA-mediated suppression of the CCR5 receptor expression also renders cells resistant to infection by the virus *in vitro* and *in vivo* [75–78]. Several proteins whose functions are not essential to T cells or macrophages and which play an important role in the life cycle of HIV-1 have been identified [79].

The optimal approach is to obtain HIV-1-resistant T cells and macrophages from their common progenitors. To achieve this objective, transduction of early hematopoietic precursor cells was carried out using lentiviral vectors that direct the synthesis of shRNAs that are specific with respect to the CCR5 or CXCR-4 gene. The descendants of these cells (T cells and macrophages) acquired resistance to the virus [80–82]. There is an approach that enables the expression of shRNA, along with the other genes. The lentiviral vector that was successfully used to provide the synthesis of a false target for the viral TAT protein in addition to the expression of shRNA specific with respect to the general region of the *tat/rev* genes is an example of the latter concept. This false target impedes the action of the TAT protein and synthesis of ribozyme, which is specific with respect to the CCR 5 receptor [83, 84]. The efficiency of this vector was tested on humanized mice; stable inhibition of HIV-1 at different stages of the life cycle was achieved using this vector [85]. Clinical trials demonstrated the safety of using this vector in autologous bone marrow transplants in patients with HIV-1 and lymphoma. The patients with lymphoma at the remission stage, which resulted from a conventional treatment regimen, exhibited no side effects associated with the introduction of shRNAs. A detectable level of shRNA expression persisted in patients for 24 months.

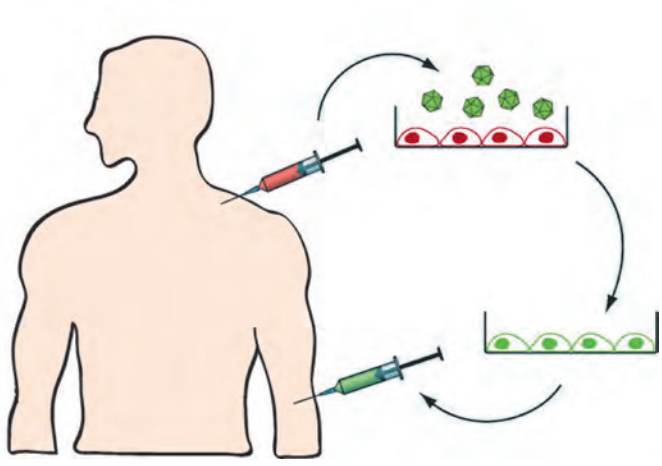


Fig. 5. *Ex vivo* transduction of bone marrow cells. During autologous transplantation bone marrow cells are transduced with lentiviral vectors that direct the synthesis of shRNAs. Transduced cells are then administered to the patient following radiation therapy [84]

Malignant tumors develop as a result of mutations leading to an abnormal expression of the genes that stimulate cell proliferation and impairing apoptosis. RNA interference is a useful tool for modulating gene expression. It is considered that methods based on the principle of RNA interference can be of considerable interest in the treatment of tumors. The classical approaches to the therapy of malignant diseases are characterized by a number of significant deficiencies associated with the nonspecificity of their action. The use of RNA interference enables to exert a specific effect on oncogenes at a relatively low cost. A total of 10 siRNA-based drugs are currently undergoing clinical trials. The main obstacle associated with the use of RNA interference in treating malignant diseases is the imperfections of the methods for siRNA delivery to tumor cells. One of the most convenient and efficient gene

transfer systems is based on the use of lentiviral vectors. These systems enable a highly specific integration of sequences encoding shRNA into the target cell's genome. Methods for pseudotyping lentiviral particles are being developed and tissue-specific promoters are used in order to achieve this objective. Systems with multiplex shRNA expression show promise as well. Multiplex shRNA expression enables specific inhibition of multiple genes involved in tumor development. The use of multiple shRNAs specific to different regions of the same activated oncogene makes it possible to improve the efficiency of these systems [86].

Many human miRNAs are capable of inhibiting the growth of malignant tumors [87, 88]. Thereby, some of the research teams are working on the use of miRNAs for treating malignant tumors whose cells are characterized by a lower expression of oncosuppressor miRNAs. It was demonstrated that restoration of the expression of oncosuppressor miRNA decreases cell growth in patients suffering from non-small-cell lung carcinoma, breast cancer, liver cancer, and chronic lymphocytic leukemia [89–92]. However, the inefficiency of *in vivo* transduction still remains the major problem when using lentiviral vectors as the main form of therapy. Leukemia therapy is considered to be the most suitable area for the use of lentiviral vectors that direct shRNA synthesis (*Fig. 5*). Autologous transplantation of hematopoietic cells transduced with a lentiviral vector which is specific with respect to one or several activated oncogenes can also be promising. The safety of this approach was demonstrated using lentiviral vectors that direct the synthesis of shRNAs capable of inhibiting HIV-1 [93].

The search for new target genes that are involved in tumor development is also regarded as a promising application for shRNAs. Nowadays, gene expression profiles in cancer cells are being actively studied. This has already enabled the discovery of several genes whose increased expression is associated with specific types of tumors. Vast libraries of lentiviral shRNA-based vectors enabling the search for the genes that are considered promising for the development of novel chemotherapeutic agents have been created [94, 95]. The shRNA-induced inhibition of oncogene expression allows one to assess the contribution of these genes to the maintenance of the malignant status of tumor cells. A similar approach was used to transduce cells derived from a patient with acute myeloid leukemia using lentiviral vectors that direct shRNA synthesis, which are specific for *c-kit* and *AML1-ETO* oncogenes. The system was successfully used to investigate the inhibitory action of binase on the tyrosine kinase KIT receptor [96].

It is thought that shRNA can be successfully used in the gene therapy of neurodegenerative diseases, such as the Alzheimer's, Parkinson's and Huntington's diseases. Their application is considered to be extremely promising, since lentiviruses are capable of overcoming the blood-brain barrier and infecting cells of the central nervous system (CNS). Lentiviral vectors enable to achieve a stable shRNA synthesis, which can be extremely important in the treatment of these chronic diseases. Pseudotyping of lentiviral vectors using the rabies virus envelope protein can increase efficiency in the infection of CNS cells. ●

This work was supported by the State contract № 16.512.11.2266, Program of the Presidium of the Russian Academy of Sciences "Molecular and Cell Biology", and the Russian Foundation for Basic Research (grant № 11-04-01365-a).

REFERENCES

1. Meister G., Tuschl T. // Nature. 2004. V. 431. № 7006. P. 343–349.
2. Tomari Y., Zamore P.D. // Genes Dev. 2005. V. 19. № 5. P. 517–529.
3. Kim K., Lee Y.S., Carthew R.W. // RNA. 2007. V. 13. № 1. P. 22–29.
4. Macrae I.J., Ma E., Zhou M., Robinson C.V., Doudna J.A. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 2. P. 512–517.
5. Collins R.E., Cheng X. // FEBS Lett. 2005. V. 579. № 26. P. 5841–5849.
6. Shen B., Goodman H.M. // Science. 2004. V. 306. № 5698. P. 997.
7. Carthew R.W., Sontheimer E.J. // Cell. 2009. V. 136. № 4. P. 642–655.
8. Lippman Z., Martienssen R. // Nature. 2004. V. 431. № 7006. P. 364–370.
9. Kim V.N. // Nat. Rev. Mol. Cell Biol. 2005. V. 6. № 5. P. 376–385.
10. Bartel D.P. // Cell. 2004. V. 116. № 2. P. 281–297.
11. Han J., Lee Y., Yeom K.H., Nam J.W., Heo I., Rhee J.K., Sohn S.Y., Cho Y., Zhang B.T., Kim V.N. // Cell. 2006. V. 125. № 5. P. 887–901.
12. Yi R., Qin Y., Macara I.G., Cullen B.R. // Genes Dev. 2003. V. 17. № 24. P. 3011–3016.
13. Gregory R.I., Chendrimada T.P., Cooch N., Shiekhattar R. // Cell. 2005. V. 123. № 4. P. 631–640.
14. Lai E.C. // Nat. Genet. 2002. V. 30. № 4. P. 363–364.
15. Wu L., Fan J., Belasco J.G. // Proc. Natl. Acad. Sci. USA. 2006. V. 103. № 11. P. 4034–4039.
16. Agrawal N., Dasaradhi P.V., Mohmmed A., Malhotra P., Bhatnagar R.K., Mukherjee S.K. // Microbiol. Mol. Biol. Rev. 2003. V. 67. № 4. P. 657–685.
17. Van den Haute C., Eggermont K., Nuttin B., Debyser Z., Baekelandt V. // Hum. Gene Ther. 2003. V. 14. № 18. P. 1799–1807.
18. Cockrell A.S., Kafri T. // Mol. Biotechnol. 2007. V. 36. № 3. P. 184–204.
19. Spirin P.V., Vil'gelm A.E., Prassolov V.S. // Mol Biol (Mosk). 2008. V. 42. № 5. P. 913–926.

20. Pluta K., Kacprzak M.M. // Acta Biochim. Pol. 2009. V. 56. № 4. P. 531–595.
21. Kotsopoulou E., Kim V.N., Kingsman A.J., Kingsman S.M., Mitrophanous K.A. // J. Virol. 2000. V. 74. № 10. P. 4839–4852.
22. Kappes J.C., Wu X., Wakefield J.K. // Methods Mol. Med. 2003. V. 76. P. 449–465.
23. Aiken C. // J. Virol. 1997. V. 71. № 8. P. 5871–5877.
24. Croyle M.A., Callahan S.M., Auricchio A., Schumer G., Linse K.D., Wilson J.M., Brunner L.J., Kobinger G.P. // J. Virol. 2004. V. 78. № 2. P. 912–921.
25. Cronin J., Zhang X.Y., Reiser J. // Curr. Gene Ther. 2005. V. 5. № 4. P. 387–398.
26. Bartosch B., Vitelli A., Granier C., Goujon C., Dubuisson J., Pascale S., Scarselli E., Cortese R., Nicosia A., Cosset F.L. // J. Biol. Chem. 2003. V. 278. № 43. P. 41624–41630.
27. Mazarakis N.D., Azzouz M., Rohll J.B., Ellard F.M., Wilkes F.J., Olsen A.L., Carter E.E., Barber R.D., Baban D.F., Kingsman S.M., et al. // Hum. Mol. Genet. 2001. V. 10. № 19. P. 2109–2121.
28. Frecha C., Szécsi J., Cosset F.L., Verhoeyen E. // Curr. Gene Ther. 2008. V. 8. № 6. P. 449–460.
29. Verhoeyen E., Cosset F.L. // J. Gene Med. 2004. V. 6. Suppl 1. P. S83–S94.
30. Mátrai J., Chuah M.K., Van den Driessche T. // Mol. Ther. 2010. V. 18. № 3. P. 477–490.
31. Funke S., Maisner A., Mühlebach M.D., Koehl U., Grez M., Cattaneo R., Cichutek K., Buchholz C.J. // Mol. Ther. 2008. V. 16. № 8. P. 1427–1436.
32. Maurice M., Verhoeyen E., Salmon P., Trono D., Russell S.J., Cosset F.L. // Blood. 2002. V. 99. № 7. P. 2342–2350.
33. Verhoeyen E., Dardalhon V., Ducrey-Rundquist O., Trono D., Taylor N., Cosset F.L. // Blood. 2003. V. 101. № 6. P. 2167–2174.
34. Verhoeyen E., Nègre D., Cosset F.L. // Methods Mol. Biol. 2008. V. 434. P. 99–112.
35. Froelich S., Ziegler L., Stroup K., Wang P. // Biotechnol. Bioeng. 2009. V. 104. № 1. P. 206–215.
36. Lei Y., Joo K.I., Zarzar J., Wong C., Wang P. // Virol. J. 2010. V. 7. P. 35.
37. Yang L., Bailey L., Baltimore D., Wang P. // Proc. Natl. Acad. Sci. USA. 2006. V. 103. № 31. P. 11479–11484.
38. Yang H., Joo K.I., Ziegler L., Wang P. // Pharm. Res. 2009. V. 26. № 6. P. 1432–1445.
39. Yang L., Yang H., Rideout K., Cho T., Joo K.I., Ziegler L., Elliot A., Walls A., Yu D., Baltimore D., et al. // Nat. Biotechnol. 2008. V. 26. № 3. P. 326–334.
40. Morizono K., Bristol G., Xie Y.M., Kung S.K., Chen I.S. // J. Virol. 2001. V. 75. № 17. P. 8016–8020.
41. Cui Y., Kelleher E., Straley E., Fuchs E., Gorski K., Levitsky H., Borrello I., Civin C.I., Schoenberger S.P., Cheng L., et al. // Nat. Med. 2003. V. 9. № 7. P. 952–958.
42. Gruh I., Wunderlich S., Winkler M., Schwanke K., Heinke J., Blömer U., Ruhparwar A., Rohde B., Li R.K., Haverich A., et al. // J. Gene Med. 2008. V. 10. № 1. P. 21–32.
43. Dykxhoorn D.M., Novina C.D., Sharp P.A. // Nat. Rev. Mol. Cell. Biol. 2003. V. 4. № 6. P. 457–467.
44. Boden D., Pusch O., Lee F., Tucker L., Shank P.R., Ramratnam B. // Nucl. Acids Res. 2003. V. 31. № 17. P. 5033–5038.
45. Tiscornia G., Singer O., Verma I.M. // Nat. Protoc. 2006. V. 1. № 1. P. 234–240.
46. Spirin P.V., Baskaran D., Rubtsov P.M., Zenkova M.A., Vlassov V.V., Chernolovskaya E.L.,Prassolov V.S. // Acta Naturae. 2009. V. 1. № 2. P. 86–90.
47. Spirin P.V., Nikitenko N.A., Lebedev T.D., Rubtsov P.M., Stocking C., Prassolov V.S.// Mol Biol (Mosk). 2011. V. 45. № 6. P. 1036–1045.
48. Grimm D., Streetz K.L., Jopling C.L., Storm T.A., Pandey K., Davis C.R., Marion P., Salazar F., Kay M.A. // Nature. 2006. V. 441. № 7092. P. 537–541.
49. Hornung V., Ellegast J., Kim S., Brzozka K., Jung A., Kato H., Poeck H., Akira S., Conzelmann K.K., Schlee M., et al. // Science. 2006. V. 314. № 5801. P. 994–997.
50. Pichlmair A., Schulz O., Tan C.P., Naslund T.I., Liljestrom P., Weber F., Reis e Sousa C. // Science. 2006. V. 314. № 5801. P. 997–1001.
51. Kenworthy R., Lambert D., Yang F., Wang N., Chen Z., Zhu H., Zhu F., Liu C., Li K., Tang H. // Nucl. Acids Res. 2009. V. 37. № 19. P. 6587–6599.
52. Scherer L.J., Frank R., Rossi J.J. // Nucl. Acids Res. 2007. V. 35. № 8. P. 2620–2628.
53. Ehsani A., Saetrom P., Zhang J., Alluin J., Li H., Snøve O. Jr., Aagaard L., Rossi J.J. // Mol. Ther. 2010. V. 18. № 4. P. 796–802.
54. An D.S., Qin F.X., Auyeung V.C., Mao S.H., Kung S.K., Baltimore D., Chen I.S. // Mol. Ther. 2006. V. 14. № 4. P. 494–504.
55. Gupta S., Maitra R., Young D., Gupta A., Sen S. // Am. J. Physiol. Heart Circ. Physiol. 2009. V. 297. № 2. P. 627–636.
56. Bot I., Guo J., van Eck M., van Santbrink P.J., Groot P.H., Hildebrand R.B., Seppen J., van Berkel T.J., Biessen E.A. // Blood. 2005. V. 106. № 4. P. 1147–1153.
57. Eren-Koçak E., Turner C.A., Watson S.J., Akil H. // Biol. Psychiatry. 2011. V. 69. № 6. P. 534–540.
58. Mäkinen P.I., Koponen J.K., Kärkkäinen A.M., Malm T.M., Pulkkinen K.H., Koistinaho J., Turunen M.P., Ylä-Herttuala S. // J. Gene Med. 2006. V. 8. № 4. P. 433–441.
59. Manjunath N., Wu H., Subramanya S., Shankar P. // Adv. Drug Deliv. Rev. 2009. V. 61. № 9. P. 732–745.
60. McBride J.L., Boudreau R.L., Harper S.Q., Staber P.D., Monteys A.M., Martins I., Gilmore B.L., Burstein H., Peluso R.W., Polisky B., et al. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 15. P. 5868–5873.
61. Zeng Y., Wagner E.J., Cullen B.R. // Mol. Cell. 2002. V. 9. № 6. P. 1327–1333.
62. Amendola M., Passerini L., Pucci F., Gentner B., Bacchetta R., Naldini L. // Mol. Ther. 2009. V. 17. № 6. P. 1039–1052.
63. Chung K.H., Hart C.C., Al-Bassam S., Avery A., Taylor J., Patel P.D., Vojtek A.B., Turner D.L. // Nucl. Acids Res. 2006. V. 34. № 7. P. 53.
64. Sano M., Li H., Nakanishi M., Rossi J.J. // Mol. Ther. 2008. V. 16. № 1. P. 170–177.
65. Watanabe T., Sudoh M., Miyagishi M., Akashi H., Arai M., Inoue K., Taira K., Yoshiba M., Kohara M. // Gene Ther. 2006. V. 13. № 11. P. 883–892.
66. Weinberg M.S., Ely A., Barichievy S., Crowther C., Mufamadi S., Carmona S., Arbuthnot P. // Mol. Ther. 2007. V. 15. № 3. P. 534–541.
67. Brake O., Hooft K.T., Liu Y.P., Centlivre M., von Eije K.J., Berkhout B. // Mol. Ther. 2008. V. 16. № 3. P. 557–564.
68. Gou D., Weng T., Wang Y., Wang Z., Zhang H., Gao L., Chen Z., Wang P., Liu L. // J. Gene Med. 2007. V. 9. № 9. P. 751–763.
69. Liu Y.P., Haasnoot J., ter Brake O., Berkhout B., Konstantinova P. // Nucl. Acids Res. 2008. V. 36. № 9. P. 2811–2824.
70. Ashfaq U.A., Yousaf M.Z., Aslam M., Ejaz R., Jahan S., Ullah O. // Virol. J. 2011. V. 8. P. 276.
71. Morris K.V., Rossi J.J. // 3 2006. V. 13. № 6. P. 553–558.

72. Zheng Z.M., Tang S., Tao M. // Ann. N.Y. Acad. Sci. 2005. V. 1058. P. 105–118.
73. ter Brake O., Berkhout B. // J. Gene Med. 2007. V. 9. № 9. P. 743–750.
74. Samson M., Libert F., Doranz B.J., Rucker J., Liesnard C., Farber C.M., Saragosti S., Lapoumeroulie C., Cognaux J., Forceille C., et al. // Nature. 1996. V. 382. № 6593. P. 722–725.
75. Anderson J., Akkina R. // Gene Ther. 2007. V. 14. № 17. P. 1287–1297.
76. Qin X.F., An D.S., Chen I.S., Baltimore D. // Proc. Natl. Acad. Sci. USA. 2003. V. 100. № 1. P. 183–188.
77. Lee M.T., Coburn G.A., McClure M.O., Cullen B.R. // J. Virol. 2003. V. 77. № 22. P. 11964–11972.
78. Butticaz C., Ciuffi A., Muñoz M., Thomas J., Bridge A., Pebernard S., Iggo R., Meylan P., Telenti A. // Antivir. Ther. 2003. V. 8. № 5. P. 373–377.
79. Brass A.L., Dykxhoorn D.M., Benita Y., Yan N., Engelman A., Xavier R.J., Lieberman J., Elledge S.J. // Science. 2008. V. 319. № 5865. P. 921–926.
80. Anderson J., Akkina R. // Retrovirology. 2005. V. 2. P. 53.
81. Banerjea A., Li M.J., Bauer G., Remling L., Lee N.S., Rossi J., Akkina R. // Mol. Ther. 2003. V. 8. № 1. P. 62–71.
82. An D.S., Donahue R.E., Kamata M., Poon B., Metzger M., Mao S.H., Bonifacino A., Krouse A.E., Darlix J.L., Baltimore D., et al. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. № 32. P. 13110–13115.
83. Li M.J., Kim J., Li S., Zaia J., Yee J.K., Anderson J., Akkina R., Rossi J.J. // Mol. Ther. 2005. V. 12. № 5. P. 900–909.
84. Tiemann K., Rossi J.J. // EMBO Mol. Med. 2009. V. 1. № 3. P. 142–151.
85. Anderson J., Li M.J., Palmer B., Remling L., Li S., Yam P., Yee J.K., Rossi J., Zaia J., Akkina R. // Mol. Ther. 2007. V. 15. № 6. P. 1182–1188.
86. Senzer N., Barve M., Kuhn J., Melnyk A., Beitsch P., Lazar M., Lifshitz S., Magee M., Oh J., Mill S.W., et al. // Mol. Ther. 2012. V. 20. № 3. P. 679–686.
87. Zhang B., Pan X., Cobb G.P., Anderson T.A. // Dev. Biol. 2007. V. 302. № 1. P. 1–12.
88. Li C., Feng Y., Coukos G., Zhang L. // AAPS J. 2009. V. 11. № 4. P. 747–757.
89. Yu F., Yao H., Zhu P., Zhang X., Pan Q., Gong C., Huang Y., Hu X., Su F., Lieberman J., et al. // Cell. 2007. V. 131. № 6. P. 1109–1123.
90. Cimmino A., Calin G.A., Fabbri M., Iorio M.V., Ferracin M., Shimizu M., Wojcik S.E., Aqeilan R.I., Zupo S., Dono M., et al. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. № 39. P. 13944–13949.
91. Kota J., Chivukula R.R., O'Donnell K.A., Wentzel E.A., Montgomery C.L., Hwang H.W., Chang T.C., Vivekanandan P., Torbenson M., Clark K.R., et al. // Cell. 2009. V. 137. № 6. P. 1005–1017.
92. Kumar M.S., Erkeland S.J., Pester R.E., Chen C.Y., Ebert M.S., Sharp P.A., Jacks T. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 10. P. 3903–3908.
93. DiGiusto D.L., Krishnan A., Li L., Li H., Li S., Rao A., Mi S., Yam P., Stinson S., Kalos M., et al. // Sci. Transl. Med. 2010. V. 2. № 36. P. 36–43.
94. Yoshino S., Hara T., Weng J.S., Takahashi Y., Seiki M., Sakamoto T. // PLoS One. 2012. V. 7. № 4. P. 35590.
95. Palchaudhuri R., Hergenrother P.J. // ACS Chem. Biol. 2011. V. 6. № 1. P. 21–33.
96. Mitkevich V.A., Petrushanko I.Y., Spirin P.V., Fedorova T.V., Kretova O.V., Tchurikov N.A., Prassolov V.S., Ilinskaya O.N., Makarov A.A. // Cell Cycle. 2011. V. 10. № 23. P. 4090–4097.

Blood Clotting Factor VIII: From Evolution to Therapy

N. A. Orlova[1,2], S. V. Kovnir[1,2], I. I. Vorobiev[1,2*], A. G. Gabibov[1,2], A. I. Vorobiev[3]
[1]Center "Bioengineering", Russian Academy of Sciences, 60-letija Oktyabrja av., 7/1, Moscow, Russia, 117312
[2]Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Str., 16/10, Moscow, Russia, 117997
[3]Research Center for Hematology, Ministry of Health and Social Development of the Russian Federation, Novij Zykovsky proezd, 4, Moscow, Russia, 125167
*E-mail: ptichman@gmail.com
Received 08.02.2013

ABSTRACT **Recombinant blood clotting factor VIII is one of the most complex proteins for industrial manufacturing due to the low efficiency of its gene transcription, massive intracellular loss of its proprotein during post-translational processing, and the instability of the secreted protein. Improvement in hemophilia A therapy requires a steady increase in the production of factor VIII drugs despite tightening standards of product quality and viral safety. More efficient systems for heterologous expression of factor VIII can be created on the basis of the discovered properties of its gene transcription, post-translational processing, and behavior in the bloodstream. The present review describes the deletion variants of factor VIII protein with increased secretion efficiency and the prospects for the pharmaceutical development of longer acting variants and derivatives of factor VIII.**
KEYWORDS **blood clotting factor VIII; hemophilia A; heterologous protein expression systems.**
ABBREVIATIONS **FVIII – blood clotting factor VIII; BDD – B domain deleted; IU – international unit; FVIII:Ag – concentration of FVIII antigen; FVIII:C – procoagulant activity of FVIII; vWF – von Willebrand factor. Addition of "a" to the number of the corresponding clotting factor denotes the activated factor.**

INTRODUCTION

Blood clotting factor VIII (FVIII) is the nonenzymatic cofactor to the activated clotting factor IX (FIXa), which, when proteolytically activated, interacts with FIXa to form a tight noncovalent complex that binds to and activates factor X (FX). *FVIII* gene defects may cause hemophilia A, the X-linked recessive genetic disorder with an incidence rate of ~ 1 case per 5,000 males. Approximately half of all hemophilia A cases are caused by inversions in intron 22 of the *FVIII* gene [1]; an additional 5% are caused by intron 1 inversions. By November 2012, a total of 2,107 various mutations in the *FVIII* gene with the hemophilia A phenotype had been described in the HAMSTeRS (The Hemophilia A Mutation, Structure, Test and Resource Site) database [2]. By July 2012, a total of 2,537 such mutations had been listed in the CHAMP (The CDC Hemophilia A Mutation Project) database [3].

Continuous substitution therapy using FVIII drugs is the only efficient treatment for hemophilia A. The conventional source of FVIII is donated blood plasma, its supply being limited. Even after a thorough screening of the prepared plasma units and numerous procedures of viral inactivation, the risk of transmission of viral [4, 5] and prion infections [6] remains when plasma is used as a source for production of therapeutic proteins. Recombinant human factor VIII for hemophilia A treatment can be produced using cultured mammalian cells or the milk from transgenic animals.

FUNCTIONS OF FACTOR VIII IN THE HEMOSTATIC SYSTEM

The tight noncovalent FVIIIa–FIXa complex is formed on the phospholipid membrane surface and additionally binds the FX molecule, which is subsequently activated by FIXa. The activated FX leaves the complex and, in turn, triggers the conversion of prothrombin to thrombin (FII to FIIa), which directly converts fibrinogen to fibrin, the major component of blood clots (*Fig. 1*). The ternary complex of clotting factors FIXa, FVIIIa, and FX, bound to the phospholipid membrane, usually is referred to as X-ase or tenase, and is the main element of the positive feedback loop in the blood clotting cascade. A complex that is function-

ally similar to tenase can be described for the extrinsic clotting pathway (FIII, FVIIa, FIX, FX); however, its enzymatic efficiency is considerably lower than that of "intrinsic" tenase. The unique feature of the tenase complex is the high degree of enhancing the catalytic activity (by approximately five orders of magnitude) of low-activity proteinase FIXa by the FVIIIa [7]. This enhancement occurs due to the changes in the active site conformation in FIXa as it binds to FVIIIa [8]. Factor V, homologous to factor VIII, potentiates the activity of FXa within the prothrombinase complex with the coefficient of enhancement of catalytic activity ×240.

The functional activity of FVIII is measured *in vitro* by determination of the clotting time of the blood plasma sample with depleted endogenous FVIII and added FVIII in the solution under study. FVIII stability in the bloodstream was studied in model animals with a defective or deleted *FVIII* gene. The animal models of hemophilia A were discussed in review [9].

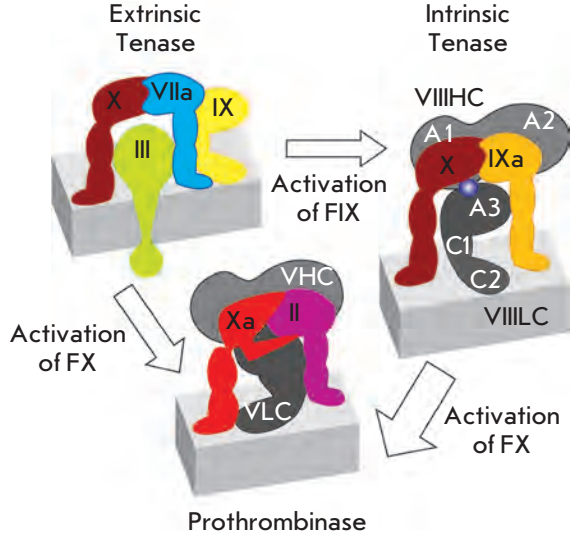


Fig. 1. Tenase complex assembly on the cell membrane. Intact proenzymes of coagulation factors are denoted by Roman numerals, activated enzymatic factors are denoted by the letter "a". Blood clotting factors are bound to the membrane surface, Factor III is the integral membrane protein. VIIIHC – heavy chain of Factor VIII, VIIILC – light chain of Factor VIII, domains A1, A2, A3, C1, C2 of Factor VIII are denoted by white letters. VHC – heavy chain of Factor V, VLC – light chain of Factor V. The thickness of the reaction arrows corresponds to the reaction rates

STRUCTURE OF THE *FVIII* GENE AND ITS EXPRESSION FEATURES

The *FVIII* gene localized on the long arm of the X chromosome occupies a region approximately 186 kbp long and consists of 26 exons (69–3,106 bp) and introns (from 207 bp to 32.4 kbp). The total length of the coding sequence of this gene is 9 kbp [10, 11] (*Fig. 2*). Expression of the *FVIII* gene is tissue-specific and is mostly observed in liver cells [12–14]. The highest level of the mRNA and FVIII proteins has been detected in liver sinusoidal cells [15, 16]; significant amounts of FVIII are also present in hepatocytes and in Kupffer cells (resident macrophages of liver sinusoids).

DOMAIN STRUCTURE

The mature factor VIII polypeptide consists of 2,332 amino acid residues (the maximum length) and includes the A1–A2–B–A3–C1-C2 structural domains [17, 18] (*Fig. 2*). Three acidic subdomains, which are denoted as a1–a3 – A1(a1)–A2(a2)–B–(a3)A3–C1–C2, localize at the boundaries of A domains and play a significant role in the interaction between FVIII and other proteins (in particular, with thrombin). Mutations in these subdomains reduce the level of factor VIII activation by thrombin [19, 20]. There currently is some controversy regarding the accurate definition of the boundaries of FVIII domains; the most common versions are listed in *Table 1*.

Table 1. Domain architecture of factor VIII with indication of the domain borders

A1	a1	A2	a2	B	a3	A3	C1	C2	Reference
1–329	331-372	380-711	700-740	741-1648	1649-1689	1649-2019	2020-2172	2173-2332	[21]
1–336	337-372	372-710	711-740	741-1648	1649-1689	1690-2019	2020-2172	2173-2332	[22]
1–336		372-710		741-1648		1896-2019	2020-2172	2173-2332	[23]
1–336	337-374	375-719	720-740		1649-1690	1691-2025			[24]

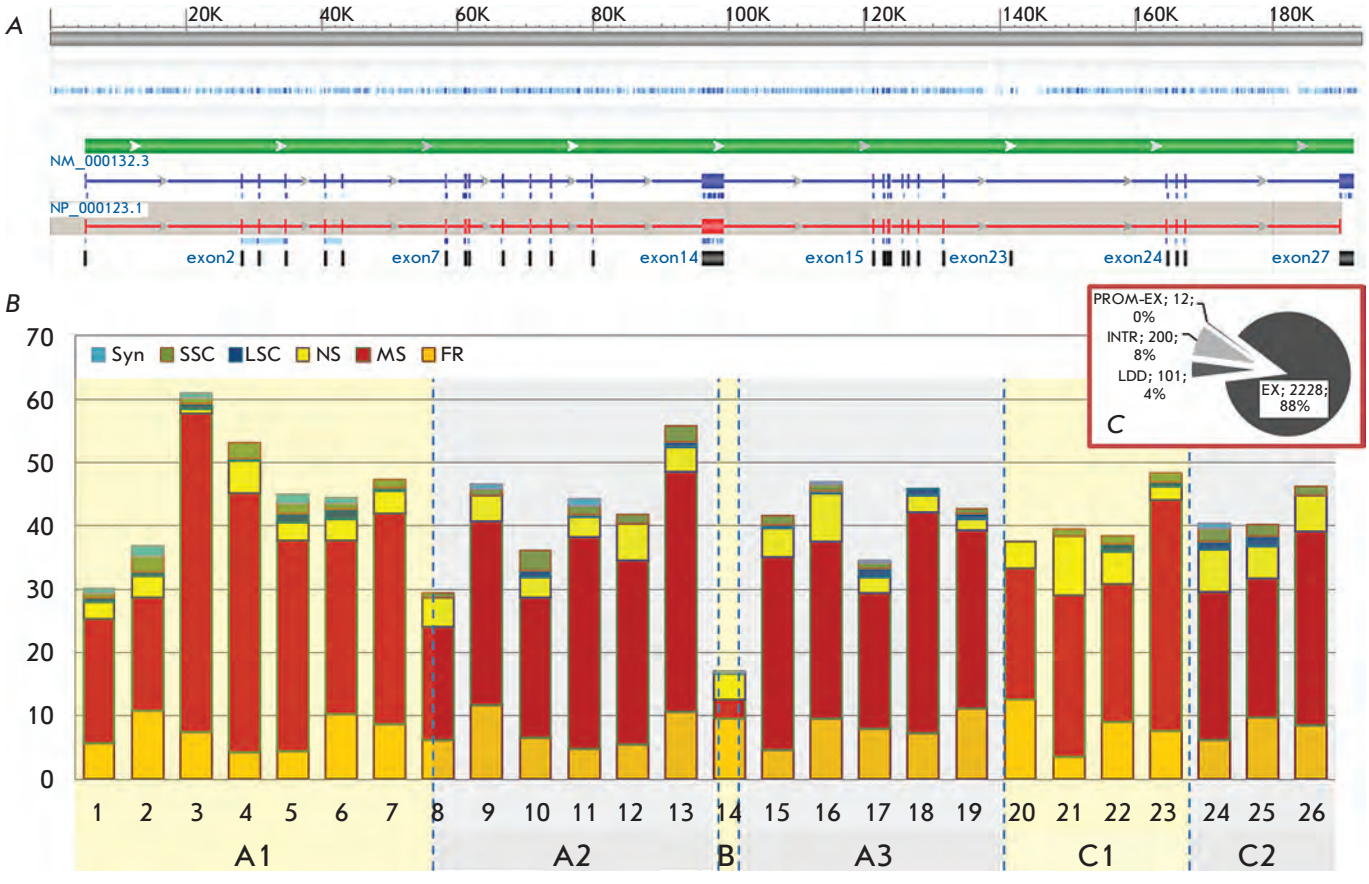


Fig. 2. FVIII gene structure and frequencies of the mutations causing hemophilia A.
Panel A: FVIII gene on the X chromosome, NCBI reference sequence number: NG_011403.1. Transcribed are two products of alternative splicing. Functional protein FVIII is coded by the transcription variant 1, reference number of mRNA NM_000132.3, reference number of the protein NP_000123.1.
Panel B: Variants of mutations in FVIII gene exons according to [3]. The number of different recorded mutations per 100 bp of the coding sequence is shown. Abbreviations: NS – nonsense mutation; MS – missense; FR – frameshift; SSC – small structural change (in-frame, <50 bp); LSC – large structural change (>50 bp). Numbers of histogram columns correspond to the exon numbers, names of protein domains are stated below the exon numbers. Length of exon 1 in mRNA is 314 b; the only coding part of this exon (including the signal peptide), 143 b, was included in the calculations; length of exon 26 in mRNA is 1965 b, the only coding part of this exon – 156 b – was used in the calculations. The lengths of the other exons are as follows: 2 – 122 b, 3 – 123 b, 4 – 213 b, 5 – 69 b, 6 – 117 b, 7 – 222 b, 8 – 262 b, 9 – 172 b, 10 – 94 b, 11 –215 b, 12 – 151 b, 13 – 210 b, 14 – 3106 b, 15 – 154 b, 16 – 213 b, 17 – 229 b, 18 – 183 b, 19 – 117 b, 20 – 72 b, 21 – 86 b, 22 – 156 b, 23 – 145 b, 24 – 149 b, 25 – 177 b.
Panel C: Variants of mutations in the FVIII gene. Abbreviations: LDD – large deletions and duplications in one or multiple domains of FVIII; INTR – distortions in the splice sites; PROM-EX – promoter area mutations and deletions in the promoter area plus the exon; EX – mutations in the exons. Primary data taken from [3], extracted July 18, 2012

The FVIII A domains show 30% homology with each other, the A domains of factor V, and the copper-binding protein of human plasma, ceruloplasmin (*Fig. 3*). The FVIII A1 domain coordinates a copper ion [17, 25–27] (*Fig. 4*). The region 558–565 of the A2 domain determines the binding of factor IXa and its conformational rearrangement within tenase [28] (*Fig. 5*).

The C1 and C2 domains in the light chain of mature FVIII are homologous to the C1 and C2 domains of FV [29], the C-terminal domains of the MFGE8 protein (milk fat globule-EGF factor 8, lactadherin) [30, 31], and the discoidin I fragment [32] (*Fig. 3*). These domains are capable of binding glycoconjugates and acidic phospholipids [33]. The C2 domain in FVIII is also required for binding to the von Willebrand factor (vWF) and ensuring selective interaction with phosphatidylserine in cell membranes [34] (*Fig. 5*).

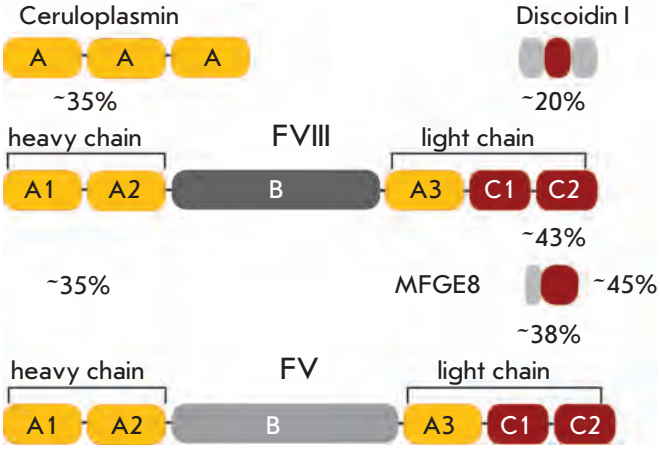


Fig. 3. Domain structure of FVIII homologues. Numbers represent the homology level of amino acids for domain groups. Discoidin I was obtained from *D. discoideum*; all other proteins were obtained from *H. sapiens*

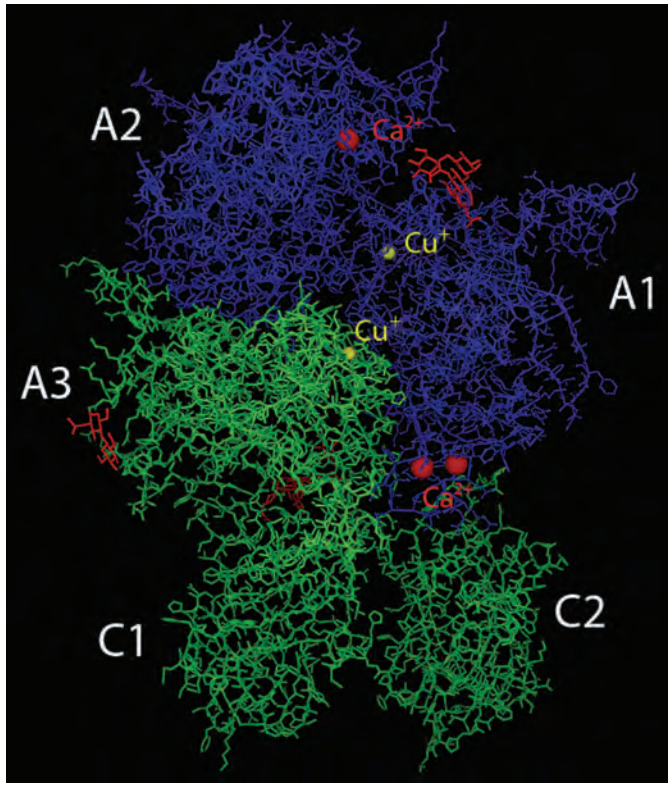


Fig. 4. 3D structure of the FVIII deletion variant according to [42]. Core residues of N-linked glycans are shown in red, FVIII heavy chain is shown in blue, light chain is shown in green

The B-domain encoded by a single long exon is partially removed from the mature protein. The B-domain contains 25 potential N-glycosylation sites, 16–19 of which are occupied and exhibit a significant level of microheterogeneity. The homology of FVIII B-domains in humans and mice is low; however, these domains are highly glycosylated in both species, which can attest to the significance of this modification for post-translational processing of a protein [35].

Proceeding from the significant homology of the factors V and VIII, a hypothesis has been put forward that the evolutionary origin of the *FVIII* gene is connected with duplication. Interestingly, the functional A and C domains of these proteins are conserved, while the similarity between the B-domains is limited to a high degree of glycosylation, which also attests to the functional significance of the high density of oligosaccharide groups in the FVIII B-domain [17, 25–27].

The highly glycosylated B-domain can participate in the intracellular transport of the FVIII precursor and its processing. However, abundant experimental data have demonstrated that deletion of the B-domain region enhances the secretion of functionally active FVIII [36, 37].

COORDINATED METAL IONS

The interaction between the FVIII polypeptide chain and metal ions determines the structural integrity of the mature protein and its cofactor function. The presence of copper ions within FVIII has been demonstrated by atomic adsorption spectrometry; dissociation of the FVIII chains results in complete dissociation of copper ions [38]. In turn, the reassociation of the split FVIII chains is possible only in the presence of copper salts [39]. It has been established by electron paramagnetic resonance (EPR) that the coordinated copper ions within FVIII are reduced to the state of +1 (Cu^+) [40]. The presence of two coordinated copper ions in direct contacts with the H267, C310, H315 and H1954, C2000, and H2005 residues (i.e., two valid type I binding sites of the copper ion) has been detected in crystals of the deletion FVIII variant (BDD SQ variant) [41] (*Fig. 4*). Both copper ion binding pockets localize near the contact surface of the A1 and A3 domains; however, they do not directly participate in the formation of noncovalent bonds between the domains. Simultaneously, evidence of functional significance has been obtained only for the binding site of copper ions in the A1 domain both via point substitution of cysteine residues [40] and by direct monitoring of the coordination of copper ions by FRET [39]. The C310F mutation in the *FVIII* gene [2] causing a severe form of hemophilia A is an additional argument in favor of the physiological significance of the copper-binding site in the A1 domain.

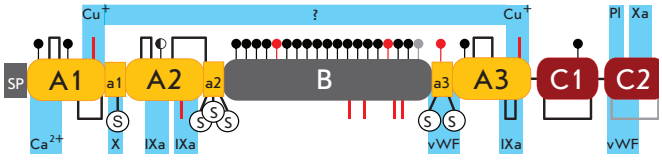


Fig. 5. Post-translational modifications and functional sites of FVIII. N-glycosylation sites are denoted by circles. Filled black circles – occupied sites, half-filled circle – partially occupied site, filled gray circle – presumably occupied site, red circles – unoccupied sites. Disulphide bonds are denoted by brackets, gray bracket – presumably existing disulphide bond. Red vertical lines – reduced Cys residues, the actual state of Cys residues in the B domain is unknown. S inside a circle – sulfated Tyr residues. Light blue marks the areas of interaction with corresponding clotting factors, phospholipids (Pl), von Willebrand factor (vWF), and copper ions (Cu^+). SP – signal peptide and propeptide

Both copper ions and calcium or manganese ions are required to recover the procoagulant activity of FVIII during chain dissociation–re-association [43, 44]. Calcium or manganese ions do not affect chain dimerization, but they ensure that the heterodimer FVIII molecule has an active conformation [39] as they bind to the sites located on both protein chains [45, 46]. The major Ca^{2+}-binding site localizes in the A1 domain (region 108–124) [45] and is homologous to the corresponding site in the FV molecule [47]. It has been ascertained by alanine scanning that Ca^{2+} binding is mediated by the D116, E122, D125, and D126 residues, while the interaction with Mn^{2+} is mediated by the D116 and D125 residues [48].

POST-TRANSLATIONAL PROCESSING OF THE FVIII PRECURSOR PROTEIN

FVIII is synthesized in the liver, which has been supported by the fact that liver transplantation can cure hemophilia A. When isolating and purifying liver cell populations, it has been ascertained that secretion of significant amounts of FVIII (0.07 IU/million cells/day) is observed in primary cultures of liver sinusoidal endothelial cells [15]. No successful attempts to immortalize cultured liver sinusoidal endothelial cells have been documented thus far; hence, all the experimental data on the features of FVIII biosynthesis have been obtained using heterologous expression systems that are usually characterized by artificially increased productivity [49].

Translocation of the growing FVIII polypeptide chain to the lumen of the endoplasmic reticulum (ER), processing of the signal 19-amino-acid-long peptide, and the primary events of disulfide bond formation and attachment of the high-mannose nuclei of N-linked oligosaccharides to the FVIII chain seem not to limit the total rate of its biosynthesis and have been thoroughly described in [47]. Meanwhile, the subsequent events of modifying oligosaccharide chains, disulfide isomerization, and folding of FVIII molecules may overload the corresponding enzyme groups in the cell and activate the systems responsible for retention of the incorrectly processed proteins in ER or the recycling systems for these proteins. The total rate of FVIII secretion is believed to be limited by translocation of the FVIII precursor from the ER to the Golgi apparatus; the FVIII polypeptide can remain in the ER for 15 min to several days.

N-GLYCOSYLATION

After the primary N-glycosylation of the FVIII chain and cleavage of the two first glucose residues from oligosaccharide groups by glucosidases I and II (GTI, GTII), polypeptide FVIII binds to the lectins calnexin (CNX) and calreticulin (CRT), which prevent the secretion of an immature protein [50] (*Fig. 6*). After the third glucose residue is eliminated, the protein is normally released from its complex with CNX and CRT and is transferred to the Golgi apparatus. Meanwhile, the unfolded or incorrectly folded FVIII remains in the ER, where it undergoes re-glucosylation by the UGT enzyme (UDP-glucose:glycoprotein glucosyltransferase) [51]. Next, it binds to CNX and CRT and undergoes shortening of GTII again (the so-called calnexin cycle).

The incorrectly folded FVIII molecules, along with the other proteins, are transferred from the ER to cytosolic proteasomes via the ERAD (ER-associated degradation) pathway; the elimination of the polypeptide from the calnexin cycle is mediated by the specialized EDEM protein [28]. Indeed, it has been demonstrated in pulse-chase experiments with proteasomes inactivated by lactacystin [50] that a significant portion of FVIII undergoes degradation via the ERAD pathway instead of being translocated to the Golgi apparatus; however, in those experiments proteasome inactivation increased the amount of intracellular FVIII but not its concentration in the culture medium. Thus, the ERAD pathway alone does not eliminate significant amounts of FVIII from the lumen of the ER and cannot be the reason for the limited transfer of FVIII from the ER to the Golgi apparatus. Since the major fraction of N-linked oligosaccharides in the FVIII molecule is localized in the B-domain, the deletion FVIII variants are less susceptible to retention in the ER during the calnexin cycle, which partially explains their increased level of secretion.

DISULFIDE BOND FORMATION

According to results of 3D modeling of FVIII and the results of most experimental studies, the FVIII molecule contains eight disulfide bonds: two in each A domain, one in each C domain, and three reduced Cys residues, one each in the A1, A2, and A3 domains (*Fig. 5*). There are no conclusive data on the state of the cysteine residues in the B-domain. Seven out of eight disulfide bonds localize within a polypeptide globule, while the C1899–C1903 bond (A3 domain) is exposed on the surface. When conducting a series of substitutions of cysteine residues by serine or glycine residues, S.W. Pipe *et al.* found that all the seven non-exposed disulfide bonds are required to maintain the structural integrity of the FVIII molecule, while the removal of the 1899–1903 bond improves secretion of FVIII twofolds without affecting its functional activity [52]. It is rather possible that the removal of the only exposed disulfide bond results in suppression of FVIII retention in the ER occurring due to the translocation control by the disulfide isomerases [53]. However, the specific mechanism of such control with respect to FVIII and the participating proteins has hardly been studied.

FOLDING AND INTERACTION WITH ER CHAPERONES

Factor VIII in the lumen of the ER forms a strong complex with the major ER chaperone GRP78 (glucose-regulated protein MW 78.000), which is also known as BiP (immunoglobulin-binding protein) [54] and is one of the key components of the UPR (unfolded protein response) signaling pathway. BiP synthesis is typically induced when cells experience glucose starvation, during N-glycosylation inhibition, and in the presence of incorrectly folded proteins in the ER [55] (in particular, during FVIII overexpression) [56]. It should be mentioned that overexpression of human FVIII in cultured cells results in total activation of UPR, which manifests itself not only as a positive regulation of BiP, but also as activation of the *ERSE* gene and an increase in the level of splicing of *XBP1* mRNA [57]. Thus, there can be other chaperones (in addition to BiP) that initiate the activation of UPR when large amounts of FVIII are transported into the lumen of the ER.

The BiP–polypeptide complex exhibits ATPase activity; hydrolysis of ATP is required to ensure disintegration of the complex. The isolation of FVIII from BiP and secretion require exceptionally high ATP expenditures [58].

Unlike FVIII, its homologue factor V does not interact with BiP. The site of FVIII binding to BiP (a hydrophobic β-sheet within the A1 domain lying near the C310 residue, which is a component of the type I copper-ion binding site) was identified using a series of chimeric FVIII–FV proteins [40, 59]. BiP forms direct contacts with hydrophobic amino acids, and the point mutation F309S inside this β-sheet results in a threefold increase in FVIII secretion, which correlates with reduced ATP expenditure [59]. Since the F309 residue is adjacent to C310, the key residue in the copper-coordination site in the A1 domain, one can assume that BiP interacts in the attachment of copper ions to FVIII as well.

Approximately one-third of FVIII molecules in the ER are aggregated to noncovalent multimers. The replacement of the region 227–336 of FVIII by the homologous region of FV reduces its degree of aggregation and affinity to BiP, as well as increases secretion [60]. The functional value of the FVIII-BiP complex presumably consists in the retention of FVIII in the ER, rather than in ensuring efficient FVIII folding prior to its translocation to the Golgi apparatus.

TRANSPORT OF FVIII FROM ER TO THE GOLGI APPARATUS

Transport of the FVIII polypeptide from the ER to the Golgi apparatus occurs via the ER-Golgi intermediate compartment (ERGIC) (*Fig. 6*). FVIII and FV are recruited to this compartment by binding to the transmembrane protein (cargo receptor) ERGIC-53, which is also known as LMAN1 (lectin, mannose-binding, 1) and ensures mannose-selective, calcium-dependent binding and transport of glycoproteins from the ER to the Golgi apparatus [61].

The mutations resulting in the loss of LMAN1 function or disturbing the interaction between LMAN1 and the component of the transport complex MCFD2 (multiple coagulation factor deficiency protein 2) cause inherited coagulopathy, a combined deficiency of factor V and factor VIII [62–64]. The FVIII level in the plasma of patients with mutant LMAN1 decreases to 5–30% of its normal level [65].

The transport of four proteins (FV, FVIII and the lysosomal proteins, the cathepsins catC and catZ) through the intermediate compartment with the participation of the LMAN1–MCFD2 has been confirmed by the cross-linking method [66, 67]. A number of other proteins interact with LMAN1, but not with MCFD2, as they are transported [68]. It has been detected by cross-linking that 5–20% of the total intracellular FVIII localizes in the complex with LMAN1 and MCFD2 [69]. Calcium ions are required for the FVIII complex with both partners to form; meanwhile, the FVIII–MCFD2 can be formed independently of LMAN1. It remains unclear whether direct interaction between FVIII and LMAN1 (which has been observed for the cathepsins catC and catZ [70]) is possible, or whether the FVIII–LMAN1 complex forms only with the participation of MCFD2. The specific FVIII motif, which can be recognized by

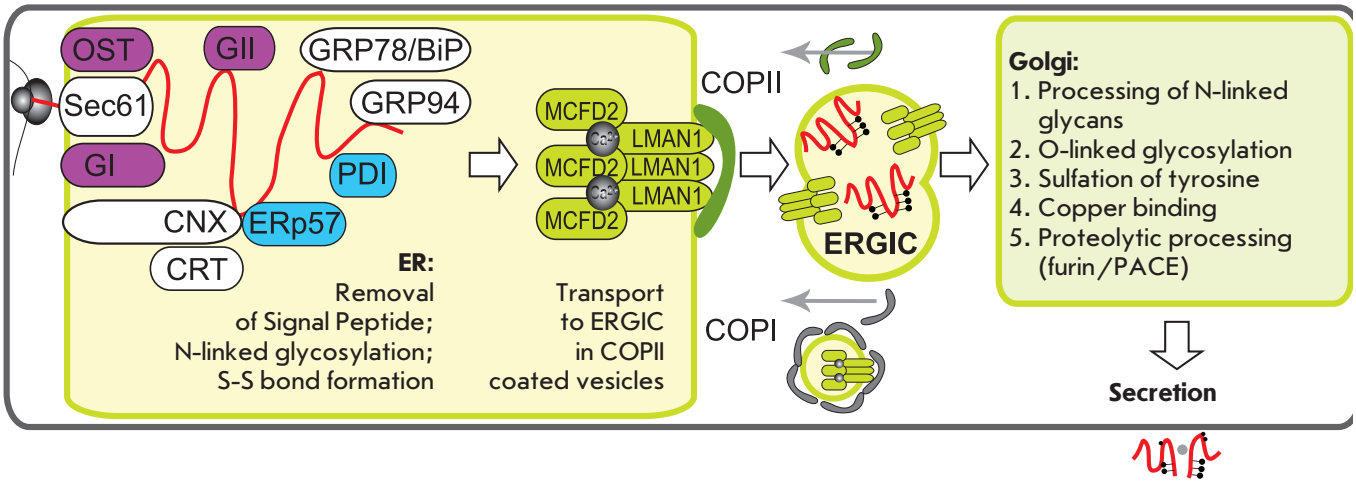


Fig. 6. Intracellular traffic of the FVIII polypeptide to be secreted. OST – Oligosaccharyltransferase, Sec61 – membrane protein translocator, GI and GII – glucosidases I and II; CNX – calnexin, CRT – calreticulin, GRP78/BiP – glucose regulated protein 78 / immunoglobulin binding protein; GRP94 – glucose regulated protein 94; PDI, ERp57 – disulphide isomerases; MCFD2 – multiple coagulation factors deficiency protein 2; LMAN1 – mannose-binding lectin 1; COPI, COPII – vesicle coat proteins I and II; ERGIC – ER-Golgi intermediate compartment

the cargo receptor, has not been identified. This motif is supposed to contain a conformational epitope and a carbohydrate moiety (to ensure that only the correctly folded and post-translationally modified proteins can be transported). The binding motif of LMAN1 has been experimentally identified in the catZ proenzyme molecule; it contains several adjacent N-glycans [37]; however, there are no regions homologous to it in FVIII and FV.

FV and FVIII have similar domain structures, including B domains, which are non-homologous but contain numerous N-glycosylation sites in both cases. Since the B-domain-deleted FVIII is characterized by reduced efficiency in the binding to the LMAN1–MCFD2 complex, a hypothesis has been put forward that LMAN1 predominantly interacts with the B domains [71]. Meanwhile, the blockage of N-glycosylation does not stop the formation of the FVIII–cargo receptor complex [69]; i.e., it is not only the carbohydrate moiety of the molecule, but also the polypeptide chain that participate in the interaction between FVIII and the cargo protein.

The LMAN1–MCFD2 complex specifically recruits FVIII and FV from the ER to COPII (coat protein II) vesicles, which are separated from the ER to subsequently bind to ERGIC (*Fig. 6*). The mechanism of release of the FVIII polypeptide from the LMAN1–MCFD2 complex during its transport has not been elucidated. It is thought to be released due to the change in the local pH value and calcium concentration [71]. The COPII proteins return to the ER outside the vesicles; the ERGIC complexes are subjected to retrograde transport within the COPI vesicles. FVIII seems to be further transported inside vesicles of unknown composition or via transport containers associated with microtubules. FVIII molecules emerge in the Golgi apparatus via the formation of new *cis*-Golgi cisterns [71–73].

PROCESSING OF FVIII IN THE GOLGI APPARATUS

High-mannose N-glycans of the FVIII molecule are modified in the Golgi apparatus; O-glycosylation and sulfation of tyrosine residues occurs in the *trans*-Golgi. Six active sites of sulfation of tyrosine residues at positions 346, 718, 719, 723, 1664, and 1680 have been detected in human FVIII; they predominantly localize near the acidic subdomains a1, a2, a3 and surround the points where FVIII is cleaved by thrombin. All the six sulfation sites are required to ensure the full activity of factor VIII; the inhibition of sulfation has resulted in a fivefold decrease in the functional activity of FVIII [74]. It has also been demonstrated that sulfation of the Y1680 residue is required to ensure efficient interaction between FVIII and the von Willebrand factor. The natural mutation of Y1680F manifests itself as moderate hemophilia A. In patients carrying this mutation, FVIII retains its normal activity level but is characterized by a decreased half-life value [75]. R.J. Kaufman *et al.* [76] employed site-directed mutagenesis to demonstrate that the presence of sulfated residues at positions 346 and 1664 increases the rate of FVIII activation by thrombin, while sulfation of residues 718, 719, and 723

increases the specific activity of FVIIIa in the tenase complex. The necessity of sulfating residue 1680 so that the complex with the von Willebrand can form has also been confirmed in [76].

The final stage of FVIII processing in the *trans*-Golgi prior to the secretion involves limited proteolysis of the single-chain precursor at residues R1313 and R1648, giving rise to a light and a heavy chain [22]. Both sites of proteolytic processing correspond to the Arg-X-X-Arg motif, which can be cleaved by protease furin/PACE (paired basic amino acid cleavage enzyme); however, it remains unclear what particular signaling protease of the PACE family is responsible for FVIII processing.

FACTOR VIII IN THE BLOODSTREAM

Mature natural FVIII, which can occur in one of several forms with a molecular weight of 170–280 kDa, is present in the blood plasma at a concentration of 0.1–0.2 μg/ml [77]. Almost all the FVIII in plasma is complexed with a chaperone, the von Willebrand factor, which is secreted by vascular endothelial cells. The FVIII regions responsible for binding to this chaperone have been mapped in the light chain: in the acidic subdomain a3 [78], domains C2 [34, 79] and C1 [80]. The von Willebrand factor stabilizes FVIII in the blood stream and is its key regulator as it allows thrombin to activate the bound FVIII [36, 81, 82] and impedes cleavage of the molecules of nonactivated FVIII by the proteases FXa [83] and activated protein C (APC) [84–86]. Furthermore, vWF prevents the nonspecific binding of FVIII to the membranes of vascular endothelial cells [61] and platelets [87]. It has been demonstrated in *in vitro* experiments that vWF facilitates the association of FVIII chains and the retention of procoagulant activity in the conditioned medium of cells producing FVIII [44, 49]. Similar data have been obtained for re-association of FVIII chains in solution [43, 44]. The dissociation constant of the vWF–FVIII complex is 0.2–0.4 nM; practical equilibrium during the *in vitro* complex formation is attained within seconds [61, 88, 89].

In a significant number of hemophilia A patients, inhibitors of injected exogenous FVIII emerge in the bloodstream, blocking its procoagulation activity [90]. Cases of development of acquired hemophilia A with the normal *FVIII* gene due to the emergence of antibodies against autologous FVIII have also been reported [91]. The etiology of emergence of inhibitory antibodies has not been elucidated; some particular correlations between the emergence of inhibitory antibodies and the HLA haplotype [92] or the nature of the mutation of the factor VIII gene [93] were recently found. IgG antibodies are the predominant class of inhibitory antibodies [94]. Alloantibodies have been found to bind predominantly to the A2 or C2 domains of factor VIII, thus impeding its interaction with factor FIX, whereas autoantibodies are likely to bind to the FVIII C2 domain, which presumably results in blockage of its interaction with phospholipids and vWF [95]. Moreover, it has been demonstrated that anti-factor VIII antibodies can specifically hydrolyze FVIII [96], the proteolytic activity of alloantibodies being in direct proportion to the level of the FVIII inhibitor [97].

INTERACTION OF FVIII WITH FIXA, FX, AND PHOSPHOLIPIDS

The protein–protein interactions between FVIII (or FVIIIa) and FIXa within tenase are ensured by two different regions; the main contacting surface localizing on the FVIII light chain (*Fig. 1*). The affinity of the free light chain to FIXa is similar to that of the full-length FVIII (K_d 14–50 nM [67, 98], while K_d of the full-length FVIII is ~ 2–20 nM [99, 100]). The main site of FVIII–FIXa interaction is a short peptide 1803–1810 [101]; the second site of FVIII–FIXa interaction is the 558–565 region [67]. The area of direct interaction between FVIII and FX has been found in the acidic C-terminal subdomain of the A1 domain (337–372) [71, 77]; however, this interaction most probably has no significant effect on the function of the tenase complex. The presence of phospholipids is required for FVIII to perform its cofactor function [7, 102, 103]. FVIII interacts *in vivo* with the phospholipids of activated platelets and damaged endothelial cells. Both non-enzymatic cofactors of the coagulation system, FVIII and FV, have been shown to bind to phosphatidylserine [38, 104]. Factor VIII predominantly binds to micelles containing 15–25% of phosphatidylserine, with the dissociation constant reaching 2–4 nM [89, 99, 102]. Platelet activation can increase the phosphatidylserine content in the platelet membrane from 2 to 13%, thus attracting FVIII. FVIII activation increases its affinity to phospholipids by 10 times [105]. The binding site of phospholipids localizes in the FVIII light chain within the C2 domain [106] (*Fig. 5*).

FVIII ACTIVATION AND FVIIIA INACTIVATION

In vivo activation of FVIII is induced by thrombin or FXa (*Fig. 7*) and involves the introduction of proteolytic breakdowns at several points. When activating FVIII by thrombin, the breakdowns are introduced at positions R372, R740, and R1689 [107] and result in removal of the B domain, in cleavage of the heavy chain into the A1 and A2 domains that remain noncovalently bound, and in elimination of the short acidic region a3 preceding the A3 domain. In a number of publications, the region a3 is referred to as the FVIII activation peptide; however, efficient activation of FVIII cannot be

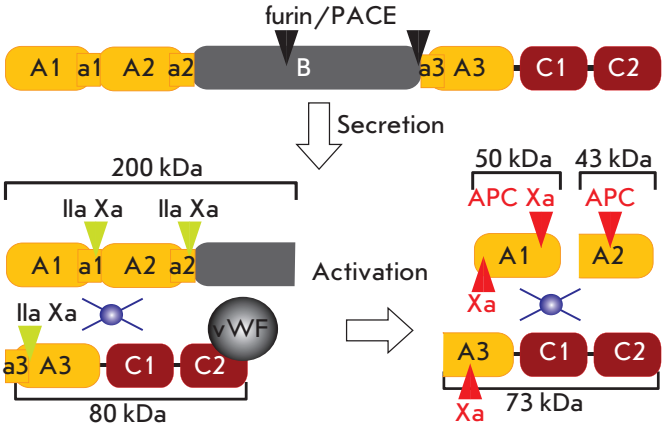


Fig. 7. Proteolytic processing of FVIII. Black triangles – sites of processing by the PACE/furin family proteases; green triangles – sites of processing during activation; red triangles – sites of processing during inactivation; and the blue circle represents the coordinated copper ion(s)

reduced simply to elimination of region a3 from the molecule. FVIII activation by FXa results in cleavage of the FVIII polypeptide chain at the sites specified above and in two or three additional breakdowns at positions R1721, R336, and K36 [63, 107]. Efficient interaction between FVIII and thrombin is mediated by sulfated tyrosine residues in FVIII, while FVIII activation by factor FXa is almost insensitive to the Y → F substitution at sulfation sites [76]. FVIIIa activated by FXa forms tenase that is considerably less productive as compared to that formed by thrombin-activated FVIIIa [108]. Thus, FVIII activation by FXa can be regarded as a side process of FVIIIa inactivation.

FVIIIa inactivation can occur spontaneously and be reduced to the dissociation of the A2 domain of the heavy chain (which is not covalently bound to the remaining FVIII domains) from the FVIIIa molecule [109, 110]. Two specific inactivators of FVIIIa are currently distinguished: APC and FXa. APC cleaves FVIIIa at positions R562 and R336 [71], disintegrating the region of interaction between FVIII and FIX and destabilizing the interaction between the A1 and A2 domains. FXa-induced inactivation of FVIIIa seems to occur *in vivo* more rapidly than the APC-induced inactivation does. It involves the introduction of breakdowns at positions R336 and K36 [73], resulting in destabilization of the A1 domain and in the accelerated dissociation of the unbound A2 domain.

ELIMINATION OF FVIII FROM THE BLOODSTREAM

The FVIII–vWF complex is mainly eliminated from the bloodstream by the specialized clearance receptor LRP (low-density lipoprotein receptor-related protein) that localizes on the hepatocyte membrane [111–113]. A 3.3-fold increase in the half-life of FVIII was observed in *in vivo* experiments in mice when blocking LRP by the receptor-associated 39 kDa protein (RAP) that binds to LRP with high affinity [111]. Three sites take part in the interaction between the FVIII and LRP: the ones in the C2 domain [112], in the A3 domain (1804–1834) [101], and in the A2 domain (484–509) [111]. Multiple sites of FVIII–LRP interaction ensure efficient elimination of unbound chains and the cleaved A2 domain from the bloodstream. The presence of vWF in complex with FVIII within the C2 domain prevents interaction between this domain and LRP, which reduces affinity to LRP by 90% [112]. The *in vivo* interaction of FVIII and its fragments with LRP is mediated by heparin sulfate proteoglycans (HSPG), which interact with the region 558–565 in the A2 domain [114].

RECOMBINANT FVIII FOR HEMOPHILIA TREATMENT

Pharmaceuticals based on recombinant full-length FVIII were developed almost simultaneously by the biotechnology companies Genetics Institute and Genentech using the *FVIII* gene expression systems in CHO and BHK cells [66, 115] and were approved to be marketed in 1992–1993 with the international nonproprietary name "octocog alfa." The recombinant FVIII secreted by CHO cells along with the recombinant vWF [49, 77] was produced under the trade names Recombinate® and Bioclate®. The recombinant FVIII secreted by BHK cells into a culture medium containing natural vWF [115] is known under the trade names Kogenate® and Helixate® (*Table 2*).

Today, there are three generations of pharmaceuticals based on recombinant blood-clotting factors [68]: the first-generation drugs contain human serum albumin and contact with animal-derived compounds during the production process; the excipients list of the second-generation drugs contains no albumin; in the third-generation drugs, contact with animal-derived compounds and components of the donated plasma is ruled out during the entire production process. The minimization of the use of plasma components and animal-derived proteins can potentially reduce the risk of transmission of viral and prion infections [116]. No confirmed evidence of transmission of infectious agents when using the first- and second-generation drugs based on recombinant FVIII has been documented thus far.

The production process of full-length recombinant FVIII comprises several stages of ion-exchange chromatography, affinity chromatography using immobilized monoclonal antibodies, and viral inactivation by solvent/detergent treatment or via pasteurization in

Table 2. Drugs based on recombinant FVIII

Name	Kogenate®, Helixate®	Kogenate FS®, Kogenate Bayer®, Helixate FS®, Helixate NexGen®	Recombinate®, Bioclate®	Advate®	ReFacto®	Xyntha®, ReFacto AF®
Manufacturer	Bayer Healthcare		Baxter		Pfizer	
Generation	1	2	1	3	2	3
Market approval in USA	1993	2000	1992	2003	2000	2008
Producer cell line	BHK		CHO		CHO	
Heterologous genes	FVIII		FVIII, vWF		FVIII BDD SQ	
Proteins in the culture medium	Human plasma proteins		BSA, aprotinin	–	BSA	–
Immunoaffinity chromatography	+		+		+	–
Stabilizing agent	HSA	Sucrose	HSA	Mannitol, trehalose	Sucrose	
Viral inactivation	SD		Pasteurization	SD	SD	SD, NF

Note. HSA – human serum albumin, BSA – bovine serum albumine, SD – treatment with a solvent and a detergent , NF – nanofiltration.

the presence of a detergent [67, 117]. The recombinant FVIII drugs inevitably contain trace amounts of proteins from the producer cells and murine IgG; thus, the emergence of antibodies against these impurities in patients and the effect of these antibodies on the effectiveness of therapy have been studied during clinical trials. The antibodies formed in most patients; however, the relationship between the immune response to impurity proteins and the effectiveness of therapy has not been elucidated [68].

B-DOMAIN DELETED RECOMBINANT FVIII

The natural FVIII circulating in the bloodstream contains multiple forms of the truncated B domain, which are formed by proteolysis of a full-length two-chain molecule. The procoagulant properties of these FVIII variants have no significant differences [118]; thus, variants of the recombinant FVIII with targeted deletion of the B domain have been obtained and characterized in a number of studies. The region encoding the amino acid residues 760–1639 (i.e., virtually the entire B domain) was deleted from FVIII cDNA in [118] (*Table 3*).

The procoagulant activity level of FVIII in a conditioned medium for COS-1 cells transfected with a plasmid with cDNA of the deleted FVIII form (LA-VIII variant) was approximately tenfold higher than that of the control cell line transfected with a similar plasmid coding the full-length FVIII. It was ascertained in further studies that the LA-VIII variant is similar to the natural FVIII in terms of its biochemical properties except for the increased sensitivity of the LA-VIII light chain to thrombin cleavage [36]. In more efficient cell lines producing FVIII, B-domain deletion (LA-VIII variant) resulted in a 17-fold increase in the level of FVIII mRNA; however, the concentration of the secreted product increased by only 30% [37]. Similar data were obtained for a FVIII delta II variant, which contained a deletion of the amino acids 771 to 1666 [121]; the level of FVIII secretion in BHK cells reached 0.6 IU/ml, while the secreted product mostly contained

Table 3. Deletion variants of FVIII

Sequence	Variant name	Deletion region	Source
…EP**R**SFSQNSRHPSTRQKQFNATTIPENDIEKTD…SQNPPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	Natural single chain	–	[11]
…EP**R**-COOH(?) NH2-EITRTTLQSDQEEIDYDDTI…	Natural 90+80, seq.1	741?–1648	[119]
…EP**R**-COOH(?) NH2-LQSDQEEIDYDDTI…	Natural 90+80, seq.2	741?–1654	[119, 120]
…EP**R**SFSQNSRHPSTRQKQFNA----------NPPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	LA-VIII	760–1639	[118]
…EP**R**SFSQNSRHPSTRQKQFNATTIPENDIEKTD----------DTI…	deltaII	771–1666	[121]
…EP**R**----------EITRTTLQSDQEEIDYDDTI…	d741-1648	741–1648	[74]
…EP**R**SFSQNSRHPSTRQKQFNATTI…LLR---DPL…QNPPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	90-142-80	797–1562	[122]
…EP**R**SFSQN-COOH NH2-DPL…QNPPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	–	746–1562	[123]
…EP**R**SFSQN----------PPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	BDD SQ	746–1639	[119]
…EP**R**SFSQNSRHPS----------QNPPVLKRHQ**R**EITRTTLQSDQEEIDYDDTI…	N0, N8	751–1637	[124, 125]
…EP**R**SFSQNS----------QAYRYRR----------Q**R**EITRTTLQSDQEEIDYDDTI…	human-cl	747–1640**	[126]
…EP**R**SFSQNSRHPSTRQKQFNATTI…SST----------**R**EITRTTLQSDQEEIDYDDTI…	226aa/N6	968–1647	[127]

Note: sign ↓ marks the sites of FVIII processing, cleavage occurs after the amino acid pointed with an arrow; * – minor processing site of natural FVIII; ** – region 1641–1647 is replaced by an artificial region.

the single-stranded 170 kDa form and two additional forms of the heavy chain (120 and 90 kDa) [128]. A similar predominant accumulation of the single-chain FVIII form was also detected in the cases of deletion of regions 741–1648, 741–1668, and 741–1689 [74]. The partial deletion of the amino acid residues 797 to 1562 in the B domain (90-142-80 variant) gave rise to a fully active FVIII [122]; however, this variant when injected into rabbits caused the emergence of specific antibodies against the protein linker region [129], which could potentially have increased the frequency of formation of inhibitory antibodies when used in therapy.

A several-fold increase in the product secretion level as compared to the full-length FVIII has been observed in most heterologous expression systems of B-domain-deleted FVIII. Such change has not been observed for the deletion of the 741–1668 region; however, when the producing CHO cell line was replaced by SK-HEP-1 cells, the secretion level of the deletion FVIII variant increased to 3.5 IU/million cells/day [130]. It is interesting to note that individual expression of the genes coding the truncated heavy and the full-length light FVIII chains with signal peptides of the heavy chain in CHO cells allows one to attain an expression level of 15 IU/million cells/day [123]. Contamination of the product with the nonprocessed form of the light chain (90 kDa), which seems to carry the C-terminal fragment of the B domain, remains the only (but unavoidable) limitation of this FVIII expression system.

In order to design a recombinant FVIII that would carry neither the B domain nor a non-natural linker region between the heavy and light chains, it is necessary to determine the points of deletion onset and termination, which would make it possible to preserve the availability of the dominant processing sites of natural FVIII for its "minimal" two-chain form (R740 and R1648). When conducting a systematic, exhaustive search, P. Lind *et al.* [119] found that a high level of processing of a single-stranded FVIII precursor at these sites was attained when the amino acid residues

746–1639 were deleted (BDD SQ variant). Meanwhile, no significant proteolytic chain cleavage at other sites has been observed. In this deletion variant, the linking point between the polypeptides of the heavy and light FVIII chains lay in the middle of the separating 14-residue-long linker peptide. The predominant cleavage of the precursor after residues R1648 and S1657 (i.e., the coincidence of the N-terminal light chain region and the natural sequence) was confirmed by N-terminal sequencing of the light chain of the secreted FVIII BDD SQ. The heavy chain of FVIII BDD SQ carried the C-terminal 729–740 region and, partially, the linker peptide [120]. Thus, the BDD SQ variant allows one to reconstruct most accurately the "minimal" two-chain FVIII variant that is present in the bloodstream. Variants of the pharmaceutical compositions of the purified FVIII BDD SQ (a solution with a high saccharide content [131] and albumin-free lyophilizate [132]) have been obtained. The lyophilized form of BDD SQ was stable for two years [133] and was used in subsequent clinical trials, which demonstrated its pharmacological effectiveness and safety [134]. Meanwhile, the half-life time of the deletion variant FVIII BDD SQ (international nonproprietary name "moroctocog alfa") in the bloodstream was slightly lower as compared to the full-length FVIII obtained from donor plasma.

The original version of the industrial production process of pharmaceutical-grade FVIII BDD SQ (trade name ReFacto®) comprised the cultivation of producer cells based on CHO in perfusion bioreactors and the purification of the target protein by five chromatography steps [135]. Viral inactivation was performed by the solvent/detergent treatment of the intermediate product. There are no data on the productivity of the industrially used clonal cell line secreting FVIII BDD SQ. The productivity of a similar cell line based on CHO cells, which had been obtained independently, was 0.5–2.0 IU/ml when grown in a serum-free medium without any induction [136] and up to 10 IU/ml when the product expression was induced by sodium propionate or sodium butyrate. The industrial process of FVIII BDD SQ production was subsequently modified: immunoaffinity chromatography on monoclonal antibodies was replaced by a safer stage comprising affinity purification on an immobilized short peptide [137] (trade names Xyntha®, ReFacto AF®) (*Table 2*).

The effectiveness and safety of FVIII BDD SQ drugs has been confirmed by clinical trials [138–140]; however, the subsequent meta-analysis of the data of numerous post-marketing studies has cast doubt. R.A. Gruppo *et al.* [141] have demonstrated that prophylactic use of FVIII BDD SQ instead of the full-length FVIII results in a statistically significant increase in the risk of bleeding under prophylaxis. The resistance to slight variations in the initial data (robustness) of the employed meta-analysis method has been discussed in a separate publication and has been confirmed for a wide range of coefficients for recalculating the number of cases of bleeding observed in different studies [142]. The inaccurately measured level of FVIII activity in patients' plasma as a result of using different coagulometry techniques and standards of procoagulant activity of FVIII [143] could have been one of the reasons for the reduced effectiveness of FVIII BDD SQ in preventive treatment. The reduced half-life time of FVIII BDD SQ [144] as a result of accelerated inactivation of activated FVIII BDD SQ by the activated proteins C and FXa can be presumably considered to be another reason [145].

As for another important safety indicator of FVIII drugs–the risk of emergence of inhibitors–the data for FVIII BDD SQ were rather inconsistent. In some studies, the frequency of emergence of inhibitors was similar for all the variants of recombinant FVIII [146–148], while in other studies an increased risk of inhibitor emergence was observed for FVIII BDD SQ [149]. Since the probability of emergence of inhibitors varies widely depending on the mutation type that caused hemophilia, certain HLA genotypes, and specific features of the substitution therapy, the data obtained in different medical centers may differ to a significant extent [150].

A variant similar to the BDD SQ variant of FVIII with the deleted residues 751–1637 (variant N8, *Table 3*) was obtained via expression in CHO cells [125] and used in clinical trials, which have shown the bioequivalence of N8 to the comparator drug, the full-length recombinant FVIII, after a single administration in a group consisting of 23 individuals [151]. Another industrially applicable gene expression system of the deletion FVIII variant was created using a human HEK293F cell line; the *FVIII* gene contained not the direct deletion of a region in the B domain but a substitution of the 747–1648 region by the "non-natural" peptide QAYRYRRQ [126] (variant human-cl, *Table 3*). A hypothesis has been put forward that the presence of the processing sites of proteinase Kex2/furin in the artificial linker peptide will allow one to increase the level of processing of the single-chain FVIII. However, the degrees of processing of the single-chain form for BDD SQ and human-cl variants turned out to be almost identical. The expression system of the deletion FVIII variant, which seemed to exhibit the highest efficiency at that moment, was obtained using special hybrid human cells HKB11. The specific productivity of clonal lines for the FVIII variant with the 90-142-80 deletion (*Table 3*) was equal to 5–10 IU/million cells/day [88, 152].

The investigation of the gene expression levels in FVIII variants with the deletion of only the C-terminal

fragments of the B domain has demonstrated that the FVIII variant carrying the first 226 amino acids of the B domain and six N-glycosylation sites (variant 226aa/N6) is secreted by CHO cells fivefold more efficiently as compared with FVIII with the deletion of the full-length B domain. This is attributable to the improved transport of the precursor protein from the ER to the Golgi apparatus [127] (as compared with the full-length form) and to a decrease in adsorption of the secreted FVIII onto the membrane surface of producer cells [124] (as compared with the regular deletion variants). The productivity of the cells producing the 226/N6 variant based on CHO cells reached 11 IU/ml without inducing target gene expression and 15.7 IU/ml when using a serum-containing medium [52].

FACTORS LIMITING THE EFFICIENCY OF THE HETEROLOGOUS EXPRESSION SYSTEMS FOR FVIII

A significant decrease in the transcription level of hybrid genes containing an open reading frame (ORF) of *FVIII* were first described in studies devoted to the cultivation of retroviral vectors [153, 154]. The presence of the ORF of *FVIII* did not affect the level of transcription initiation; however, a 1.2 kbp ORF fragment reduced the efficiency of transcription elongation by 30–100 times. The observed effect depended on orientation and, to a significant extent, on the position. It was delocalized as the removal of different parts of the ORF fragment under study resulted in partial recovery of the transcription elongation level. An orientation-independent 305-bp-long transcriptional silencer was later detected in another region of FVIII ORF [155]; its activity was suppressed by sodium butyrate [156], which made it possible to enhance the level of FVIII secretion in the cell culture approximately sixfold. The presence of this controllable element for regulating the transcription level in *FVIII* ORF impedes the production of effective therapeutic viral vectors but has some advantages during FVIII biosynthesis in cell cultures. Stress induced by processing of the FVIII precursor can be limited by suppression of *FVIII* gene transcription in the dividing culture, while only the dense non-dividing cell culture is exposed to stress during subsequent induction of *FVIII* expression by adding sodium butyrate.

The codon optimization of the coding region of *FVIII* mRNA was studied in [157]. The replacement of some codons by ones characterized by the highest frequency for *H. sapiens* and the simultaneous elimination of internal TATA-boxes, CHI sites, ribosomal binding sites, cryptic splice sites, etc. from the encoding mRNA region increased the FVIII:C level by 7–30 times. An increase in the ratio between the level of FVIII antigen and its procoagulant activity (the ratio between FVIII:Ag and FVIII:C) from 1.27± 0.3 to 2.35 ± 0.49 was simultaneously observed for the deletion variant FVIII BDD SQ, which can attest to the fact that the practical threshold of productivity of cell line 293 was reached and the nonfunctional protein had appeared in the culture medium.

The FVIII precursor in the lumen of ER forms a stable complex with BiP, the major chaperone [54] and one of the key participants in the UPR signaling pathway. Overexpression of the *FVIII* gene induces transcription of the *BiP* gene [56]; the intracellular level of BiP is proportional to the level of factor FVIII secretion within an appreciably wide range [13, 158]. The suppression of BiP expression by short hairpin RNAs (shRNAs) increased the secretion level of human FVIII [159] by ~ 2 times, while the number of copies of *FVIII* mRNA simultaneously decreased by ~ 65%. A similar effect was observed for the overexpression of the *XBP1* gene, whose product also participates in UPR.

The overexpression of the chaperone Hsp70 was found to reduce induction of apoptosis in a dense culture of FVIII-producing BHK cells and to increase the level of FVIII secretion [160]. Similar data were obtained for overexpression of the anti-apoptotic genes *Aven* and *E1B-19K* [72]. Meanwhile, no significant changes in the level of *Hsp70* expression and the anti-apoptotic genes *Bcl-2* and *Bcl-xL* among clones with different levels of FVIII secretion have been observed for the FVIII-producing human hybrid cell line HKB11 [152]. These data attest to the fact that this pathway of anti-apoptosis re-engineering of FVIII producers can be efficient only for a super-dense BHK cell culture.

The suppression of oxidative stress in the ER (as well as the UPR and apoptosis induced by it) in CHO cells overexpressing FVIII by antioxidants was demonstrated in [51]. The addition of the antioxidant butylated hydroxyanisole to the culture medium simultaneously with sodium butyrate (an agent inducing *FVIII* gene expression) enabled a fourfold increase in the secretion level of the full-length FVIII. Manifestations of oxidative stress were also observed for the FVIII with a fully deleted B-domain but not for the variant 226/N6.

The increase in FVIII secretion by suppressing the intensity of UPR, oxidative stress, and apoptosis of the producer cells may be associated with the changes in the level of FVIII adsorption onto the outer cell membrane. In CHO cells secreting the full-length FVIII, FVIII concentration in a serum-free supernatant increased by a factor of 4 after porcine vWF had been added to the culture medium [77]; i.e., at least three quarters of the total FVIII secreted in the absence of vWF remained bound to the cell membrane. FVIII

predominantly binds to phospholipid membranes containing phosphatidylserine. For the membranes of activated platelets, the dissociation constants of FVIII, FVIII BDD SQ, and FVIIIa are equal to 10.4, 5.1, and 1.7 nM, respectively [161]. An increase in FVIII adsorption on the membrane of apoptotic cells (which also contains an elevated fraction of phosphatidylserine) has been shown by flow cytometry. The suppression of apoptosis in producer cells via overexpression of the *Hsp70* gene resulted in a drop in the level of absorption of the full-length FVIII on the membrane and an increase in its concentration in the culture medium [161]. In expression systems of the FVIII gene variants with B-domain deletion, adsorption on the cell membrane is more pronounced and can exceed 90% for the variant N0 (which is identical to N8). Partial (instead of complete) deletion of the B domain reduces adsorption to ~ 50%, while the total level of FVIII expression decreases almost twofold [124]. The loss of secreted FVIII N0 on the membrane of the producer cells can also be reduced via the inhibition of its interaction with phosphatidylserine by adding vWF, annexin V, or o-phospho-*L*-serine into the culture medium [162].

TRANSGENIC ORGANISMS

The expression systems of recombinant proteins of the hemostatic system, which are based on cultured cells, can potentially be completely replaced with technologies that allow production of these proteins in the milk of transgenic animals. Antithrombin III exemplifies the successful implementation of such an approach. For its production, transgenic goats have been developed with antithrombin III productivity of over 1 g/l; industrial processes for purifying the target protein have been elaborated [163]. The level of FIX production in the milk of transgenic pigs was considerably lower [164], which is typically attributed to the insufficient degree of γ-carboxylation of the product. Polypeptide stability and accurate processing of the single-chain form to the heterodimer are considered to be the main factors limiting the productivity of transgenic animals that secrete FVIII in the milk [165]. Functionally active full-length FVIII was produced in the milk of mice [166], rabbits [167], sheep [168], and pigs [169]; however, in all cases the level of product secretion was of no practical interest (*Table 4*). When the deletion variant FVIII 226/N6 was used and the von Willebrand factor was coexpressed, the level of FVIII:C in the milk of transgenic mice reached 678 IU/ml, which attests to the possibility of producing large transgenic animals that can secrete industrially significant amounts of FVIII in their milk. However, the simultaneous introduction of two transcriptionally active transgenes into the cattle genome will require a significant effort.

VARIANTS OF LONG-ACTING FVIII

Despite the fact that the risk of transmission of viral infections has been considerably reduced thanks to drugs

Table 4. Main properties of transgenic animals capable of secreting FVIII in milk

Name	Milk volume, l*,**	Estimated maximum productivity, g*,**	FVIII:Ag, μg/ml	FVIII:C, IU/ml	Specific activity, IU/mg, [for plasma 5 000+]	Productivity per doe per year, mg/IU	Comments & references
Mouse	0.0015	0.01–0.02	50.21	13.41	267	0.075 / 20	Fl [166]
			122–183	555–678	3705–4549	0.183–0.275 / 833–1017	Variant 226/N6 + vWF [165]
Rabbit	2–5	20	0.117***	0.521	4500***	0.234–0.585 / 1042–2605	Fl [167]
Sheep	200–500	2500	N/A	0.02–0.03****	N/A	N/A / 4000–15000	Fl [168]
Pig	200–400	1500	2.66	0.62	233	532–1064 / 124 000–248 000	Fl [169]

Fl – full-length FVIII.
*Per doe per year.
**According to [170] and [171].
***In the paper cited, the FVIII:Ag content is given in μg/ml, which seems to be a misprint. The data is listed in Table as ng/ml.
****In the paper cited, FVIII:C was measured vs. the standard of natural FVIII and expressed as ng/ml; the data are listed in Table as IU/ml under an assumption that the specific activity of the standard compound is 5,000 IU/mg.

based on recombinant FVIII, modern substitution therapy with hemophilia A remains far from perfect, as it denies a decent quality of life to hemophiliacs. The reasons limiting the effectiveness of substitution therapy include the immunogenicity of the injected FVIII, resulting in the appearance of inhibitory antibodies and in FVIII instability in the bloodstream, which requires injections every 2–3 days when FVIII is used as a preventive agent. Since the risk of emergence of anti-FVIII inhibitor antibodies is determined, among other factors, by the number of injections made, production of FVIII derivatives with a prolonged half-life would help to solve both these problems.

The following trends in the study of FVIII derivatives with a prolonged effect can be mentioned: production of FVIII conjugates to hydrophilic polymers, the introduction of point mutations, production of fusion proteins, and design of hybrid human–porcine FVIII variants.

The conjugation of therapeutic proteins to polyethylene glycol (PEG) molecules usually makes it possible to increase their circulation time in the bloodstream by several times. In some cases, it also reduces immunogenicity and prevents proteolytic degradation. Meanwhile, the non-specific attachment of PEG molecules to the therapeutic protein can cause its inactivation [172]. For FVIII, blockage of its interaction with vWF can also result in a significant decrease in its half-life time. The feasibility of site-specific attachment of PEG molecules to non-paired cysteine residues inserted at various domains of the deletion variant FVIII BDD SQ via site-directed mutagenesis was studied [173] (*Fig. 8*). For the FVIII variant containing two additional cysteine residues at positions 491 and 1804, which are conjugated to 60 kDa PEG molecules (variant BDD FVIII 60 kDa di-PEG-L491C/K1804C, *Fig. 8A*), an increase in the survival rate of knockout mice after the transection of the tail vein was observed (from 60% for the intact FVIII to 86%) [173]. The non-directed attachment of PEG molecules to the full-length FVIII at lateral amino groups of lysine residues has also allowed to obtain a conjugate (code BAX 855, Fig. 8B) characterized by an average degree of attachment of PEG residues = 2:1, unchanged procoagulant activity, and a lifetime in the bloodstream increased by approximately two times [174].

Since the intact FVIII circulates in the bloodstream within a multimeric high-molecular-weight complex with vWF, there is little promise in increasing the half-life time of the recombinant FVIII by designing proteins fused with long-acting proteins of the blood plasma (e.g., with serum albumin). Meanwhile, linking FVIII in frame to the domains of other proteins, which specifically protect them against elimination from circulation, can considerably increase FVIII stability. Thus, in experiments on knockout mice and dogs with a model of hemophilia A, the protein FVIII–immunoglobulin Fc-region (FVIII-Fc, *Fig. 8C*) provided protection against uncontrollable bleeding twice as long-lasting as the intact FVIII [175]. The prolonged effect of FVIII-Fc was entirely determined by the interaction with the neonatal Fc-receptor (FcRn). Clinical trials of FVIII-Fc conducted with 16 patients have demonstrated that the time of retention of FVIII-Rc in the bloodstream (time between the injection of the drug and the decrease in the FVIII:C level below 1%) increases by a factor of 1.53–1.68 [176]. It should be noted that the prophylactic use of FVIII drugs usually includes three injections per week. Meanwhile, the duration of the effect of FVIII needs to be increased at least twice so that a single injection per week is sufficient [177]. Hence, an increase in the duration of the effect of modified FVIII variants by approximately 1.5 times as compared to the intact FVIII can reduce the risk of bleeding to a certain extent for the existing prophylaxis regimens (the so-called "third-day problem"), but it does not allow one to make the injections less frequent.

The alteration of the properties of FVIII via point mutagenesis has been described in several independent studies; however, none of the mutant proteins (muteins) has undergone clinical trials. The introduction of three point substitutions R336I/R562K/R740A to the gene of the deletion variant FVIII 741–1689 (variant IR8, *Fig. 8D*) has enabled the production of a protein characterized by normal procoagulant activity, loss of its affinity to vWF, and high resistance to proteolytic inactivation of FVIIIa by the activated protein C [178]. However, no significant differences in the termination of bleeding in knockout mice have been observed when using this gene variant for targeted FVIII expression on the platelet membrane [179].

The introduction of a pair of cysteine residues to the spatially juxtaposed regions of the A2 and A3 domains forms a disulfide bond between them, which stabilizes the activated FVIII, thus increasing its procoagulant activity [180]. Muteins of the deletion FVIII variant containing the cysteine pair C664–C1826 or C662–C1828 (*Fig. 8D*) in *in vitro* experiments exhibited specific activity tenfold higher than that of the intact FVIII [181].

The stability of FVIIIa can also be enhanced by substituting the amino acids on the interface surfaces between the A2, A1, and A3 domains. The point substitution of E1984V (*Fig. 8F*) resulted in an increase in the lifetime of the activated FVIII by 4–8 times, while its normal procoagulant activity was retained [182].

Most inhibitor antibodies emerging in patients with hemophilia A are oriented toward the epitopes within the A2 and C2 domains. Anti-A2-domain antibod-

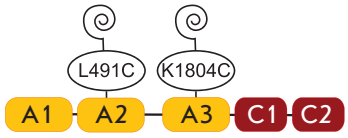
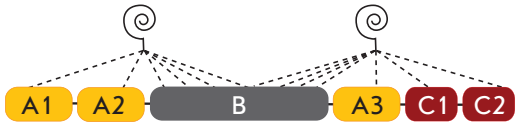
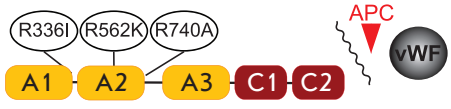
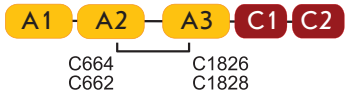
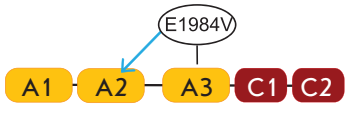
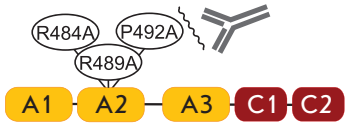
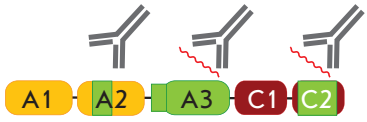
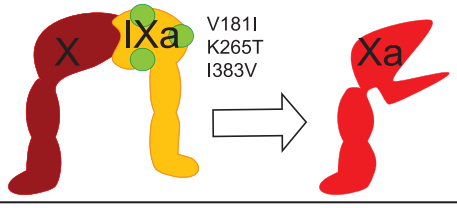
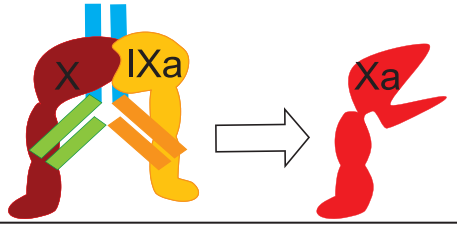
A		**BDD FVIII 60 kDa di-PEG- L491C/ K1804C** Increase in half-life and absence of loss of specific activity due to site-specific modification
B		**BAX-855** Increase in half-life and absence of loss in specific activity due to prevalent modification of B domain residues
C		**FVIII-Fc** Retardation in the bloodstream by the reversible interaction with the FcRn receptor
D		**IR8** FVIIIa is resistant to inactivation by APC but cannot bind to vWF
E		**FVIII C664—C1826, C662—C1828** Stabilization of FVIIIa by the absence of inactivation due to dissociation of the A2 domain
F		**E1984V** Stabilization of the FVIIIa by the increase in the affinity of A2 domain to the A3 domain
G		**R489A R484A P492A** Loss of dominant epitope of alloantibodies and no changes in functional properties
H		**HP32** Replacement of major alloantibodies epitopes by homologous areas from porcine FVIII
I		**FIX V181I/ K265T/ I383V** Increase in enzymatic activity of the FIXa sufficient for the direct activation of the FX
J		**hBS23** Mimicking of the FVIIIa by the bi-specific antibody against FIXa and FX. The complex is anchored on the cell membrane via the interaction of the Gla-domains in FIXa and FX with the membrane

Fig. 8. FVIII long-acting variants and functional mimetics. Spirals represent the covalently attached PEG groups; dashed lines – unknown conjugation sites; arrows – noncovalent interaction; and wavy lines – blocking of interactions. Protein parts from porcine FVIII are shown in green in panel *H*

ies mostly interact with a short region, 484–508; thus, the replacement of several amino acids in this region of FVIII can reduce its immunogenicity. It turned out to be sufficient to introduce the triple substitution R484A/R489A/P492A (*Fig. 8G*) to reduce the average inhibitor level from 670 to 310 Beteshda U/ml in knockout mice that had received seven sequential FVIII injections at an interval of 14 days [183].

It was demonstrated in *in vitro* experiments using a number of hybrid FVIII molecules with the deleted B domain 741–1648, which contained alternating fragments of porcine and human FVIII, that the replacement of the 484–508 fragment of the A2 domain of human FVIII by a homologous fragment of porcine FVIII and complete replacement of the human FVIII A3- and C2 domains by the corresponding domains of porcine FVIII (variant HP32, *Fig. 8H*) allow one to produce a FVIII molecule that is resistant to the inhibitor effect of most antibodies isolated from hemophilia A patients [184]. For this reason, the porcine recombinant FVIII with a deleted B domain (variant OBI-1), which is currently undergoing clinical trials [185], can be potentially replaced by a hybrid molecule that inhibits lower immunogenicity as compared to xenogenic porcine FVIII [186] but carries no immunodominant epitopes of human FVIII.

FUNCTIONAL ANALOGUES OF FVIII

Since the function of FVIIIa can be confined to increasing FIXa activity, the high-activity analogue of FIXa, which can produce a sufficient amount of thrombin, will ensure efficient blood clotting without the participation of FVIII. Unlike FVIII, FIX modified in this manner can also be used to treat the inhibitor form of hemophilia A. The introduction of a gene therapy plasmid encoding mutein FIX with the triple substitution V181I, K265T, and I383V into *FVIII* gene knockout mice improved the blood clotting indicators [187], thus attesting to the fact that the hemostatic function in patients with hemophilia A can be recovered without using FVIII drugs (*Fig. 8I*). The tenase complex can also be reconstructed by replacing the FVIIIa molecule by a bispecific antibody against FIXa and FX. This antibody was selected among 40,000 molecules composed of fragments of monoclonal antibodies against FIX and FX via high-throughput screening [188]. After the optimization of the structure of the leading molecules, the bispecific humanized antibody hBS23 was obtained, which is capable of increasing the catalytic efficiency of FIXa by 19,800 times (272,000 times for FVIII) due to the 20-fold reduction in K_m of the reaction of FX activation and a 1,000-fold increase in k_{cat} (*Fig. 8J*). A single injection of 0.3 mg/kg hBS23 to macaques used as a model of acquired hemophilia A ensured virtually identical bleeding control as therapy using porcine FVIII.

CONCLUSIONS

Efficient bleeding control in hemophilia A patients is based on continuous substitution therapy using FVIII preparations and prophylaxis of bleeding in pediatric practice. Since the current state of studies and elaboration of long-acting FVIII derivatives provides no grounds to expect considerably improved drugs in the near future, the development of new cell lines producing FVIII (with allowance for the accumulated knowledge on the FVIII structure and the factors affecting the levels of its biosynthesis and secretion) may enable a several-fold increase in its production. It can be assumed that a simple increase in the production volume of third-generation biosimilar drugs based on recombinant FVIII will make it possible to increase the volume of substitution therapy for hemophilia A, while the current costs remain unchanged (i.e., to improve patients' quality of life and increase their longevity without reallocating the limited healthcare resources). ●

REFERENCES

1. Antonarakis S.E., Kazazian H.H.,Tuddenham E.G. // Hum. Mutat. 1995. V. 5. № 1. P. 1–22.
2. Kemball-Cook G. Haemophilia A Mutation Database. University College London. 1994. *http://hadb.org.uk*
3. Payne A.B., Miller C.H., Kelly F.M., Michael Soucie J., Craig Hooper W. // Hum. Mutat. 2013. V. 34. № 2. P. E2382–2391.
4. Blumel J., Schmidt I., Effenberger W., Seitz H., Willkommen H., Brackmann H.H., Lower J., Eis-Hubinger A.M. // Transfusion. 2002. V. 42. № 11. P. 1473–1481.
5. Yokozaki S., Fukuda Y., Nakano I., Katano Y., Toyoda H., Takamatsu J. // Blood. 1999. V. 94. № 10. P. 3617.
6. Evatt B.L. // Haemophilia. 1998. V. 4. № 4. P. 628–633.
7. Brooks A.R., Harkins R.N., Wang P., Qian H.S., Liu P., Rubanyi G.M. // J. Gene Med. 2004. V. 6. № 4. P. 395–404.
8. Kim M., O'Callaghan P.M., Droms K.A., James D.C. // Biotechnol. Bioengin. 2011. V. 108. № 10. P. 2434–2446.
9. Sabatino D.E., Nichols T.C., Merricks E., Bellinger D.A., Herzog R.W., Monahan P.E. // Prog. Mol. Biol. Transl. Sci. 2012. V. 105. P. 151–209.
10. Gitschier J., Wood W.I., Goralka T.M., Wion K.L., Chen E.Y., Eaton D.H., Vehar G.A., Capon D.J., Lawn R.M. // Nature. 1984. V. 312. № 5992. P. 326–330.
11. Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L., Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C., et al. // Nature. 1984. V. 312. № 5992. P. 342–347.
12. Cho M.S., Tran V.M. // Virology. 1993. V. 194. № 2. P. 838–842.
13. Dunne M.J., Kane C., Shepherd R.M., Sanchez J.A., James R.F., Johnson P.R., Aynsley-Green A., Lu S., Clement J.P., et al. // N. Engl. J. Med. 1997. V. 336. № 10. P. 703–706.

14. Stel H.V., van der Kwast T.H., Veerman E.C. // Nature. 1983. V. 303. № 5917. P. 530–532.
15. Do H., Healey J.F., Waller E.K., Lollar P. // J. Biol. Chem. 1999. V. 274. № 28. P. 19587–19592.
16. Hollestelle M.J., Thinnes T., Crain K., Stiko A., Kruijt J.K., van Berkel T.J., Loskutoff D.J., van Mourik J.A. // Thromb. Haemost. 2001. V. 86. № 3. P. 855–861.
17. Vehar G.A., Keyt B., Eaton D., Rodriguez H., O'Brien D.P., Rotblat F., Oppermann H., Keck R., Wood W.I., Harkins R.N., et al. // Nature. 1984. V. 312. № 5992. P. 337–342.
18. Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J., Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L., et al. // Nature. 1984. V. 312. № 5992. P. 330–337.
19. Kjalke M., Heding A., Talbo G., Persson E., Thomsen J., Ezban M. // Eur. J. Biochem. 1995. V. 234. № 3. P. 773–779.
20. Donath M.J., Lenting P.J., van Mourik J.A., Mertens K. // Eur. J. Biochem. 1996. V. 240. № 2. P. 365–372.
21. Colman R.W. Hemostasis and Thrombosis: Basic Principles and Clinical Practice. Baltimore:Lippincott Williams & Wilkins. 2012. 1578 p.
22. Lenting P.J., van Mourik J.A., Mertens K. // Blood. 1998. V. 92. № 11. P. 3983–3996.
23. Andrew P.J., Auer M., Lindley I.J., Kauffmann H.F., Kungl A.J. // FEBS Lett. 1997. V. 408. № 3. P. 319–323.
24. Pemberton S., Lindley P., Zaitsev V., Card G., Tuddenham E.G., Kemball-Cook G. // Blood. 1997. V. 89. № 7. P. 2413–2421.
25. McCune J.S., Liles D., Lindley C. // Pharmacotherapy. 1997. V. 17. № 4. P. 822–826.
26. Garrigues C., Loubiere P., Lindley N.D., Cocaign-Bousquet M. // J. Bacteriol. 1997. V. 179. № 17. P. 5282–5287.
27. Harvey M.J., Cauvin A., Dale M., Lindley S., Ballabio R. // J. Small Anim. Pract. 1997. V. 38. № 8. P. 336–339.
28. Molinari M., Calanca V., Galli C., Lucca P., Paganetti P. // Science. 2003. V. 299. № 5611. P. 1397–1400.
29. Liu M. L., Shen B.W., Nakaya S., Pratt K.P., Fujikawa K., Davie E.W., Stoddard B.L., Thompson A.R. // Blood. 2000. V. 96. № 3. P. 979–987.
30. Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M., Ceriani R.L. // Cancer Res. 1991. V. 51. № 18. P. 4994–4998.
31. Lin L., Huai Q., Huang M., Furie B., Furie B.C. // J. Mol. Biol. 2007. V. 371. № 3. P. 717–724.
32. Johnson J.D., Edman J.C., Rutter W.J. // Proc. Natl. Acad. Sci. USA. 1993. V. 90. № 12. P. 5677–5681.
33. Stubbs J.D., Lekutis C., Singer K. L., Bui A., Yuzuki D., Srinivasan U., Parry G. // Proc. Natl. Acad. Sci. USA. 1990. V. 87. № 21. P. 8417–8421.
34. Shima M., Scandella D., Yoshioka A., Nakai H., Tanaka I., Kamisue S., Terada S., Fukui H. // Thromb. Haemost. 1993. V. 69. № 3. P. 240–246.
35. Elder B., Lakich D., Gitschier J. // Genomics. 1993. V. 16. № 2. P. 374–379.
36. Pittman D.D., Alderman E.M., Tomkinson K.N., Wang J.H., Giles A.R., Kaufman R.J. // Blood. 1993. V. 81. № 11. P. 2925–2935.
37. Pittman D.D., Marquette K.A., Kaufman R.J. // Blood. 1994. V. 84. № 12. P. 4214-4225.
38. Cho M.S., Yee H., Brown C., Jeang K.T., Chan S. // Cytotechnology. 2001. V. 37. № 1. P. 23–30.
39. Wakabayashi H., Koszelak M.E., Mastri M., Fay P.J. // Biochemistry. 2001. V. 40. № 34. P. 10293–10300.
40. Tagliavacca L., Moon N., Dunham W.R., Kaufman R.J. // J. Biol. Chem. 1997. V. 272. № 43. P. 27428–27434.
41. Ngo J.C., Huang M., Roth D.A., Furie B. C., Furie B. // Structure. 2008. V. 16. № 4. P. 597–606.
42. Shen B.W., Spiegel P.C., Chang C.H., Huh J.W., Lee J.S., Kim J., Kim Y.H., Stoddard B.L. // Blood. 2008. V. 111. № 3. P. 1240–1247.
43. Fay P.J. // Arch. Biochem. Biophys. 1988. V. 262. № 2. P. 525–531.
44. Wise R.J., Dorner A.J., Krane M., Pittman D.D., Kaufman R.J. // J. Biol. Chem. 1991. V. 266. № 32. P. 21948–21955.
45. Wakabayashi H., Schmidt K.M., Fay P.J. // Biochemistry. 2002. V. 41. № 26. P. 8485–8492.
46. Wakabayashi H., Zhen Z., Schmidt K.M., Fay P.J. // Biochemistry. 2003. V. 42. № 1. P. 145–153.
47. Mukhopadhyay C.K., Mazumder B., Lindley P.F., Fox P.L. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. № 21. P. 11546–11551.
48. Wakabayashi H., Freas J., Zhou Q., Fay P.J. // J. Biol. Chem. 2004. V. 279. № 13. P. 12677–12684.
49. Kaufman R.J., Wasley L.C., Dorner A.J. // J. Biol. Chem. 1988. V. 263. № 13. P. 6352–6362.
50. Pipe S.W., Morris J.A., Shah J., Kaufman R.J. // J. Biol. Chem. 1998. V. 273. № 14. P. 8537–8544.
51. Malhotra J.D., Miao H., Zhang K., Wolfson A., Pennathur S., Pipe S.W., Kaufman R.J. // Proc. Natl. Acad. Sci. USA. 2008. V. 105. № 47. P. 18525–18530.
52. Selvaraj S.R., Scheller A.N., Miao H.Z., Kaufman R.J., Pipe S.W. // J. Thromb. Haemost. 2012. V. 10. № 1. P. 107–115.
53. Hagiwara M., Nagata K. // Antioxid. Redox Signal. 2012. V. 16. № 10. P. 1119–1128.
54. Liles D., Landen C.N., Monroe D.M., Lindley C.M., Read M.S., Roberts H.R., Brinkhous K.M. // Thromb. Haemost. 1997. V. 77. № 5. P. 944–948.
55. Malhi H., Kaufman R.J. // J. Hepatol. 2011. V. 54. № 4. P. 795–809.
56. Dorner A.J., Wasley L.C., Kaufman R.J. // J. Biol. Chem. 1989. V. 264. № 34. P. 20602–20607.
57. Brown H.C., Gangadharan B., Doering C.B. // J. Biol. Chem. 2011. V. 286. № 27. P. 24451–24457.
58. Murphy M.E., Lindley P.F., Adman E.T. // Protein Sci. 1997. V. 6. № 4. P. 761–770.
59. Swaroop M., Moussalli M., Pipe S.W., Kaufman R.J. // J. Biol. Chem. 1997. V. 272. № 39. P. 24121–24124.
60. Tagliavacca L., Wang Q., Kaufman R.J. // Biochemistry. 2000. V. 39. № 8. P. 1973–1981.
61. Moussalli M., Pipe S.W., Hauri H.P., Nichols W.C., Ginsburg D., Kaufman R.J. // J. Biol. Chem. 1999. V. 274. № 46. P. 32539–32542.
62. Zakas P.M., Gangadharan B., Almeida-Porada G., Porada C.D., Spencer H.T., Doering C.B. // PLoS One. 2012. V. 7. № 11. P. e49481.
63. Nogami K., Wakabayashi H., Schmidt K., Fay P.J. // J. Biol. Chem. 2003. V. 278. № 3. P. 1634–1641.
64. Zhang B., McGee B., Yamaoka J.S., Guglielmone H., Downes K.A., Minoldo S., Jarchum G., Peyvandi F., de Bosch N.B., Ruiz-Saez A., et al. // Blood. 2006. V. 107. № 5. P. 1903–1907.
65. Fay P.J., Smudzin T.M. // J. Biol. Chem. 1992. V. 267. № 19. P. 13246–13250.
66. White G.C., 2nd, McMillan C.W., Kingdon H.S., Shoemaker C.B. // N. Engl. J. Med. 1989. V. 320. № 3. P. 166–170.
67. Lee D.C., Miller J.L., Petteway S.R., Jr. // Haemophilia. 2002. V. 8. Suppl 2. P. 6–9.
68. Yoshioka A. // Textbook of Hemophilia. New York:Wiley-Blackwell, 2010. P. 146-152.

69. Zhang B., Kaufman R.J., Ginsburg D. // J. Biol. Chem. 2005. V. 280. № 27. P. 25881–25886.
70. Jankowski M.A., Patel H., Rouse J.C., Marzilli L.A., Weston S.B., Sharpe P.J. // Haemophilia. 2007. V. 13. № 1. P. 30–37.
71. Regan L.M., O'Brien L.M., Beattie T.L., Sudhakar K., Walker F.J., Fay P.J. // J. Biol. Chem. 1996. V. 271. № 8. P. 3982–3987.
72. Nivitchanyong T., Martinez A., Ishaque A., Murphy J.E., Konstantinov K., Betenbaugh M.J., Thrift J. // Biotechnol. Bioeng. 2007. V. 98. № 4. P. 825–841.
73. Nogami K., Wakabayashi H., Fay P.J. // J. Biol. Chem. 2003. V. 278. № 19. P. 16502–16509.
74. Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P., Mertens K., van Mourik J.A. // J. Biol. Chem. 1991. V. 266. № 2. P. 740–746.
75. Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K., Kazazian H.H., Jr., Antonarakis S.E. // Genomics. 1990. V. 6. № 1. P. 65–71.
76. Michnick D.A., Pittman D.D., Wise R.J., Kaufman R.J. // J. Biol. Chem. 1994. V. 269. № 31. P. 20095–20102.
77. Adamson R. // Ann. Hematol. 1994. V. 68. Suppl 3. P. S9–14.
78. Leyte A., Verbeet M.P., Brodniewicz-Proba T., van Mourik J.A., Mertens K. // Biochem. J. 1989. V. 257. № 3. P. 679–683.
79. Saenko E.L., Scandella D. // J. Biol. Chem. 1997. V. 272. № 29. P. 18007–18014.
80. Jacquemin M., Benhida A., Peerlinck K., Desqueper B., Vander Elst L., Lavend'homme R., d'Oiron R., Schwaab R., Bakkus M., Thielemans K., et al. // Blood. 2000. V. 95. № 1. P. 156–163.
81. Hamer R.J., Koedam J.A., Beeser-Visser N.H., Sixma J.J. // Eur. J. Biochem. 1987. V. 167. № 2. P. 253–259.
82. Hill-Eubanks D.C., Lollar P. // J. Biol. Chem. 1990. V. 265. № 29. P. 17854–17858.
83. Koedam J.A., Hamer R.J., Beeser-Visser N.H., Bouma B.N., Sixma J.J. // Eur. J. Biochem. 1990. V. 189. № 2. P. 229–234.
84. Koedam J.A., Meijers J.C., Sixma J.J., Bouma B.N. // J. Clin. Invest. 1988. V. 82. № 4. P. 1236–1243.
85. Fay P.J., Coumans J.V., Walker F.J. // J. Biol. Chem. 1991. V. 266. ; 4. P. 2172–2177.
86. Nogami K., Shima M., Nishiya K., Hosokawa K., Saenko E.L., Sakurai Y., Shibata M., Suzuki H., Tanaka I., Yoshioka A. // Blood. 2002. V. 99. № 11. P. 3993–3998.
87. Nesheim M., Pittman D.D., Giles A.R., Fass D.N., Wang J.H., Slonosky D., Kaufman R.J. // J. Biol. Chem. 1991. V. 266. № 27. P. 17815–17820.
88. Cho M.S., Yee H., Brown C., Mei B., Mirenda C., Chan S. // Biotechnol. Prog. 2003. V. 19. № 1. P. 229–232.
89. Gilbert G.E., Drinkwater D., Barter S., Clouse S.B. // J. Biol. Chem. 1992. V. 267. № 22. P. 15861–15868.
90. Gilles J.G., Jacquemin M.G., Saint-Remy J.M. // Thromb. Haemost. 1997. V. 78. № 1. P. 641–646.
91. Franchini M., Lippi G. // Blood. 2008. V. 112. № 2. P. 250–255.
92. Tuddenham E.G., McVey J.H. // Haemophilia. 1998. V. 4. № 4. P. 543–545.
93. Gouw S.C., van den Berg H.M., Oldenburg J., Astermark J., de Groot P.G., Margaglione M., Thompson A.R., van Heerde W., Boekhorst J., Miller C.H., le Cessie S., van der Bom J.G. // Blood. 2012. V. 119. № 12. P. 2922–2934.
94. Towfighi F., Gharagozlou S., Sharifian R.A., Kazemnejad A., Esmailzadeh K., Managhchi M.R., Shokri F. // Acta Haematol. 2005. V. 114. № 2. P. 84–90.
95. Green D. // Haemophilia. 2011. V. 17. № 6. P. 831–838.
96. Lacroix-Desmazes S., Moreau A., Sooryanarayana, Bonnemain C., Stieltjes N., Pashov A., Sultan Y., Hoebeke J., Kazatchkine M.D., Kaveri S.V. // Nat. Med. 1999. V. 5. № 9. P. 1044–1047.
97. Lacroix-Desmazes S., Bayry J., Misra N., Horn M.P., Villard S., Pashov A., Stieltjes N., d'Oiron R., Saint-Remy J.M., Hoebeke J., et al. // N. Engl. J. Med. 2002. V. 346. № 9. P. 662–667.
98. Lenting P.J., Donath M.J., van Mourik J.A., Mertens K. // J. Biol. Chem. 1994. V. 269. № 10. P. 7150–7155.
99. Duffy E.J., Parker E.T., Mutucumarana V.P., Johnson A.E., Lollar P. // J. Biol. Chem. 1992. V. 267. № 24. P. 17006–17011.
100. Curtis J.E., Helgerson S.L., Parker E.T., Lollar P. // J. Biol. Chem. 1994. V. 269. № 8. P. 6246–6251.
101. Bovenschen N., Boertjes R.C., van Stempvoort G., Voorberg J., Lenting P.J., Meijer A. B., Mertens K. // J. Biol. Chem. 2003. V. 278. № 11. P. 9370–9377.
102. Gilbert G.E., Furie B.C., Furie B. // J. Biol. Chem. 1990. V. 265. № 2. P. 815–822.
103. Kemball-Cook G., Barrowcliffe T.W. // Thromb. Res. 1992. V. 67. № 1. P. 57–71.
104. Cho M.S.,Chan S.Y. Vectors having terminal repeat sequence of Epstein-Barr virus: Patent of USA № 6180108. 2001 (http://www.google.com/patents/US6180108).
105. Saenko E.L., Scandella D., Yakhyaev A.V., Greco N.J. // J. Biol. Chem. 1998. V. 273. № 43. P. 27918–27926.
106. Arai M., Scandella D., Hoyer L.W. // J. Clin. Invest. 1989. V. 83. № 6. P. 1978–1984.
107. Eaton D., Rodriguez H., Vehar G.A. // Biochemistry. 1986. V. 25. № 2. P. 505–512.
108. Lollar P., Knutson G.J., Fass D.N. // Biochemistry. 1985. V. 24. № 27. P. 8056–8064.
109. Fay P.J., Haidaris P.J., Smudzin T.M. // J. Biol. Chem. 1991. V. 266. № 14. P. 8957–8962.
110. Hockin M.F., Jones K.C., Everse S.J., Mann K.G. // J. Biol. Chem. 2002. V. 277. № 21. P. 18322–18333.
111. Saenko E.L., Yakhyaev A.V., Mikhailenko I., Strickland D.K., Sarafanov A.G. // J. Biol. Chem. 1999. V. 274. № 53. P. 37685–37692.
112. Lenting P.J., Neels J.G., van den Berg B.M., Clijsters P.P., Meijerman D.W., Pannekoek H., van Mourik J.A., Mertens K., van Zonneveld A.J. // J. Biol. Chem. 1999. V. 274. № 34. P. 23734–23739.
113. Turecek P.L., Schwarz H.P., Binder B.R. // Blood. 2000. V. 95. № 11. P. 3637–3638.
114. Sarafanov A.G., Ananyeva N.M., Shima M., Saenko E.L. // J. Biol. Chem. 2001. V. 276. № 15. P. 11970–11979.
115. Schwartz R.S., Abildgaard C.F., Aledort L.M., Arkin S., Bloom A.L., Brackmann H.H., Brettler D.B., Fukui H., Hilgartner M.W., Inwood M.J., et al. // N. Engl. J. Med. 1990. V. 323. № 26. P. 1800–1805.
116. Josephson C.D., Abshire T. // Clin. Adv. Hematol. Oncol. 2004. V. 2. № 7. P. 441–446.
117. Gomperts E., Lundblad R., Adamson R. // Transfus. Med. Rev. 1992. V. 6. № 4. P. 247–251.
118. Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C., Kaufman R.J. // Proc. Natl. Acad. Sci. USA. 1986. V. 83. № 16. P. 5939–5942.
119. Lind P., Larsson K., Spira J., Sydow-Backman M., Almstedt A., Gray E., Sandberg H. // Eur. J. Biochem. 1995. V. 232. № 1. P. 19–27.
120. Sandberg H., Almstedt A., Brandt J., Gray E., Holmquist L., Oswaldsson U., Sebring S., Mikaelsson M. // Thromb. Haemost. 2001. V. 85. № 1. P. 93–100.

121. Meulien P., Faure T., Mischler F., Harrer H., Ulrich P., Bouderbala B., Dott K., Sainte Marie M., Mazurier C., Wiesel M. L., et al. // Protein Eng. 1988. V. 2. № 4. P. 301–306.
122. Eaton D.L., Wood W.I., Eaton D., Hass P.E., Hollingshead P., Wion K., Mather J., Lawn R.M., Vehar G.A., Gorman C. // Biochemistry. 1986. V. 25. № 26. P. 8343–8347.
123. Yonemura H., Sugawara K., Nakashima K., Nakahara Y., Hamamoto T., Mimaki I., Yokomizo K., Tajima Y., Masuda K., Imaizumi A., et al. // Protein Eng. 1993. V. 6. № 6. P. 669–674.
124. Kolind M.P., Norby P.L., Flintegaard T.V., Berchtold M.W., Johnsen L.B. // J. Biotechnol. 2010. V. 147. № 3–4. P. 198–204.
125. Thim L., Vandahl B., Karlsson J., Klausen N.K., Pedersen J., Krogh T.N., Kjalke M., Petersen J.M., Johnsen L.B., Bolt G., Norby P.L., Steenstrup T.D. // Haemophilia. 2010. V. 16. № 2. P. 349–359.
126. Sandberg H., Kannicht C., Stenlund P., Dadaian M., Oswaldsson U., Cordula C., Walter O. // Thromb. Res. 2012. V. 130. № 5. P. 808–817.
127. Miao H.Z., Sirachainan N., Palmer L., Kucab P., Cunningham M.A., Kaufman R.J., Pipe S.W. // Blood. 2004. V. 103. № 9. P. 3412–3419.
128. Krishnan S., Kolbe H.V., Lepage P., Faure T., Sauerwald R., de la Salle H., Muller C., Bihoreau N., Paolantonacci P., Roitsch C., et al. // Eur. J. Biochem. 1991. V. 195. № 3. P. 637–644.
129. Esmon P.C., Kuo H.S., Fournel M.A. // Blood. 1990. V. 76. № 8. P. 1593–1600.
130. Herlitschka S.E., Schlokat U., Falkner F.G., Dorner F. // J. Biotechnol. 1998. V. 61. № 3. P. 165–173.
131. Fatouros A., Osterberg T., Mikaelsson M. // Pharm. Res. 1997. V. 14. № 12. P. 1679–1684.
132. Osterberg T., Fatouros A., Mikaelsson M. // Pharm. Res. 1997. V. 14. № 7. P. 892–898.
133. Osterberg T., Fatouros A., Neidhardt E., Warne N., Mikaelsson M. // Semin. Hematol. 2001. V. 38. № 2. Suppl 4. P. 40–43.
134. Fijnvandraat K., Berntorp E., ten Cate J.W., Johnsson H., Peters M., Savidge G., Tengborn L., Spira J., Stahl C. // Thromb. Haemost. 1997. V. 77. № 2. P. 298–302.
135. Eriksson R.K., Fenge C., Lindner-Olsson E., Ljungqvist C., Rosenquist J., Smeds A.L., ostlin A., Charlebois T., Leonard M., Kelley B.D., Ljungqvist A. // Semin. Hematol. 2001. V. 38. № 2. Suppl 4. P. 24–31.
136. Chun B.H., Park S.Y., Chung N., Bang W.G. // Biotechnol. Lett. 2003. V. 25. № 4. P. 315–319.
137. Kelley B.D., Tannatt M., Magnusson R., Hagelberg S., Booth J. // Biotechnol. Bioeng. 2004. V. 87. № 3. P. 400–412.
138. Lusher J.M., Lee C.A., Kessler C.M., Bedrosian C.L. // Haemophilia. 2003. V. 9. № 1. P. 38–49.
139. Smith M.P., Giangrande P., Pollman H., Littlewood R., Kollmer C., Feingold J. // Haemophilia. 2005. V. 11. № 5. P. 444–451.
140. Recht M., Nemes L., Matysiak M., Manco-Johnson M., Lusher J., Smith M., Mannucci P., Hay C., Abshire T., O'Brien A., Hayward B., Udata C., Roth D.A., Arkin S. // Haemophilia. 2009. V. 15. № 4. P. 869–880.
141. Gruppo R.A., Brown D., Wilkes M.M., Navickis R.J. // Haemophilia. 2004. V. 10. № 6. P. 747–750.
142. Gruppo R.A., Brown D., Wilkes M.M., Navickis R.J. // Haemophilia. 2004. V. 10. № 5. P. 449–451.
143. Morfini M., Cinotti S., Bellatreccia A., Paladino E., Gringeri A., Mannucci P.M. // J. Thromb. Haemost. 2003. V. 1. № 11. P. 2283–2289.
144. Johnston A. // Ther. Drug Monit. 2012. V. 34. № 1. P. 110–117.
145. Khrenov A.V., Ananyeva N.M., Saenko E.L. // Blood Coagul. Fibrinolysis. 2006. V. 17. № 5. P. 379–388.
146. Lusher J.M. // Semin. Thromb. Hemost. 2002. V. 28. № 3. P. 273–276.
147. Pollmann H., Externest D., Ganser A., Eifrig B., Kreuz W., Lenk H., Pabinger I., Schramm W., Schwarz T.F., Zimmermann R., et al. // Haemophilia. 2007. V. 13. № 2. P. 131–143.
148. Gouw S.C., van den Berg H.M., le Cessie S., van der Bom J.G. // J. Thromb. Haemost. 2007. V. 5. № 7. P. 1383–1390.
149. Aledort L.M., Navickis R.J., Wilkes M.M. // J. Thromb. Haemost. 2011. V. 9. № 11. P. 2180–2192.
150. Lee C.A., Kessler C.M., Varon D., Martinowitz U., Heim M., Vermylen J. // Haemophilia. 1998. V. 4. № 4. P. 538–542.
151. Martinowitz U., Bjerre J., Brand B., Klamroth R., Misgav M., Morfini M., Santagostino E., Tiede A., Viuff D. // Haemophilia. 2011. V. 17. № 6. P. 854–859.
152. Mei B., Chen Y., Chen J., Pan C.Q., Murphy J.E. // Mol. Biotechnol. 2006. V. 34. № 2. P. 165–178.
153. Lynch C.M., Israel D.I., Kaufman R.J., Miller A.D. // Hum. Gene Ther. 1993. V. 4. № 3. P. 259–272.
154. Koeberl D.D., Halbert C.L., Krumm A., Miller A.D. // Hum. Gene Ther. 1995. V. 6. № 4. P. 469–479.
155. Hoeben R.C., Fallaux F.J., Cramer S.J., van den Wollenberg D.J., van Ormondt H., Briet E., van der Eb A.J. // Blood. 1995. V. 85. № 9. P. 2447–2454.
156. Fallaux F.J., Hoeben R.C., Cramer S.J., van den Wollenberg D.J., Briet E., van Ormondt H., van Der Eb A.J. // Mol. Cell. Biol. 1996. V. 16. № 8. P. 4264–4272.
157. Ward N.J., Buckley S.M., Waddington S.N., Vandendriessche T., Chuah M.K., Nathwani A.C., McIntosh J., Tuddenham E.G., Kinnon C., Thrasher A.J., et al. // Blood. 2011. V. 117. № 3. P. 798–807.
158. Roe S.M., Gormal C., Smith B.E., Baker P., Rice D., Card G., Lindley P. // Acta Crystallogr. D Biol. Crystallogr. 1997. V. 53. Pt 2. P. 227–228.
159. Lindley E.J. // EDTNA ERCA J. 1997. V. 23. № 2. P. 50–51.
160. Ishaque A., Thrift J., Murphy J.E., Konstantinov K. // Biotechnol. Bioeng. 2007. V. 97. № 1. P. 144–155.
161. Li X., Gabriel D.A. // Biochemistry. 1997. V. 36. № 35. P. 10760–10767.
162. Kolind M.P., Norby P.L., Berchtold M.W., Johnsen L.B. // J. Biotechnol. 2011. V. 151. № 4. P. 357–362.
163. Levy J.H., Weisinger A., Ziomek C.A., Echelard Y. // Semin. Thromb. Hemost. 2001. V. 27. № 4. P. 405–416.
164. Gil G.C., Velander W.H., van Cott K.E. // Glycobiology. 2008. V. 18. № 7. P. 526–539.
165. Pipe S.W., Miao H., Butler S.P., Calcaterra J., Velander W.H. // J. Thromb. Haemost. 2011. V. 9. № 11. P. 2235–2242.
166. Chen C.M., Wang C.H., Wu S.C., Lin C.C., Lin S.H., Cheng W.T. // Transgenic Res. 2002. V. 11. № 3. P. 257–268.
167. Chrenek P., Ryban L., Vetr H., Makarevich A.V., Uhrin P., Paleyanda R.K., Binder B.R. // Transgenic Res. 2007. V. 16. № 3. P. 353–361.
168. Niemann H., Halter R., Carnwath J. W., Herrmann D., Lemme E., Paul D. // Transgenic Res. 1999. V. 8. № 3. P. 237–247.
169. Paleyanda R.K., Velander W.H., Lee T.K., Scandella D.H., Gwazdauskas F.C., Knight J.W., Hoyer L.W., Drohan W.N., Lubon H. // Nat. Biotechnol. 1997. V. 15. № 10. P. 971–975.
170. Dove A. // Nat. Biotechnol. 2000. V. 18. № 10. P. 1045–1048.

171. Panno J. Animal Cloning: The Science of Nuclear Transfer, New York:Facts on File, 2004. 176 p.
172. Rostin J., Smeds A.L., Akerblom E. // Bioconjug. Chem. 2000. V. 11. № 3. P. 387–396.
173. Mei B., Pan C., Jiang H., Tjandra H., Strauss J., Chen Y., Liu T., Zhang X., Severs J., Newgren J., Chen J., Gu J.M., Subramanyam B., Fournel M.A., Pierce G.F., Murphy J.E. // Blood. 2010. V. 116. № 2. P. 270–279.
174. Turecek P.L., Bossard M.J., Graninger M., Gritsch H., Hollriegl W., Kaliwoda M., Matthiessen P., Mitterer A., Muchitsch E.M., Purtscher M., et al. // Hamostaseologie. 2012. V. 32. Suppl 1. P. S29–38.
175. Dumont J.A., Liu T., Low S.C., Zhang X., Kamphaus G., Sakorafas P., Fraley C., Drager D., Reidy T., McCue J., et al. // Blood. 2012. V. 119. № 13. P. 3024–3030.
176. Powell J.S., Josephson N.C., Quon D., Ragni M.V., Cheng G., Li E., Jiang H., Li L., Dumont J.A., Goyal J., et al. // Blood. 2012. V. 119. № 13. P. 3031–3037.
177. Pipe S.W. // Blood. 2010. V. 116. № 2. P. 153–154.
178. Pipe S.W., Kaufman R.J. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. № 22. P. 11851–11856.
179. Greene T.K., Wang C., Hirsch J.D., Zhai L., Gewirtz J., Thornton M.A., Miao H. Z., Pipe S.W., Kaufman R.J., Camire R.M., et al. // Blood. 2010. V. 116. № 26. P. 6114–6122.
180. Gale A.J., Pellequer J.L. // J. Thromb. Haemost. 2003. V. 1. № 9. P. 1966–1971.
181. Radtke K.P., Griffin J.H., Riceberg J., Gale A.J. // J. Thromb. Haemost. 2007. V. 5. № 1. P. 102–108.
182. Wakabayashi H., Varfaj F., Deangelis J., Fay P.J. // Blood. 2008. V. 112. № 7. P. 2761–2769.
183. Parker E.T., Healey J.F., Barrow R.T., Craddock H.N., Lollar P. // Blood. 2004. V. 104. № 3. P. 704–710.
184. Barrow R.T., Healey J.F., Gailani D., Scandella D., Lollar P. // Blood. 2000. V. 95. № 2. P. 564–568.
185. Toschi V. // Curr. Opin. Mol. Ther. 2010. V. 12. № 5. P. 617–625.
186. Kempton C.L., Abshire T.C., Deveras R.A., Hoots W.K., Gill J.C., Kessler C.M., Key N.S., Konkle B.A., Kuriakose P., Macfarlane D.E., et al. // Haemophilia. 2012. V. 18. №5. P. 798-804.
187. Milanov P., Ivanciu L., Abriss D., Quade-Lyssy P., Miesbach W., Alesci S., Tonn T., Grez M., Seifried E., Schuttrumpf J. // Blood. 2012. V. 119. № 2. P. 602-611.
188. Kitazawa T., Igawa T., Sampei Z., Muto A., Kojima T., Soeda T., Yoshihashi K., Okuyama-Nishida Y., Saito H., Tsunoda H., et al. // Nat. Med. 2012. V. 18. № 10. P. 1570–1574.

The Evolutionary Pathway of X Chromosome Inactivation in Mammals

A.I. Shevchenko[1, 2, 3], I.S. Zakharova[1, 2, 3], S.M. Zakian[1, 2, 3*]
[1]Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Prospekt Lavrentyeva, 10, Novosibirsk, Russia, 630090
[2]State Research Institute of Circulation Pathology, Rechkunovskaya Str., 15, Novosibirsk, Russia, 630055
[3]Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, Russian Academy of Sciences, Prospekt Lavrentyeva, 8, Novosibirsk, Russia, 630090
*E-mail: zakian@bionet.nsc.ru
Received 02.11.2012

ABSTRACT **X chromosome inactivation is a complex process that occurs in marsupial and eutherian mammals. The process is thought to have arisen during the differentiation of mammalian sex chromosomes to achieve an equal dosage of X chromosome genes in males and females. The differences in the X chromosome inactivation processes in marsupial and eutherian mammals are considered, and the hypotheses on its origin and evolution are discussed in this review.**
KEYWORDS **mammals; X chromosome inactivation; *Xist*.**
ABBREVIATIONS **XIC – X inactivation center; PAR – pseudoautosomal region of the mammalian X chromosome.**

INTRODUCTION

The class Mammalia (mammals) is divided into two subclasses: Prototheria (monotremes) and Theria. In turn, the infraclasses Metatheria (marsupial mammals) and Eutheria (placental mammals) are distinguished in the Theria subclass. The divergence between monotremes and marsupial mammals took place 166.2 million years ago; the divergence between marsupial and placental mammals occurred 147.7 million years ago [1].

The ontogenesis of female marsupial and placental mammals is accompanied by a unique epigenetic phenomenon, heterochromatization of one X chromosome (out of two) and inactivation of its transcription, which is maintained in cell generations [2, 3]. This mechanism is believed to have arisen due to the necessity of gene dosage compensation in heteromorphic sex chromosomes in individuals of the opposite sex. In the subclass Theria, sex is determined by two heteromorphic sex chromosomes, X and Y. Males have the XY combination of sex chromosomes, while females have the XX combination. Since the Y chromosome contains only several tens of genes, as opposed to the X chromosome that contains approximately a thousand genes, most genes in the X chromosome are represented as a single copy in males (XY) and two copies in females (XX). As a result of inactivation of a single X chromosome in females, only one gene copy of the X chromosome is transcriptionally active in individuals of both sexes; thus, approximately equal amounts of the products of X-linked genes are synthesized in cells. X chromosome inactivation occurs due to the effect of specific nuclear RNAs and chromatin modifications that repress transcription and differ in marsupial and eutherian mammals [3, 4]. The evolution of X chromosome inactivation is discussed in this review.

PHENOMENOLOGY OF X-INACTIVATION IN MAMMALS

Monotremes use a mechanism different from X chromosome inactivation for dosage compensation

The living representatives of the most ancient mammalian subclass Prototheria, one platypus and four echidna species, are merged into the order of monotremes (Monotremata). Unlike the rest of mammals, the monotremes have a complex sex-determination system. The male platypus (*Ornithorhychus anatinus*) has five X and five Y chromosomes; five X and four Y chromosomes have been detected in male echidna (*Tachuglossus aculeatus*) [5–7]. The genes typical of the X chromosomes of marsupial and eutherian mammals have autosomal localization [7–9]. However, the genes typical of the sex chromosome Z of birds (including the *Dmrt1* gene, which presumably plays the key role in sex determination in birds) have been found on the X chromosomes of monotremes. The most extensive region homologous to the chicken Z chromosome

Table 1. The ratio between the gene expression levels in the X chromosomes in female and male platypus cells and frequency of their monoallelic expression [10]

Chromosome	Gene	Ratio between the gene expression levels in females and males	Fraction of nuclei with monoallelic expression
Complete compensation			
X_1	*Ox_plat_124086*	1.10	46
X_5	*ZNF474*	1.01	53
X_5	*LOX*	1.06	53
X_3	*APC*	1.17	48
X_5	*SHB*	1.23	53
Partial compensation			
X_5	*FBXO10*	1.37	50
X_5	*EN14997*	1.40	61
No compensation			
X_5	*SEMA6A*	1.82	74
X_5	*DMRT2*	2.04	47
X_5	*SLC1A1*	2.78	45

has been detected on the platypus X5 chromosome; less extensive regions of homology are localized on the X_1, X_2, and X_3 chromosomes (*Fig. 1*).

All the X and Y chromosomes of monotremes contain homologous pseudoautosomal regions that enable conjugation between the X and Y chromosomes in meiosis [5–7]. However, the extensive regions of the platypus X_1–X_5 chromosomes (corresponding to ~12% of the genome) are nonhomologous and show no similarity to Y_1–Y_5. It is reasonable to expect that a mechanism of dosage compensation for the genes localized in these regions exists. A quantitative analysis of the transcription of the genes localized in the differentiated regions of different platypus X chromosomes [10] has demonstrated that some of them have identical transcription levels both in female and male cells, while expression of the remaining genes is either compensated partially or is not compensated at all (i.e., expression in female cells turns out to be twice as high as that in male cells) (*Table 1*). Thus, dosage compensation in monotremes presumably functions only for individual genes of the sex chromosome, resembling incomplete and variable dosage compensation in birds [11, 12]. In cell nuclei of female platypus, transcription of the genes exhibiting dosage compensation is revealed only for one of the homologous X chromosomes with a frequency of 50–70%. Nevertheless, total mRNA contains equal amounts of transcripts corresponding to each homologue. These data provide grounds for assuming that dosage compensation in monotremes occurs due to a decrease in the transcription level of one of the alleles (selected in each cell in a random manner) [10]. Since each pair of X chromosomes in female platypus has no visible distinctions in chromatin modifications at the cytological level, it is assumed that the dosage compensation in monotremes affects individual genes rather than chromosomes [13].

The pseudoautosomal region of the echidna X_1 chromosome in some cell types is characterized by late replication [14], which can be regarded as an indicator of inactive chromatin, although the genes localized in this region are present both on X_1 and Y_1 and require no dosage compensation. Taking into account its susceptibility to inactivation, this region was previously regarded as an ancestral region when the mechanism of silencing of an entire chromosome could have presumably been formed. However, since the genes contained in this region in marsupial and eutherian mammals have autosomal localization and are not involved in inactivation, this assumption has been refuted.

Thus, it is an obvious fact that monotremes, unlike marsupial and eutherian mammals, use a mechanism that differs from X chromosome inactivation for dosage compensation.

Table 2. Status of gene expression in the X chromosomes in different marsupial species

Gene	Species	Method	Inactivation in somatic tissues
G6pd	*Macropus robustus*	Isoenzyme analysis, SNuPE	Complete
	Macropus rufogriseus	Isoenzyme analysis	Complete
	Didelphis virginiana	Isoenzyme analysis	Partial
	Monodelphis domestica	RT-PCR	Complete
Gla	*Antechinus stuarttii*	Isoenzyme analysis	Complete
	Kangaroo hybrids	«	Complete
Pgk1	*Macropus giganteus*	«	Tissue-specific
	Macropus parryi	«	«
	Trichosurus vulpecula	«	«
	Didelphis virginiana	«	«
	Monodelphis domestica	SNuPE	Partial

Imprinted, incomplete and tissue-specific X chromosome inactivation in marsupial mammals

Infraclass Metatheria (marsupials) comprises 270 species, 200 of which live in Australia; 69, in South America; and 1, in North America. The evolutionary segregation between Australian and American marsupials occurred 70 million years ago [9, 15]. The sex chromosomes in marsupial and eutherian mammals have a common origin. The X chromosome in marsupials represents 2/3 of the X chromosome of eutherian mammals; the remaining third of the genes are localized on the autosome (*Fig. 1*). Marsupials are the most ancient mammals; dosage compensation in female marsupials occurs due to X chromosome inactivation; however, the inactivation processes in marsupial and eutherian mammals differ significantly.

Nonrandom imprinted inactivation is typical of all marsupial tissues; it involves suppression of gene transcription and establishment of late replication in the S phase of the cell cycle, exclusively on the X chromosome inherited from the father [16, 17]. The untranslated nuclear RNA *Rsx* (RNA-on-the-silent X), which can propagate over the inactive X chromosome and repress gene transcription, is presumably responsible for the inactivation process at the chromosomal level [4]. The imprinted inactivation of three genes of the X chromosome has been studied in tissues of eight species (*Table 2*). It was found that the inactive status of the X chromosome inherited from the father is unstable, and that genes are frequently reactivated. It turns out that inactivation in marsupials does not affect all genes to the same extent (i.e., is incomplete). Moreover, the same loci of the X chromosome can be inactivated to different extents depending on a particular tissue. Thus, the phosphoglycerate kinase A (*Pgk1*) gene in the Virginia (North American) opossum *Didelphis virginiana* is completely inactivated in all tissues, whereas no stable repression of the paternal allele of the glucoso-6-phosphate dehydrogenase (*G6pd*) gene is observed in most tissues [18]. In the gray short-tailed opossum *Monodelphis domestica*, unlike the Virginia opossum, the paternal *G6pd* allele is stably inactivated, whereas *Pgk1* exhibits incomplete inactivation in all tissues [19]. Thus, orthological genes can be inactivated to different extents in different marsupial species.

It should be mentioned that X chromosome inactivation is not the only mechanism of dosage compensation in marsupials. In the members of the bandicoot family (Paramelidae), the Y chromosome in males and one of the two X chromosomes in females are eliminated at different ontogenic stages in somatic cells [20]. The elimination of sex chromosomes in different tissues can be observed either in all cells or in some of them. The investigation of the expression of the alleles of the X-linked *Pgk1* gene in the southern brown bandicoot *Isoodon obesulus* shows that only the X chromosome inherited from the father is eliminated in females [21]. In the cells where sex chromosomes have not been eliminated, the X chromosome of paternal origin in females and the Y chromosome in males are late-replicating. The mechanism of elimination of sex chromosome is unknown; however, the preferential elimina-

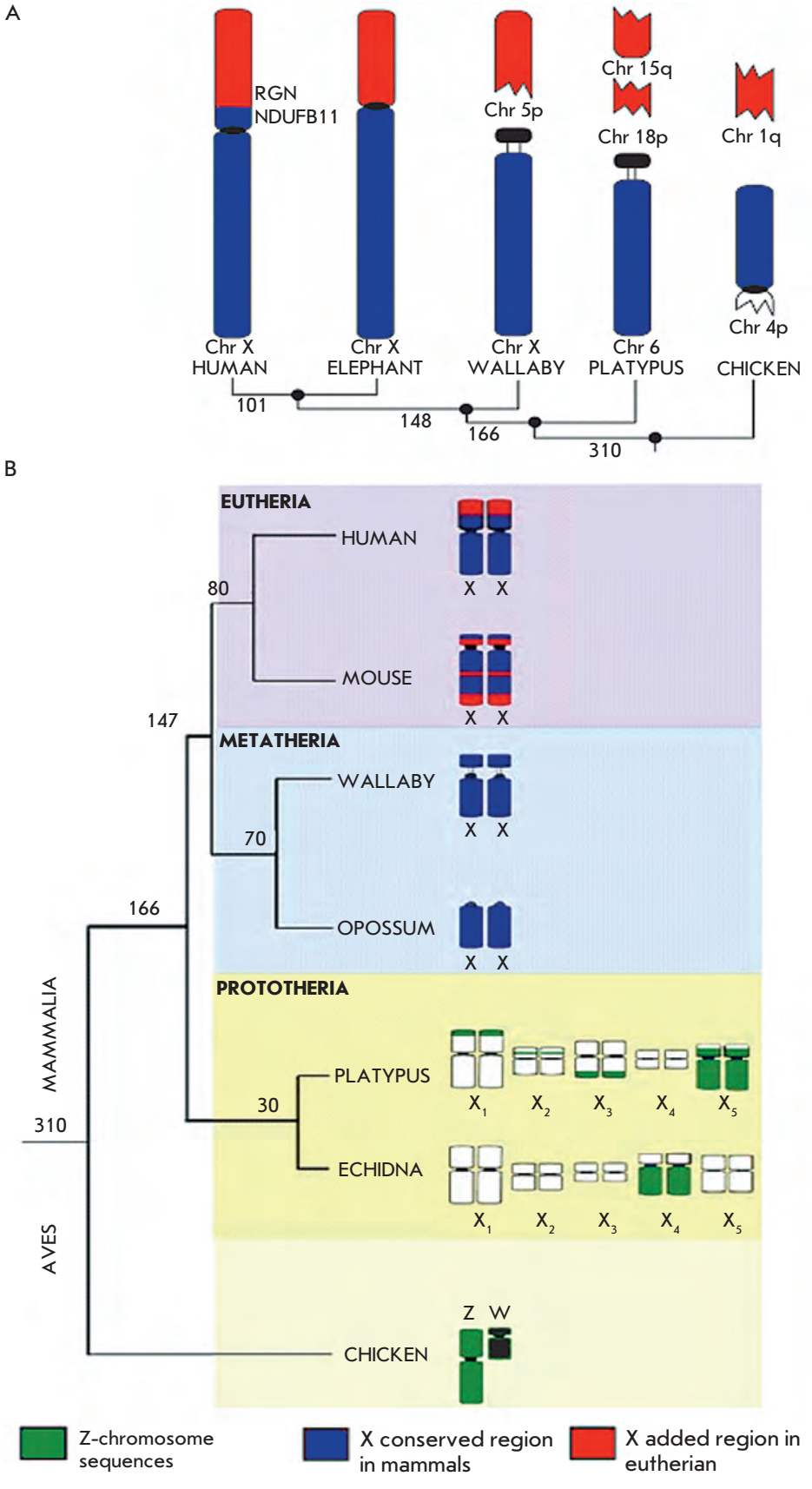


Fig. 1. The origin and evolution of the mammalian X chromosome. A) Genes of the mammalian X chromosome have autosomal localization in birds (chicken) and monotremes (platypus, echidna). The X chromosome of marsupials (wallaby, opossum) represents the most ancient part of the mammalian X (shown in blue) and comprises 2/3 of the genes of the eutherian X chromosome. The eutherian X chromosome contains an added region (shown in red), which has autosomal localization in marsupials [7]. B) Monotremes have five X chromosomes, which show nothing in common with eutherian X but contain sequences homologous to the Z chromosome of birds [9]. The divergence time of the taxa (Mya) is shown on the branches of the phylogenetic tree

tion of the X chromosome inherited from the father and asynchronous replication of the X chromosomes in females attest to the fact that this process emerged in marsupials as a trend in the evolution of the X chromosome inactivation process.

Eutherian mammals have imprinted and random X chromosome inactivation, which are controlled by the inactivation center and the *Xist* gene

Infraclass Eutheria (placental mammals), which is subdivided into the four superorders Afrotheria, Xenarthra, Euarchontoglires and Laurasiatheria, is the most numerous, diverse, and common mammalian infraclass. The X chromosome in eutherian mammals consists of the genes constituting the X chromosome in marsupials by 2/3 and contains an added region, which has autosomal localization in marsupials [9] (Fig. 1). As opposed to marsupial mammals, the X chromosomes of paternal and maternal origins are inactivated with equal probabilities in the cells of adult female eutherians; hence, on average half of the cells express the genes of the paternal X chromosome, while the remaining half express the genes of the maternal X chromosome. Unlike imprinted inactivation, random inactivation embraces most genes on the X chromosome and is stably maintained through cell generations. It should be mentioned that the genes in the added region of the X chromosome in eutherian mammals, which were localized on the autosome in marsupials and did not participate in the inactivation process, are inactivated with a lower efficiency and are capable of avoiding inactivation [22]. The random inactivation in eutherians comprises several stages: counting the number of X chromosomes per diploid genome, choice of an X chromosome for inactivation, initiation of activation, and propagation of the inactive status and its maintenance through cell generations [3, 23]. It is possible that the stage involving the choice of the X chromosomes (during which the mutually exclusive choices of the future active and inactive X chromosomes (like in a mouse) occurs) is typical not of all eutherian species. Thus, inactivation in early ontogenesis of the rabbit occurs stochastically, resulting in the formation of different cells, where 1) neither one of the X chromosomes is inactivated, 2) both X chromosomes are inactivated, or 3) one X chromosome out of two is randomly inactivated. Due to the disrupted gene dosage, the former two cell types subsequently die, while the remaining cells with normal inactivation form the organs and tissues of the organism [24].

In certain taxa of eutherian mammals (e.g., in rodents and artiodactyles), in addition to the random inactivation there also exists imprinted, incomplete and unstable inactivation of the X chromosome inherited from the father (however, this occurs exclusively at the pre-implantation stages of embryogenesis and remains in cells resulting in extraembryonic organs (placenta and vitelline sac) [25, 26].

Both the random and imprinted inactivation in eutherians are controlled by the inactivation center (XIC) and the *Xist* gene, which have not been detected in monotremes and marsupials [3, 23]. During the random inactivation, the *Xist* gene ensures initiation of inactivation and propagation of the inactive status, while the other elements of the inactivation center function at the stage of the counting of X chromosomes and choice of the chromosome to undergo inactivation.

The evolution of complete and stable inactivation was accompanied by substitution of the noncoding RNA *Rsx* by *Xist* and the emergence of *Xist*-dependent modifications in the histones on the inactive X chromosome, along with DNA methylation in promoters

Despite the differences, there are a number of common features between the X chromosome inactivation in marsupial and eutherian mammals, which presumably reflect the fundamental and the most ancient mechanisms underlying this process (*Fig. 2*) [13, 27, 28]. Both in marsupials and eutherian mammals, the inactive X chromosome is revealed in female interphase nuclei in the form of a cytologically discernible compact chromatin mass known as the Barr body. The DNA-dependent RNA polymerase II responsible for gene transcription is almost completely eliminated from the chromosomal area of the inactive X chromosome in interphase nuclei. The inactive X chromosome is late-replicating; during the replication stage, it migrates to the perinucleolar region of the nucleus, which is enriched in the enzymes required to reproduce the inactive chromatin structure. Covalent histone modifications typical of transcriptionally active chromatin are eliminated in the inactive X chromosome, while modifications typical of transcriptionally inactive chromatin are present. Chromatin of the inactive X chromosome contains untranslatable nuclear RNA, which is expressed only from the inactive X chromosome and propagates over it, resulting in gene inactivation.

It should be emphasized that marsupial and eutherian mammals use completely different, unrelated in terms of their origin and nuclear noncoding RNAs of *Rsx* and *Xist*, which exhibit similar properties and behavior during the inactivation process [4]. Both noncoding RNAs are enriched in microsatellite repeats, which are significant functional domains required for the repression of transcription, propagation over the inactive X chromosome, and binding of the protein complexes responsible for chromatin modification (as has been demonstrated for *Xist* RNA) [29] (*Fig. 3*). The

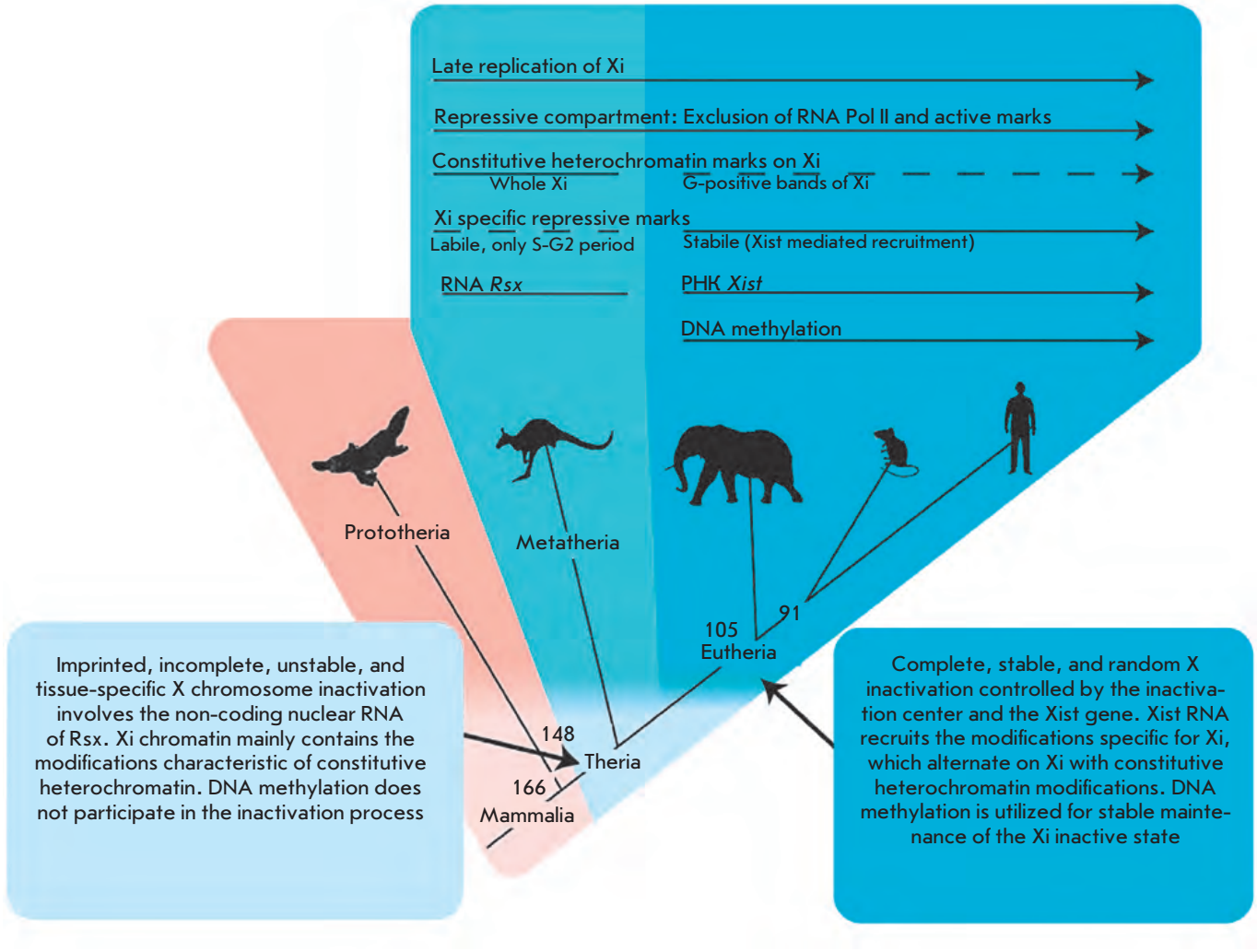


Fig. 2. The evolution of the epigenetic mechanisms underlying X chromosome inactivation in mammals [28]. Xi is the inactive X chromosome. The divergence time of the taxa (Mya) is shown on the branches of the phylogenetic tree

evolutionary conserved minisatellite A-repeats localized in the first exon of the *Xist* gene play a significant role in the inactivation of the transcription of X chromosome genes [30]. Deletion of the A-repeats renders *Xist* RNA incapable of inducing inactivation of the transcription of X-linked genes, although it can still normally propagate along the X chromosome [29, 31]. The propagation of *Xist* RNA along the X chromosome is controlled by the cumulative action of the microsatellite repeats B, C, D, and E [32]. The area of minisatellite C-repeats is responsible for the binding of Xist RNA to the chromatin of the inactive X chromosome via the hnRNP U protein, which is also known as SP120 and SAF-A (scaffold attachment factor A) [31, 33–35]. hnRNP U (heterogeneous nuclear ribonucleoprotein U) is a protein that contains three conserved domains:

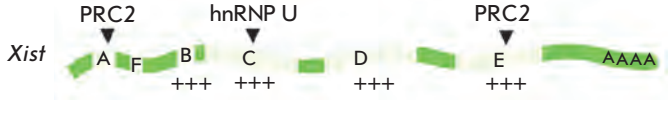


Fig. 3. Functional RNA domains of the *Xist* gene. A, B, C, D, E, F – minisatellite repeats included in *Xist* RNA. **(+++)** – sequences responsible for *Xist* RNA spreading on the X chromosome. Arrows indicate the A- and E-repeat regions involved in binding of the PRC2 protein complex and the C-repeat region responsible for *Xist* RNA binding to the inactive X chromosome by the hnRNP U (SP120/SAF-A) protein. A-repeats are also necessary for transcriptional gene silencing and organization of the inactive X chromosome compartment [3]

SAF-Box, which can bind to the AT-rich DNA region known as S/MAR (scaffold- or matrix-attachment region); the SPRY domain (Spla and Ryanodine receptor) with an unknown function; and the RNA-binding domain RGG (arginine-glycine-glycine). The presence of these domains makes it possible for hnRNP U to interact with *Xist* DNA and RNA, which facilitates its retention in the inactivated X chromosome [35].

It should also be noted that during the whole cell cycle the inactive X chromosome in marsupials is stably associated with heterochromatin protein HP1, histone H3 trimethylated at lysine K9, and histone H4 trimethylated at lysine K20, which are typical of the centromeric and telomeric regions of constitutive heterochromatin [13, 28, 36]. Some modifications specific to the inactive X chromosome in eutherians (e.g., histone H3 trimethylated at lysine 27) may temporarily emerge on the inactive X chromosome of marsupial mammals during the period between the S- to and the early G2-phase of the cell cycle.

In eutherians (similarly to marsupials), the repression of the entire X chromosome at the early stages of imprinted inactivation may occur exclusively as a result of the modifications typical of constitutive heterochromatic regions [37]. At the later stages of imprinted inactivation, as well as in the case of random inactivation, these modifications occur on the inactive X chromosome only in the regions enriched in repeats, which correspond to the G-positive bands. The regions of the inactive X chromosome enriched in genes are stably repressed during the whole cell cycle via the trimethylation of H3 at lysine K27, monoubiquitination of H2A at lysine K119, and insertion of the histone macroH2A1.2 (which are colocalized with *Xist* RNA) into chromatin [38–42]. The emergence of modifications capable of colocalizing with the *Xist* gene depends on its expression; repression of *Xist* and disturbances in the propagation of its RNA result in elimination of these modifications from the inactive X chromosome [29, 31, 43]. Moreover, it has been revealed that *Xist* RNA contains two sites that are capable of binding to the protein complex PRC2 (Polycomb repressive complex 2), whose proteins function as histone methyltransferases responsible for the trimethylation of H3K27 [44].

Methylation of DNA in the inactive X chromosome is another epigenetic difference during inactivation in marsupial and eutherian mammals. The DNA of the inactive X chromosome in the embryonic tissues of eutherian mammals (as opposed to that of the active chromosome) is hypermethylated at the CpG dinucleotides localized in the promoters and 5'-untranslated regions of the genes during random inactivation [45]. The methylation is detectable during unstable imprinted inactivation neither in the extraembryonic tissues of eutherians nor in the somatic tissues of marsupials [18, 19, 46]. Methylation of promoter DNA during random inactivation has presumably emerged in eutherians as an additional stage of stabilization of the inactive status of the X chromosome in somatic cells.

HYPOTHESES CONCERNING THE ORIGIN AND EVOLUTION OF X CHROMOSOME INACTIVATION

Imprinted inactivation is likely to be more ancient

Imprinted X chromosome inactivation, which occurs in all marsupial tissues and organs and in extraembyonic organs (placenta, vitelline sac) in a number of eutherian mammals, is considered to be the most ancient and primitive X chromosome inactivation. Imprinted inactivation has further evolved into the more preferable process of random inactivation as it incorporates the mechanisms of counting the number of X chromosomes per diploid set and choosing the future inactive X chromosome, which are controlled by the inactivation center.

Imprinted inactivation in certain eutherian taxa could have been retained or emerged again as it incorporated the new mechanisms offered by the inactivation center and the *Xist* gene. Thus, at least in mice, imprinted inactivation involves XIC and *Xist*. Imprinting preventing *Xist* expression and protecting the X chromosome inherited from the mother against inactivation has been detected in XIC [23]. However, imprinted inactivation has been completely eliminated in the other taxa (e.g., in humans) [47].

The inactivation process could have originated from the mechanisms of imprinted or random monoallelic expression of autosomal genes and from meiotic silencing of sex chromosomes

There is at present no satisfactory explanation for the origin of the X chromosome inactivation. The inactivation mechanism could have emerged *de novo* on the X chromosome or could have been borrowed from the existing silencing process.

There is a hypothesis that the mechanism that is used for imprinted monoallelic expression of the genes on one of the two homologous autosomes could underlie imprinted X chromosome inactivation [48]. Imprinting of gene expression on autosomes is a common conserved process among marsupial and eutherian mammals. Nuclear RNAs, whose expression causes transcriptional gene silencing *in cis*, elimination of the modifications typical of active chromatin, and recruitment of the modifications specific to inactive chromatin, are involved both in autosomal genomic imprinting and in X chromosome inactivation in eutherian mammals. In eutherians, both these processes occur at the early stages

of embryonic development, are retained in placenta, and lost in the embryo.

It should be mentioned that the randomly established monoallelic expression of autosomal genes is also a rather common phenomenon. Thus, the genes of immunoglobulins, factory receptors, T-cell receptors, and natural killer cell receptors exemplify the genes with monoallelic expression, which is determined stochastically. A number of genes with random monoallelic expression are characterized by asynchronous replication: they are early-replicating on one homologue and late-replicating on the other one during the S-phase of the cell cycle. The asynchronous replication of these genes is likely to take root during early development. Clusters of different genes with monoallelic expression localized on the same chromosome at a considerable distance from one another are characterized by equal replication times within the same homologue [49]. This fact provides grounds to assume that each homologue within a pair has its own specifically organized chromosomal area, which is similar to the region of the inactive X chromosome that can be cytologically detected in the interphase nuclei of marsupial and eutherian mammals as a compact chromatin mass known as the Barr body [23]. Thus, it is possible that X chromosome inactivation originates from the mechanism of stochastic monoallelic gene expression, with imprinting introduced later [50].

It has also been assumed that imprinted inactivation of the X chromosome inherited from the father either originates from meiotic inactivation of sex chromosomes in spermatogenesis or is its extension [18]. During spermatogenesis, meiotic inactivation of sex chromosomes at the pachytene stage of meiosis results in transcriptional silencing of sex chromosomes, giving rise to the sex body (XY body). The assumption of the fact that imprinted inactivation of the X chromosome may be related to the process of meiotic inactivation of sex chromosomes in spermatogenesis is supported by the data indicating that chromatin modifications identical to those formed during meiotic inactivation are formed during imprinted inactivation in marsupial mammals and at the early stages of imprinted inactivation in eutherians [37]. The tentative cognation between meiotic and imprinted inactivation provides grounds to believe that X chromosome inactivation could have occurred at the early evolutionary stages without the participation of nuclear noncoding RNA (and if this RNA did exist, it did not play the key role in transcriptional repression). This assumption is based on the data indicating that similar modifications ensuring chromatin repression are not specific to the inactive X chromosome in case of meiotic and imprinting inactivation but are typical of all the regions of constitutive heterochromatin in the genome, and that their emergence (at least on the eutherian X chromosome) is independent of *Xist* expression. Moreover, meiotic inactivation and the early stages of imprinted inactivation in eutherian mammals can successfully occur in the absence of *Xist* RNA, as well [51, 52]. In marsupials, meiotic gene repression in spermatogenesis is also independent of the *Rsx* gene, which is not expressed at this stage [4]. Thus, it can be assumed that the role of nuclear RNA in X chromosome inactivation could have originally consisted in organization of the specific chromosomal area or in relocation of the inactive chromosome to the perinucleolar compartment in order to ensure its replication (these processes occur with the immediate participation of *Xist* RNA) [53–55]. It was not until some time later that nuclear RNAs started to be used directly for transcriptional repression and recruitment of the protein complexes repressing chromatin. However, one should bear in mind that the core histones (along with the epigenetic data regarding the transcriptional status of chromatin) are in most cases replaced by protamines as chromosomes are packaged in sperm cells, while methylation of the CpG islands employed for the inheritance of the inactive status in X-linked genes has not been detected [19]. Hence, it remains unclear how the inactive status of chromatin can be transmitted to the zygote. Furthermore, since the molecular mechanisms of both meiotic and imprinted inactivation remain unknown, it is difficult to determine the actual cognation between these processes.

ORIGIN AND EVOLUTION OF THE X INACTIVATION CENTER AND THE *XIST* GENE

The genes of the X inactivation center originate from the protein-coding genes and mobile elements

The X inactivation process in eutherian mammals is controlled by a complex genetic locus of the X chromosome, the X inactivation center (XIC). Along with *Xist*, the XIC of evolutionarily distant eutherian species contain two more genes that encode nuclear RNAs – *Enox* (*Jpx*) and *Ftx*; it has been shown in experiments on mice that these genes activate *Xist* expression [56–59]. The XIC also contains the protein-coding genes *Tsx* and *Cnbp2*, whose products are not involved in inactivation [56]. It has been demonstrated that several protein-coding genes in the region of synteny with the XIC on chicken chromosome 4 exhibit homology with the genes of the inactivation center and could be their ancestors [60]. The Lnx3 gene, whose protein product contains the ubiquitin-ligase domain PDZ, underlies the formation of *Xist* (*Fig. 4*). It has been shown by comparing these genes that the promoter region and at least three exons of the *Xist* gene originate from the

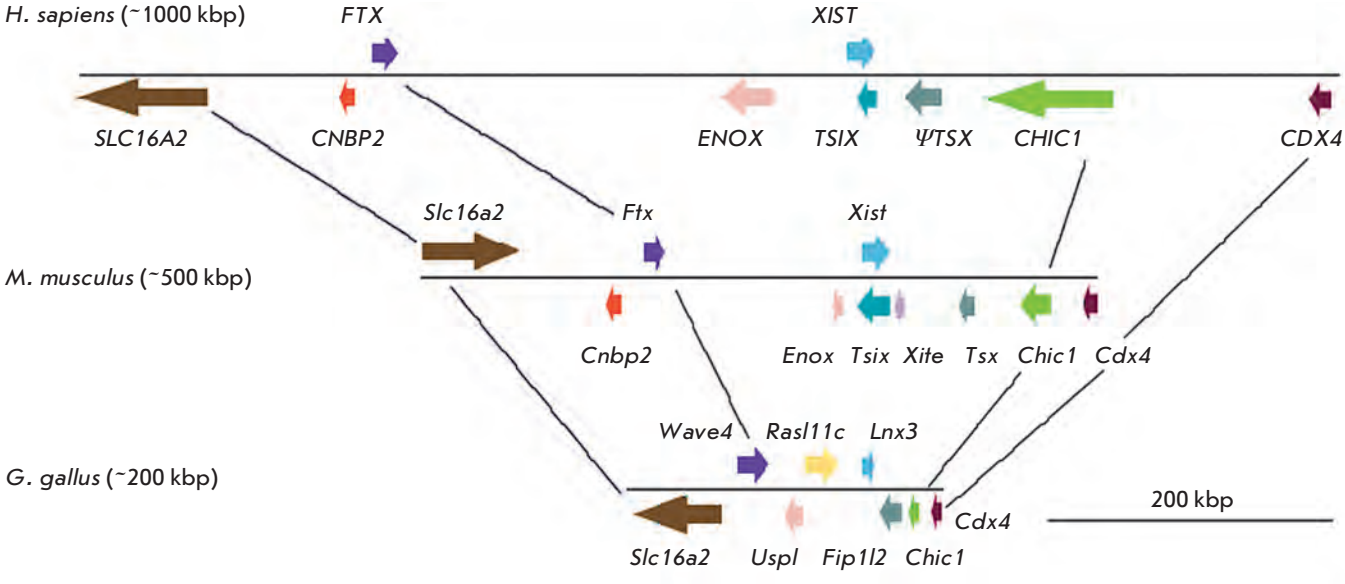


Fig. 4. Comparison of the human and mouse X inactivation centres with its homologous region in chicken. Colored boxes represent genes; arrows show their transcription direction. Homologous genes in different species are shown in the same color. Lines connect the same homologous genes in the cognate loci of chicken, mouse and human. *Cdx4*, *Chic1* and *Slc16a2* are the conserved protein-coding genes that flank both eutherian XIC and its homologous locus in chicken. *Cnbp2* is a protein-coding gene, which was retrotransposed to the XIC locus in the eutherian lineage. *Tsx* is a testis-specific protein-coding gene which partially evolved from the cognate chicken protein-coding gene *Fip1l2*. Note that human *TSX* is no longer functional and represents a pseudogene. *Xist*, *Enox* (*Jpx*) and *Ftx* are the genes of XIC-produced nuclear RNA; they show homology to the cognate chicken protein-coding genes *Lnx3*, *Uspl* and *Wave4*, respectively. The remainder of the chicken protein-coding gene *Rasl11c* is found in eutherian XIC between the genes *Rtx* and *Enox* (*Jpx*)

sequences of the *Lnx3* gene. The largest first exon of the *Xist* gene presumably descended from endogenous retroviruses, whose fragments (after having been inserted into the locus) were amplified, producing simple tandem repeats of several types, which have been identified within it. The remaining exons of the *Xist* gene are syntenic to mobile elements of various classes (*Fig. 5*) [61]. The protein-coding genes surrounding the *Lnx3* gene produced the other genetic elements of the inactivation center in mammals (*Fig. 4*). The *Tsx* gene descended from the *Fip1l2* gene. The two other genes, *Uspl* and *Wave4*, gave rise to *Enox* (*Jpx*) and *Ftx*, respectively. It can be noted that the *Enox* (*Jpx*) gene (as well as *Xist*) contains exons descending from mobile elements, which correspond to various types of repeats in different species [56, 57, 61].

Monotremes and marsupials have no *Xist* gene; the region homologous to the X inactivation center in eutherians is separated by chromosomal rearrangements

No direct orthologues of the Xist gene or other XIC sequences have been detected in monotremes and eutherians as a result of screening of the genome libraries and of a thorough search for homology in the sequenced genomes [62]. Moreover, protein-coding genes ancestral to XIC separated by independent chromosomal partitions and localizing as two individual groups (on the X chromosome in marsupials and on chromosome 6 in monotremes) have been detected in them [60, 62–64]. *Lnx3* RNA in marsupials has a native reading frame, is expressed both in males and females, and obviously functions as a protein-coding gene rather than as an untranslated nuclear RNA that is similar to *Xist*. Thus, protein-coding genes ancestral to XIC were transformed into the genes of the inactivation center only in eutherian mammals; the inactivation process in marsupials involves neither *Xist* nor XIC. The *Rsx* gene in marsupial mammals, which presumably has functions similar to those of the *Xist* gene in eutherians, flanks the protein-coding gene *Hprt* of the X chromosome and does not share a common origin with *Xist* and XIC [4].

The *Xist* gene and the X inactivation center rapidly accumulate species-specific differences during evolution

The *Xist* gene has been detected in the genomes of representatives of all four mammalian superorders,

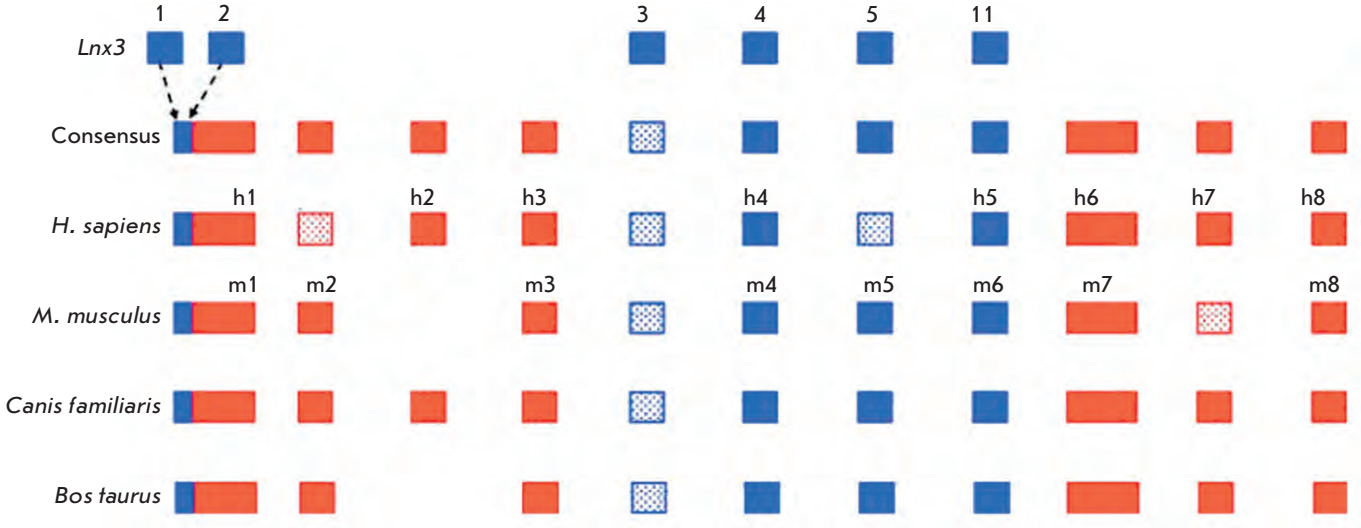


Fig. 5. The origin of the *Xist* gene from the sequences of the protein-coding gene *lnx3* and various classes of mobile elements [61]. Blue rectangles denote the exons that evolved from the gene L*nx3;* red rectangles denote the exons originating from mobile elements; hatched blue and red rectangles denote the exon sequences detectable in the genome but not contained in the *Xist* transcript in the corresponding species. Consensus is a putative ancestral structure of the *Xist* gene. Exon numbering is given for the human (*Homo sapiens*) and mouse (*Mus musculus*) *Xist* genes: m1–m8 for mouse and h1–h8 for human

including the most ancient Afrotheria and Xenarthra [62]. However, the *Xist* gene is not conserved and evolves very rapidly [56, 60, 61, 65]. The exons of the *Xist* gene evolve slower than introns do. The most conserved, exon 4, bears the best resemblance with the exon of the *Lnx3* gene. Paradoxically, the first exon with some functions (and, in particular, the A-repeat region required to establish transcriptional gene silencing) evolves quicker than exon 4, whose deletion has no effect on inactivation. The number of exons per gene in different eutherian species varies from six to eight (*Fig. 6*). The sequences that are exons in certain species may constitute introns in other species. The size of certain exons may vary due to the formation of new exon–intron borders. The size of the largest first exon of the *Xist* gene varies due to amplification and deletions of the tandem repeats within it and insertions/deletions of taxon-specific mobile elements. Because of this variability, the length of *Xist* RNA in representatives of different orders may differ approximately twofold. The differences in the *Xist* gene in terms of RNA size, presence of exons, repeats, and mobile elements are believed to be attributable to its adaptation to functioning in the genome and to the X chromosome in each particular species.

The mouse XIC has two additional genes that encode nuclear RNA: *Tsix,* which is expressed from the antisense chain of the *Xist* gene, and *Xite* (X-inactivation intergenic transcriptional element). These genes control *Xist* expression during imprinted and random inactivation; they are involved in the counting of the number of X chromosomes per diploid autosomal set and choice of the future inactive X chromosome [66]. These genes are less conserved. Not all rodents possess the *Xite* gene; it has not been detected in humans [67]. Antisense transcription with respect to the *Xist* gene (similar to that for *Tsix*) has been detected in humans; however, it does not exhibit the same functions it does in mice [68, 69]. Thus, no conserved elements of XIC responsible for the functions of "counting" and "choice" have been found; hence, the functional elements of XIC, *Xist* regulation, and the inactivation process are at least partially species-specific [67].

In general, it can be noted that the genes of nuclear RNAs involved in the inactivation process in XIC of eutherian mammals evolve very quickly. Their exon–intron structure and borders are changed; some noncoding RNAs participating in the inactivation process are lost, while some others emerge during the evolution. Against this background, the replacement of the *Rsx* gene in marsupials by the *Xist* gene in eutherian mammals seems to be a trivial phenomenon, which properly complies with the general evolutionary trends of the inactivation process.

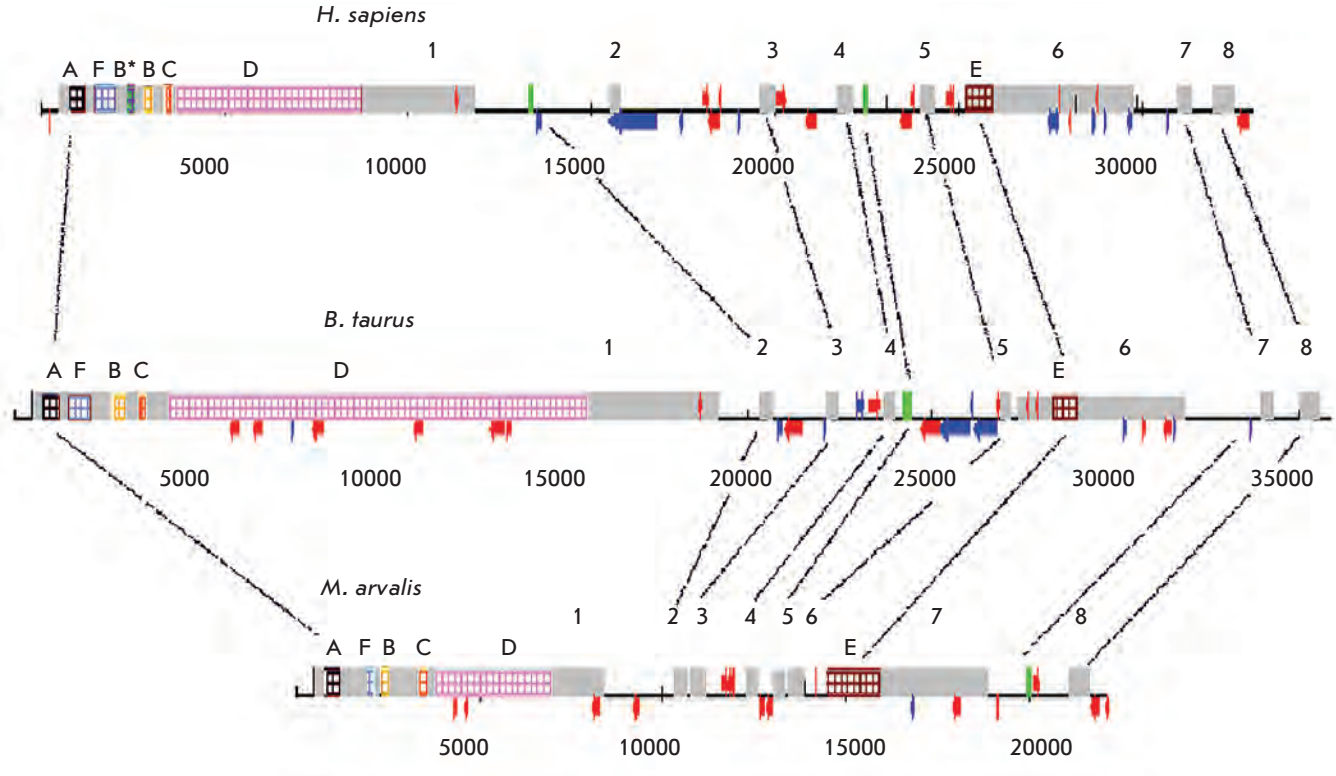


Fig. 6. Comparison of the *Xist* gene structures in vole *M. arvalis*, *B. taurus* and *H. sapiens*. Grey rectangles represent exons (1–8). Green rectangles indicate parts of introns, which are exons in the *Xist* of other species. Lines connect the homologous sequences. Colored rectangles indicate arrays of tandem repeats, named A, B, C, D, E and F, which are present in the *Xist* exons of allthree species, and B*-repeats specific for humans. Yang species-specific LINE and SINE (short interspersed nuclear elements) mobile elements are indicated by blue and red arrows, respectively

COEVOLUTION OF THE X CHROMOSOME AND THE X INACTIVATION PROCESS

The X inactivation process limits the exchange of genetic material between the X chromosome and autosomes

The evolution of mammalian sex chromosomes and X chromosome inactivation occur in an interrelated manner. The necessity of dosage compensation of X-linked genes emerged in mammals during the differentiation of sex chromosomes that had originally been a homologous autosomal pair. The process of X chromosome inactivation emerged after the Y chromosome started to lose homologues of the X chromosome genes and to accumulate the genes that participate in male gametogenesis as a result of recombination repression between the proto-X and proto-Y chromosomes [70]. The homology on the X and Y chromosomes was retained within a short region referred to as the pseudoautosomal region (PAR), which is required to ensure correct conjugation between the X and Y chromosomes during male meiosis. The PAR genes, which are homologous on the X and Y chromosomes, require no dosage compensation and avoid inactivation. The inactivation process presumably emerged in the common ancestor of marsupial and eutherian mammals on the X chromosome, which was compositionally close to the marsupial X chromosome. Further translocations of autosomal material to the ancestral X chromosome, which are observed in eutherian mammals, are supposed to have occurred in such a manner as not to disturb the dosage compensation. Otherwise, these rearrangements would have been eliminated by selection. It has been assumed that dosage compensation had not been disturbed when autosomal material had been added to the PAR of the X chromosome, followed by translocation to the PAR in the Y chromosome via recombination. At the initial stages, the autosomal genes newly added to the PAR in the X and Y chromosomes required no dosage compensation. Then, along with gene degradation in PAR on the Y chromosome, their homologues on the X chromosome became involved in the inactivation process. There were five sequential translocations on the mammalian X chromosome, resulting in the addition

of autosomal genes to the ancestral X chromosome and the formation of younger evolutionary strata. In modern mammals, the lowest number of active homologues of the Y chromosomes has been retained in the most ancient (conserved) part of the X chromosome (*Fig. 1*), while the added regions contain more genes that avoid inactivation and have an active homologue on the Y chromosome [71]. Nevertheless, the eutherian X chromosome contains genes that avoid inactivation despite the fact that their homologue on the Y chromosome has been eliminated. Thus, the involvement of the genes in the inactivation process presumably takes some time; it appears that it does not take place immediately after the Y homologues are eliminated. Moreover, it has been noted that a twofold decrease in the amount of the product of one gene may have no adverse effects on a cell and the organism; hence, there is no need for gene dosage adjustment [72]. The dosage gene compensation in sex chromosomes seems to be aimed at maintaining the collective functions of the genes (e.g., the total protein concentration per cell), which depends on a number of expressible genes. Significant changes in the concentration of cytoplasmic proteins may disturb the ion concentration gradient on the cell membrane. Excess of protein products of X chromosome genes due to disturbance of inactivation results in the development of autoimmune diseases. Thus, disturbance of the collective gene functions may act as the driving force behind the evolution of dosage compensation.

An interesting solution to the problem of translocation of autosomal material to the X chromosome has been revealed in the common shrew *Sorex araneus*. Common shrews have not experienced the recombinational transfer of the translocated autosomal fragment to the Y chromosome; hence, the X chromosome in modern common shrews has two homologues: one corresponding to the ancestral Y chromosome (Y_1), while the other one corresponds to the translocated autosome (Y_2) [73]. The major part of the short arm of the X chromosome (original X) behaves as a typical eutherian X chromosome: it conjugates to the true Y chromosome during male meiosis and undergoes inactivation in female somatic cells. The added region, which occupies the long arm and the small pericentromeric region of the short arm, is identical to the autosome in terms of its behavior: it conjugates to Y_2 and does not undergo inactivation.

X chromosomes in eutherian mammals are enriched in LINE1 retrotransposons that participate in propagation and/or maintenance of the inactive status

The autosomal genes linked to the inactivation center were found to be inactivated less efficiently as compared to X chromosome genes. An assumption has been put forward that the X chromosome has presumably accumulated specific sequences participating in propagation and/or maintenance of the inactive status. M.F. Lyon [74] has mentioned that this role can be played by LINE1 retrotransposon, whose density on the mouse X chromosome is higher than that on autosomes. This hypothesis has been further supported by data obtained by an analysis of sequenced mammalian genomes. The LINE1 content on the X chromosomes in mice, rats, and humans is twice as high as that on autosomes. LINE1 are distributed rather uniformly along the eutherian X chromosome; their fraction is reduced only in the regions containing the genes that avoid inactivation [75, 76]. In the gray short-tailed opossum *M. domestica*, the fractions of LINE1 localized in the X chromosome and autosomes do not significantly differ. This fact agrees with the data on incomplete and instable inactivation in marsupial mammals and demonstrates that an increased LINE1 content is associated with their role in the inactivation process rather than being caused by the less efficient negative selection of LINE1, due to the decrease in the frequency of meiotic recombinations of the X chromosome as compared to autosomes. The experimental data demonstrate that LINE1 can participate in the arrangement of the chromosomal area of the inactive X chromosome; evolutionarily, the youngest LINE1 are expressed on the inactivated X chromosome and promote propagation of the inactive status [77].

CONCLUSIONS

Thus, it can be said that the process of X chromosome inactivation in marsupial and eutherian mammals has common epigenetic and, possibly, molecular mechanisms (*Fig. 2*). The key feature of the inactivation process in mammals, the coordinated gene repression at the level of the X chromosome, is presumably a result of the propagation of the nuclear noncoding RNA along it. However, the *Rsx* gene of nuclear noncoding RNA was replaced in eutherians during evolution by *Xist*, which is better, as compared to its ancestor, at attracting modifications, providing stable gene inactivation, to the X chromosome. The inactivation center with elements capable of counting and choosing the future active and inactive chromosomes was formed around the *Xist* gene, which made random repression of one of the two X chromosome possible. Furthermore, the formation of the more complete and stable inactivation in eutherians was promoted by the involvement of the mechanisms of DNA methylation in the maintenance of the inactive status and enrichment of the X chromosome in LINE1 sequences, which increase the efficiency of propagation of the inactive state. Nevertheless, the evolution of X

chromosome inactivation in mammals remains poorly studied. The hypotheses about its origin and evolution presented in this review are sometimes illogical and too speculative, since they are mainly based on phenomenological data, rather than on the knowledge of the mechanisms, which may differ even when being phenomenologically identical. Further research into the molecular and epigenetic mechanisms of this process could make it possible to better reconstruct the picture of the evolution of the dosage compensation system in mammals. ●

This work was supported by the Russian Foundation for Basic Research (grant № 12-04-31465mol_a), Program of the Russian Academy of Sciences "Molecular and Cellular Biology", and the Ministry of Education and Science of the Russian Federation (Agreement № 8264).

REFERENCES

1. Bininda-Emonds O.R., Cardillo M., Jones K.E., MacPhee R.D., Beck R.M., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. // Nature. 2007. V. 446. № 7135. P. 507–512.
2. Lyon M.F. // Nature. 1961. V. 190. № 4773. P. 372–373.
3. Escamilla-Del-Arenal M., da Rocha S.T., Heard E. // Hum. Genet. 2011. V. 130. № 2. P. 307–327.
4. Grant J., Mahadevaiah S.K., Khil P., Sangrithi M.N., Royo H., Duckworth J., McCarrey J.R., Vandeberg J.L., Renfree M.B., Taylor W., et al. // Nature. 2012. V. 487. № 7406. P. 254–258.
5. Grützner F., Rens W., Tsend-Ayush E., El-Mogharbel N., O'Brien P.C., Jones R.C., Ferguson-Smith M.A., Graves J.A.M. // Nature. 2004. V. 432. № 7019. P. 913–917.
6. Rens W., Grützner F., O'Brien P.C., Fairclough H., Graves J.A., Ferguson-Smith M.A. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 46. P. 16257–16261.
7. Rens W., O'Brien P.C., Grützner F., Clarke O., Graphodatskaya D., Tsend-Ayush E., Trifonov V.A., Skelton H., Wallis M.C., Johnston S., et al. // Genome Biol. 2007. V. 8. № 11. P. R243.
8. Veyrunes F., Waters P.D., Miethke P., Rens W., McMillan D., Alsop A.E., Grützner F., Deakin J.E., Whittington C.M., Schatzkamer K., et al. // Genome Res. 2008. V. 18. № 6. P. 965–973.
9. Deakin J.E., Chaumeil J., Hore T.A., Marshall Graves J.A. // Chromosome Res. 2009. V. 17. № 5. P. 671–685.
10. Deakin J.E., Hore T.A., Koina E., Graves J.A.M. // PLoS Genet. 2008. V. 4. № 7. P. e1000140.
11. Ellegren H., Hultin-Rosenberg L., Brunstrom B., Dencker L., Kultima K., Scholz B. // BMC Biol. 2007. V. 5. P. 40.
12. Itoh Y., Melamed E., Yang X., Kampf K., Wang S., Yehya N., van Nas A., Replogle K., Band M.R., Clayton D.F., et al. // J. Biol. 2007. V. 6. № 1. P. 2.
13. Rens W., Wallduck M.S., Lovell F.L., Ferguson-Smith M.A., Ferguson-Smith A.C. // Proc. Natl. Acad. Sci. USA. 2010. V. 107. № 41. P. 17657–17662.
14. Wrigley J.M., Graves J.A. // J. Hered. 1988. V. 79. № 2. P. 115–118.
15. Kirsch J.A.W., Lapointe F.J., Springer M.S. // Aust. J. Zool. 1997. V. 45. № 3. P. 211–280.
16. Richardson B.J., Czuppon A.B., Sharman G.B. // Nat. New Biol. 1971. V. 230. № 13. P. 154–155.
17. Sharman G.B. // Nature. 1971. V. 230. № 5291. P. 231–232.
18. Cooper D.W., Johnston P.G., Graves J.A.M. // Sem. Dev. Biol. 1993. V. 4. № 2. P. 117–128.
19. Hornecker J.L., Samollow P.B., Robinson E.S., Vandeberg J.L., McCarrey J.R. // Genesis. 2007. V. 45. № 11. P. 696–708.
20. Hayman D.L., Martin P.G. // Genetics. 1965. V. 52. № 6. P. 1201–1206.
21. Johnston P.G., Watson C.M., Adams M., Paull D.J. // Cytogenet. Genome Res. 2002. V. 99. № 1–4. P. 119–124.
22. Dementyeva E.V., Shevchenko A.I., Zakian S.M. // Bioessays. 2009. V. 31. № 1. P. 21-28.
23. Heard E., Disteche C.M. // Genes Dev. 2006. V. 20. № 14. P. 1848–1867.
24. Okamoto I., Patrat C., Thépot D., Peynot N., Fauque P., Daniel N., Diabangouaya P., Wolf J.P., Renard J.P., Duranthon V., Heard E. // Nature. 2011. V. 474. № 7350. P. 239–240.
25. Takagi N., Sasaki M. // Nature. 1975. V. 256. № 5519. P. 640–642.
26. Dindot S.V., Kent K.C., Evers B., Loskutoff N., Womack J., Piedrahita J.A. // Mamm. Genome. 2004. V. 15. № 12. P. 966–974.
27. Mahadevaiah S.K., Royo H., VandeBerg J.L., McCarrey J.R., Mackay S., Turner J.M. // Curr. Biol. 2009. V. 19. № 17. P. 1478–1484
28. Chaumeil J., Waters P.D., Koina E., Gilbert C., Robinson T.J., Graves J.A.M. // PLoS One. 2011. V. 6. № 4. P. e19040.
29. Kohlmaier A., Savarese F., Lachner M., Martens J., Jenuwein T., Wutz A. // PLoS Biol. 2004. V. 2. № 7. P. e171.
30. Wutz A., Rasmussen T. P., Jaenisch R. // Nat. Genet. 2002. V. 30. № 2. P. 167–174.
31. Pullirsch D., Hartel R., Kishimoto H., Leeb M., Steiner G., Wutz A. // Development. 2010. V. 137. № 6. P. 935–943.
32. Wutz A., Jaenisch R. // Mol. Cell. 2000. V. 5. № 4. P. 695–705.
33. Helbig R., Fackelmayer F.O. // Chromosoma. 2003. V. 112. № 4. P. 173–182.
34. Sarma K., Levasseur P., Aristarkhov A., Lee J.T. // Proc. Natl. Acad. Sci. USA. 2010. V. 107. № 51. P. 22196–22201.
35. Hasegawa Y., Brockdorff N., Kawano S., Tsutsui K., Nakagawa S. // Dev. Cell. 2010. V. 19. № 3. P. 469–476.
36. Zakharova I.S., Shevchenko A.I., Shilov A.G., Nesterova T.B., Vandeberg J.L., Zakian S.M. // Chromosoma. 2011. V. 120. № 2. P. 177–183.
37. Dementyeva E.V. Status of gene expression and chromatin modifications in the active and the inactive X chromosomes in common voles. A summary of PhD thesis. Novosibirsk: The Institute of cytology and genetics SB RAS. 2010. 17 P.
38. Chadwick B.P., Willard H.F. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 50. P. 17450–17455.
39. Brinkman A.B., Roelofsen T., Pennings S.W., Martens J.H., Jenuwein T., Stunnenberg H.G. // EMBO Rep. 2006. V. 7. № 6. P. 628–634.
40. Chadwick B.P. // Chromosoma. 2007. V. 116. № 2. P. 147–157.
41. Coppola G., Pinton A., Joudrey E.M., Basrur P.K., King W.A. // Sex Dev. 2008. V. 2. № 1. P. 12–23.
42. Shevchenko A.I., Pavlova S.V., Dementyeva E.V., Zakian S.M. // Mamm. Genome. 2009. V. 20. № 9–10. P. 644–653.
43. Csankovszki G., Nagy A., Jaenisch R. // J. Cell Biol. 2001.

V. 153. № 4. P. 773–784.
44. Zhao J., Sun B.K., Erwin J.A., Song J.J., Lee J.T. // Science. 2008. V. 322. № 5902. P. 750–756.
45. Hellman A., Chess A. // Science. 2007. V. 315. № 5815. P. 1141–1143.
46. Kratzer P.G., Chapman V.M., Lambert H., Evans R.E., Liskay R.M. // Cell. 1983. V. 33. № 1. P. 37–42.
47. Zeng S.M., Yankowitz J. // Placenta. 2003. V. 24. № 2–3. P. 270–275.
48. Reik W., Lewis A. // Nat. Rev. Genet. 2005. V. 6. № 5. P. 403–410.
49. Singh N., Ebrahimi F.A., Gimelbrant A.A., Ensminger A.W., Tackett M.R., Qi P., Gribnau J., Chess A. // Nat. Genet. 2003. V. 33. № 3. P. 339–341.
50. Ohlsson R., Paldi A., Graves J.A.M // Trends Genet. 2001. V. 17. № 3. P. 136–141.
51. Turner J.M., Mahadevaiah S.K., Elliott D.J., Garchon H.J., Pehrson J.R., Jaenisch R., Burgoyne P.S. // J. Cell Sci. 2002. V. 115. № 21. P. 4097–4105.
52. Kalantry S., Purushothaman S., Bowen R.B., Starmer S., Magnuson T. // Nature. 2009. V 460. № 7255. P. 647–651.
53. Chaumeil J., Le Baccon P., Wutz A., Heard E. // Genes Dev. 2006. V. 20. № 16. P. 2223–2237.
54. Clemson C.M., Hall L.L., Byron M., McNeil J., Lawrence J.B. // Proc. Natl. Acad. Sci. USA. 2006. V. 103. № 20. P. 7688–7693.
55. Zhang L.F., Huynh K.D., Lee J.T. // Cell. 2007. V. 129. № 4. P. 693–706.
56. Chureau C., Prissette M., Bourdet A., Barbe V., Cattolico L., Jones L., Eggen A., Avner P., Duret L. // Genome Res. 2002. V. 12. № 6. P. 894–908.
57. Johnston C.M., Newall A.E., Brockdorff N., Nesterova T.B. // Genomics. 2002. V. 80. № 2. P. 236–244.
58. Tian D., Sun S., Lee J.T. // Cell. 2010. V. 143. № 3. P. 390–403.
59. Chureau C., Chantalat S., Romito A., Galvani A., Duret L., Avner P., Rougeulle C. // Hum. Mol. Genet. 2011. V. 20. № 4. P. 705–718.
60. Duret L., Chureau C., Samain S., Weissenbach J., Avner P. // Science. 2006. V. 312. № 5780. P. 1653–1655.
61. Elisaphenko E.A., Kolesnikov N.N., Shevchenko A.I., Rogozin I.B., Nesterova T.B., Brockdorff N., Zakian S.M. // PLoS One. 2008. V. 3. № 6. P. e2521.
62. Hore T.A., Koina E., Wakefield M.J., Graves J.A. // Chromosome Res. 2007. V. 15. № 2. P. 147–161.
63. Davidow L.S., Breen M., Duke S.E., Samollow P.B., McCarrey J.R., Lee J.T. // Chromosome Res. 2007. V. 15. № 2. P. 137–146.
64. Shevchenko A.I., Zakharova I.S., Elisaphenko E.A., Kolesnikov N.N., Whitehead S., Bird C., Ross M., Weidman J.R., Jirtle R.L., Karamysheva T.V., et al. // Chromosome Res. 2007. V. 15. № 2. P. 127–136.
65. Nesterova T.B., Slobodyanyuk S.Y., Elisaphenko E.A., Shevchenko A.I., Johnston C., Pavlova M.E., Rogozin I.B., Kolesnikov N.N., Brockdorff N., Zakian S.M. // Genome Res. 2001. V. 11. № 5. P. 833–849.
66. Lee J.T. // Science. 2005. V. 309. № 5735. P. 768–771.
67. Shevchenko A.I., Malakhova A.A., Elisaphenko E.A., Mazurok N.A., Nesterova T.B., Brockdorff N., Zakian S.M. // PLoS One. 2011. V. 6. № 8. P. e22771.
68. Migeon B.R., Lee C.H., Chowdhury A.K., Carpenter H. // Am. J. Hum. Genet. 2002. V. 71. № 2. P. 286–293.
69. Migeon B.R. // Nat. Genet. 2003. V. 33. № 3. P. 337–338.
70. Graves J.A. // Comp. Biochem. Physiol. A Comp. Physiol. Part A. 1991. V. 99. № 1–2. P. 5–11.
71. Carell L., Willard H.F. // Nature. 2005. V. 434. № 7031. P. 400–404.
72. Forsdyke D.R. // Bioessays. 2012. V. 34. № 11. P. 930–933.
73. Pack S.D., Borodin P.M., Serov O.L., Searle J.B. // Chromosoma. 1993. V. 102. № 5. P. 355–360.
74. Lyon M.F. // Cytogenet. Cell Genet. 1998. V. 80. № 1–4. P. 133–137.
75. Bailey J.A., Carrel L., Chakravarti A., Eichler E.E. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. № 12. P. 6634–6639.
76. Mikkelsen T.S., Wakefield M.J., Aken B., Amemiya C.T., Chang J.L., Duke S., Garber M., Gentles A.J., Goodstadt L., Heger A., Jurka J., Kamal M., Mauceli E., Searle S.M., Sharpe T., Baker M.L., Batzer M.A., Benos P.V., et al. // Nature. 2007. V. 447. № 7141. P. 167–177.
77. Chow J.C., Ciaudo C., Fazzari M.J., Mise N., Servant N., Glass J.L., Attreed M., Avner P., Wutz A., Barillot E., et al. // Cell. 2010. V. 141. № 6. P. 956–969.

Late Replication of the Inactive X Chromosome Is Independent of the Compactness of Chromosome Territory in Human Pluripotent Stem Cells

A. V. Panova, E. D. Nekrasov, M. A. Lagarkova, S. L. Kiselev, A. N. Bogomazova*
Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina Str., 3, Moscow, Russia, 119991
*E-mail: bogomazova@vigg.ru
Received 20.12.2012

ABSTRACT **Dosage compensation of the X chromosomes in mammals is performed via the formation of facultative heterochromatin on extra X chromosomes in female somatic cells. Facultative heterochromatin of the inactivated X (Xi), as well as constitutive heterochromatin, replicates late during the S-phase. It is generally accepted that Xi is always more compact in the interphase nucleus. The dense chromosomal folding has been proposed to define the late replication of Xi. In contrast to mouse pluripotent stem cells (PSCs), the status of X chromosome inactivation in human PSCs may vary significantly. Fluorescence *in situ* hybridization with a whole X-chromosome-specific DNA probe revealed that late-replicating Xi may occupy either compact or dispersed territory in human PSCs. Thus, the late replication of the Xi does not depend on the compactness of chromosome territory in human PSCs. However, the Xi reactivation and the synchronization in the replication timing of X chromosomes upon reprogramming are necessarily accompanied by the expansion of X chromosome territory.**
KEYWORDS **reprogramming; ESCs; iPS cells; chromosome territories; the X chromosome; late replication.**
ABBREVIATIONS **BrdU – 5-bromo-2'-deoxyuridine; DAPI – 4', 6-diamidino-2-phenylindole; ESCs – embryonic stem cells; FISH – fluorescent *in situ* hybridization; H3K27me3 – trimethylation of histone H3 at lysine 27; H3K4me2 – dimethylation of histone H3 at lysine 4; H3K9me3 – trimethylation of histone H3 at lysine 9; HUVEC – human umbilical vein endothelial cells; iPS cells – induced pluripotent stem cells; Xa – active X chromosome; Xi – inactive X chromosome, PSC - pluripotent stem cells.**

INTRODUCTION

The chromatin structure and architecture of the nucleus are the crucial elements in the regulation of transcription and replication, the key genetic processes that occur in nuclei. Chromosomes occupy certain non-overlapping regions in the interphase nucleus, forming the so-called chromosome territories. The densely packed chromatin mostly localizes in the peripheral and perinucleolar regions of the nucleus [1]. The replication of dispersed euchromatin and densely packed heterochromatin is separated both in space and in time. Dispersed euchromatin replicates during the early S-phase, while the condensed heterochromatin replicates in the late S-phase [2]. T. Ryba *et al.* have put forward a hypothesis that late replication of densely packed chromatin domains can be attributed to the fact that access for the replication initiation factors to these regions is hindered [3].

An inactivated X chromosome (Xi), which becomes transcriptionally silent as a result of the dosage compensation, forms the compact structure known as the Barr body on the nuclear periphery and replicates in the late S-phase, is an example of a bulk heterochromatin domain inside the nucleus in female mammalian somatic cells [4].

The variability of the status of X chromosome inactivation in female human pluripotent stem cells (PSCs) provides an interesting opportunity for studying the relationship between the different epigenetic states of chromatin, the architecture of the chromosome territories in the interphase nucleus, and the regulation of replication [5–7].

Up to now, female human embryonic stem cell (hESCs) lines with two active X chromosomes (Xa), one inactivated X chromosome, and hESC lines without any conventional cytological indicators of X inactivation have been described. As for human induced pluripotent stem cells (iPSCs), there is no clear opinion about the possibility of complete reactivation of the X chromosome and the possibility of long-term maintenance

of the active status *in vitro*. However, the X chromosome during reprogramming undoubtedly undergoes a number of significant epigenetic changes associated at least with partial reactivation [8–10]. Our study was aimed at searching for a relationship between the replication timing of the X chromosomes in human PSCs with different statuses of X chromosome inactivation and the degree of compactness of their chromosome territories.

Our results demonstrate that replication of the inactive X chromosome in the late S-phase of the cell cycle can be unrelated to the compactness of the chromosome area and that the late-replicating and transcriptionally silent Xi can be present in female PSCs in the relaxed state. Nevertheless, the X chromosome territory relaxes as the X chromosome becomes active, which is accompanied by synchronization of the replication of homologous X chromosomes.

MATERIALS AND METHODS

Cell cultures

Human ESC lines hESM01 and hESM04 have been described earlier [11]. The cell line HUES 9 was created and kindly provided by D. Melton (Harvard University, USA) [12]. A human umbilical vein endothelial cell (HUVEC) line was obtained in accordance with [13]. iPSC lines (incompletely reprogrammed clones iPS-6 and iPS-7 and completely reprogrammed clone iPS-12) were obtained from HUVEC cells by lentiviral transfection with four transcription factors (KLF4, OCT4, SOX2 and C-MYC) [14] and described in [15]. The iPSC MA-02 line was obtained from dermal fibroblasts according to the previously described procedure [16].

PSC lines were cultured in mTeSR1 medium (StemCell Technologies) on Petri dishes coated with BD Matrigel. The HUVEC line was cultured in DMEM/F12 supplemented with 15% FBS, 5 ng/ml hrbFGF (Peprotech), 20 ng/ml hrVEGF (Peprotech), 1% nonessential amino acids, 2 mM *L*-glutamine, 50 U/ml penicillin, and 50 ng/ml streptomycin (all reagents purchased from Hyclone). All cell lines were cultured in 5% CO_2 at 37˚C.

Immunostaining

The nuclei and metaphase chromosomes were immunostained as described in [8]. The following primary antibodies were used: polyclonal rabbit anti-H3K27me3 antibodies (Millipore, dilution 1 : 500); polyclonal rabbit anti-H3K4me2 antibodies (Abcam, 1 : 200); monoclonal mouse anti-H3K4me2 antibodies (Abcam, 1 : 100); and polyclonal rabbit anti-H3K9me3 antibodies (Abcam, 1 : 200). Alexa Fluor 546 secondary goat anti-rabbit IgG antibodies (Invitrogen, 1 : 1000) or Alexa Fluor 488 goat anti-mouse IgG antibodies (Invitrogen, 1 : 1000) were also used. The DNA of nuclei and metaphase chromosomes was stained with DAPI; Vectashield solution (Vector Laboratories) was then spotted onto the specimen, and the specimen was covered with a cover slip.

RNA FISH

RNA FISH was conducted using fluorescently labeled DNA probes derived from BAC-clones (Empire Genomics) according to the procedure described in [17]. The following BAC clones were used: RP11-13M9 for the *XIST* locus and RP11-1104L9 for the *POLA1* locus.

In situ hybridization with whole chromosome probe (painting)

In order to prepare the specimens of interphase nuclei, cells were treated with trypsin (0.05% Trypsin, Hyclone). Trypsin was inactivated using FBS (Hyclone), and cells were treated with a hypotonic solution (0.075 M KCl) for 18 min at 37˚C. The cells were fixed using two fixatives (6 : 1 and 3 : 1 mixtures of methanol and glacial acetic acid, respectively). The fixed cells were stored in the 3 : 1 fixative at –20˚C. The suspension of fixed cells was dropped onto ice-cold wet glass slides and air-dried for 24 h. In order to improve FISH quality, the specimens were incubated in 0.25% paraformaldehyde for 10 min at room temperature; the pretreated specimens were sequentially dehydrated in 70, 80, and 96% ethanol, followed by treatment with 0.002% pepsin solution (Sigma) in 0.01 M HCl for 30 s at 37˚C and another cycle of dehydration in ethanol. Denaturation was carried out in 70% formamide in the 2xSSC buffer for 5 min at 75˚C, followed by dehydration in ethanol. Whole chromosome probes to chromosomes X and 8 were purchased from MetaSystems. The specimens were subsequently stained with DAPI; the Vectashield solution (Vector Laboratories) was then spotted onto the specimen, and the specimen was covered with a cover slip.

Detection of the replication timing using 5-bromo-2-deoxyuridine (BrdU)

The cells were incubated in the presence of 5-bromo-2-deoxyuridine at a final concentration of 10 μM for 20 min 6–10 h prior to fixation. Colcemid (Demicolcine solution, Sigma) at a final concentration of 0.2 μg/ml was added to the cultivation medium 1 h prior to fixation. The metaphase spreads were prepared according to the above-described procedure for interphase nuclei. For DNA denaturation, the metaphase spreads were treated with 70% formamide in the 2x SSC buffer for 5 min at 75˚C. Next, they were dehydrated in ethanol and incubated with primary mouse anti-BrdU antibodies (Sigma, 1 : 1000) for 2 h at 37˚C. The cells were washed in a PBS-0.1% Tween 20 solution. The specimens were subsequently incubated with Alexa Fluor 546 sec-

ondary goat anti-mouse IgG antibodies (Invitrogen, 1 : 1000) for 1 h at room temperature and stained with DAPI. The X chromosome was identified on the basis of inverted DAPI-banding or X-specific FISH probe.

Microscopy and photography

The metaphase spreads and interphase nuclei were analyzed on a Axio Imager A1 epifluorescence microscope (Carl Zeiss). Pseudo color images of the micro-objects were obtained using the AxioVision software (Carl Zeiss).

Comparison of the degrees of chromosome compactness, calculation of variance

In order to objectively compare the compactness of two X chromosomes in each individual cell we used software that had been designed especially for this experiment. The calculation algorithm is explained below.

Let us consider the channel of a micro-image of interphase nuclei, where the data on the fluorescence intensity of the stained chromosome territories is stored.

Let us assume that x, y are the coordinates on the plane formed by the points of the image, and $P_s(x, y)$ is the intensity of an image point as a function of its coordinate. First, the background level of fluorescence was intercepted by a linear transformation:

$$P(x,y) = k \cdot P_s(x,y) + b.$$

Let us denote the image area where the chromosome territory under analysis lies as G. The parameters P_0, x_0, y_0 were calculated for each chromosome territory as follows:

$$P_0 = \iint\limits_G P\left(x,y\right)dxdy$$

$$x_0 = \frac{1}{P_0}\iint\limits_G \left(x \cdot P\left(x,y\right)\right)dxdy$$

$$y_0 = \frac{1}{P_0}\iint\limits_G \left(y \cdot P\left(x,y\right)\right)dxdy.$$

Next, substitution of variables $P(x, y) \rightarrow P(r, \varphi)$ was performed according to the transformation formulas given below:

$$\begin{cases} x - x_0 = r \cdot \cos(\varphi) \\ y - y_0 = r \cdot \sin(\varphi). \end{cases}$$

This substitution of variables is nothing but a transition to polar coordinates with the center at point (x_0, y_0).

Function $P(r, \varphi)$ was then averaged over the variable φ so that $P(r, \varphi) \rightarrow P(r)$. The formula used for averaging is presented below.

$$P(r) = \frac{\int\limits_0^{2\pi} P(r,\varphi)d\varphi}{2\pi}$$

Function $P(r)$ was normalized by $P(r) \rightarrow P_n(r)$.

$$P_n(r) = \frac{P(r)}{\int\limits_r P(r)dr}.$$

Function $P_n(r)$ was approximated to the normal distribution of $N(r)$ using the least-squares procedure:

$$N\left(r\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{r^2}{2\sigma^2}},$$

in other words, a particular σ value was selected, for which

$$\int\limits_r \left(P_n(r) - N(r)\right)^2 dr \rightarrow \min.$$

The parameter σ^2 is the variance of the normal distribution. This very parameter was output by the software as the analysis result and was subsequently used to estimate the compactness of chromosome territories. The σ^2 values for each X chromosome image in each nucleus in all cell lines were obtained in this study.

The ratio between a chromosome territory with a higher σ^2 value and a chromosome territory with a lower σ^2 value was individually calculated for each nucleus. The comparison of the σ_1^2/σ_1^2 values for two autosomes in the same cell was used as a control and to determine the threshold values of the ratio between the variance of two chromosomes σ_1^2/σ_1^2. The threshold value $\sigma_1^2/\sigma_1^2 = 2.1$ was obtained by comparing autosomes. Thus, all cells with the σ_1^2/σ_1^2 value ≤ 2.1 for the X chromosomes were considered to possess two dispersed chromosome territories. An analysis of the ratio between X chromosome dispersions in pluripotent cells (the line that has previously been described as a line with one inactive X chromosome, hESM04 [11]) allowed to reveal the second threshold value $\sigma_1^2/\sigma_1^2 = 3$. Thus, all the cells with a σ_1^2/σ_1^2 value ≥ 3 for the X chromosomes were considered to possess one compact and one dispersed chromosome territory. The data obtained by this analysis are listed in *Table 2*. A total of 50–100 nuclei were analyzed in each cell line.

Table 1. Summary of X chromosome inactivation in the human pluripotent and somatic cell lines used.

Cell line	H3K4me2 Euchromatin mark	H3K27me3 Heterochromatin mark	H3K9me3 Heterochromatin mark	RNA-FISH		X inactivation status
				XIST-coating	Expression of POLA1	
HUES9	+	-	Some bands were immunopositive	-	Biallelic	XaXa*
hESM01	+	-	+	-	Monoallelic	XaXi *
hESM04	-	+	+	+	Monoallelic	XaXi
iPS MA02	+	-	-	-	Biallelic	XaXa
HUVEC	-	+	+	+	Monoallelic	XaXi
IPS-6	-	+	+	+	Monoallelic	XaXi
IPS-7	-	+	+	+	Monoallelic	XaXi
IPS-12	+	+	+	+	Monoallelic	XaXi

* XaXa – two active X chromosomes in cells; XaXi – one active and one inactivated X chromosome in cells.

RESULTS AND DISCUSSION

Characterization of the status of X chromosome inactivation in human ESCs and iPSCs

The status of the X chromosome in all the cell lines used in this study was characterized previously using the regular criteria: presence/absence of the XIST-RNA cloud in the interphase nucleus, presence/absence of focal staining with anti-H3K27me3 antibodies in the interphase nuclei; and monoallelic/biallelic expression of the *POLA1* gene. The data, including those published earlier for some cell lines [8, 11], are summarized in *Table 1*. Monoallelic or biallelic expression of the *POLA1* gene was the main and determining criterion of the X chromosome status.

It is clear from *Table 1* that identically to the original somatic cells, the incompletely reprogrammed (the so-called imperfect iPSC clones, iPS-6 and iPS-7) cells exhibited all the features of the Xi: the XIST-RNA cloud and focus of H3K27me3 in the interphase nucleus; in addition, monoallelic expression of the *POLA1* gene was observed in these cells. The completely reprogrammed clone iPS-12 exhibited the features of partial reactivation (the presence of H3K4me2 on both X chromosomes) [8] but was characterized by monoallelic *POLA1* expression. The ESC line hESM01 contained the transcriptionally silent X; however, it had lost such inactivation markers as the XIST-RNA cloud and focus of H3K27me3 in the interphase nuclei.

Based on all these criteria, the iPSC line MA-02 and ESC line HUES 9 can be classified as lines with Xa. Neither the XIST cloud was detected via RNA-FISH in all these cell lines nor focal staining with anti-H3K27me3 antibodies in the interphase nuclei. Staining with antibodies against the active chromatin marker H3K4me2 on both X chromosomes and biallelic expression of the *POLA1* gene were observed. It should be mentioned that X chromosome reactivation during the reprogramming of human cells is a rather infrequent event. Most clones produced by regular reprogramming have a single Xi [9, 18, and our own observations].

An analysis of the replication timing of the X chromosomes was carried out for all the ESC and iPSC cell lines listed in *Table 1*.

Replication timing of the X chromosomes

Late replication is known to be typical of heterochromatin; namely, replication in the late S-phase of the

cell cycle after the euchromatin replication. In particular, it is also typical of facultative heterochromatin of the Xi [2]. In order to determine the replicating timing of the X chromosomes in the S-phase of the cell cycle, we conducted an experiment consisting in the incorporation of BrdU into the newly synthesized DNA chains during the replication.

After cell fixation and preparation of metaphase spreads, the incorporated BrdU was detected via immunocytochemical staining with anti-BrdU antibodies.

The patterns of late replication of the X chromosome in all cell lines with the Xi (in the somatic HUVEC cells, in ESC (hESM01 and hESM04) and iPSC (iPS-6, iPS-7, iPS-12) lines) were obtained. *Figure 1* shows two types of incorporation of BrdU into DNA which are observed during the late replication of Xi. Incorporation of BrdU in all chromosomes but one of the X chromosomes is observed in the first variant (*Fig. 1A*). In the second variant, BrdU is incorporated into the pericentromeric heterochromatin and in the p- and q-arms of one of the chromosomes in the metaphase plate – one of the X chromosomes, according to FISH or by inverted DAPI-banding (*Fig. 1B*). The simultaneous replication with the pericentromeric constitutive heterochromatin supports the fact that the observed incorporation type of BrdU corresponds to replication in the late S-phase.

The homologous X chromosomes in PSC lines with two Xa (HUES 9, MA-02) replicate almost synchronously; hence, the homologous X chromosomes are almost indiscernible from one another in terms of the type of BrdU incorporation. *Figure 1C* shows an example of synchronous replication of the X chromosomes.

Next, we decided to estimate the correlation between the replication timing of the X chromosomes and the degree of compactness of their chromosome territories.

The degree of compactness of the X chromosomes in the interphase nucleus does not necessarily correlate with the X chromosome status in human pluripotent stem cells and the timing of their replication

In order to determine the degrees of chromatin condensation, the territories of two X chromosomes of the same nucleus on flat specimens of interphase nuclei were compared. This method of specimen preparation has been used previously when studying chromosome territories [19]. A pair of autosomal chromosomes (chromosome 8) was used as a control. In mammalian cells, the autosomes have the same epigenetic status and the size of their territories is identical. The chromosome territories were compared using the algorithm (see the Materials and Methods section) and by comparing the variance values for each individual

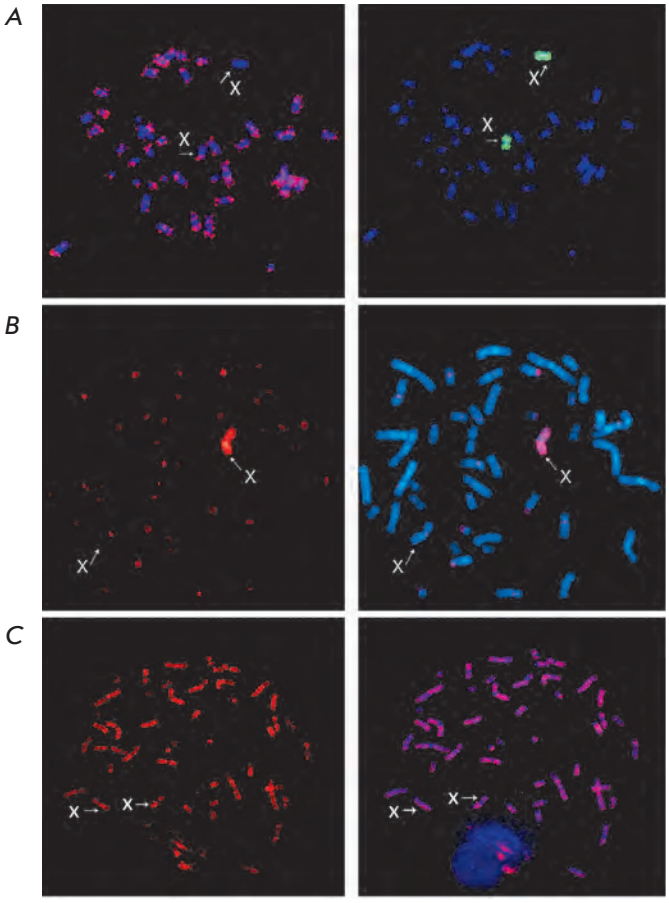


Fig. 1. Replication pattern of X chromosomes in human pluripotent stem cells. X chromosomes are indicated by arrows and letters. A, B – Representative images of asynchronous replication of the X chromosomes. Late-replicating Xi in metaphase spread of hESM04 is shown. A – BrdU (red, left image) is incorporated in all but one chromosome. The right-hand side image represents the same metaphase after FISH with whole X chromosome probe (green). Chromosomes were stained with DAPI (blue). B – BrdU (red, left image) is incorporated only in pericentromeric constitutive heterochromatin and p- and q-arms of a single chromosome. The merged image (right) consists of BrdU (red) and DAPI (blue). X chromosomes were identified by inverted DAPI-banding (not shown). C – Representative images of the synchronous replication of the X chromosomes in the metaphase spreads of HUES9. BrdU (red, left image) is incorporated in all chromosomes but not in pericentromeric heterochromatin. The merged image (right) consists of BrdU (red) and DAPI (blue). X chromosomes were identified by inverted DAPI-banding (not shown)

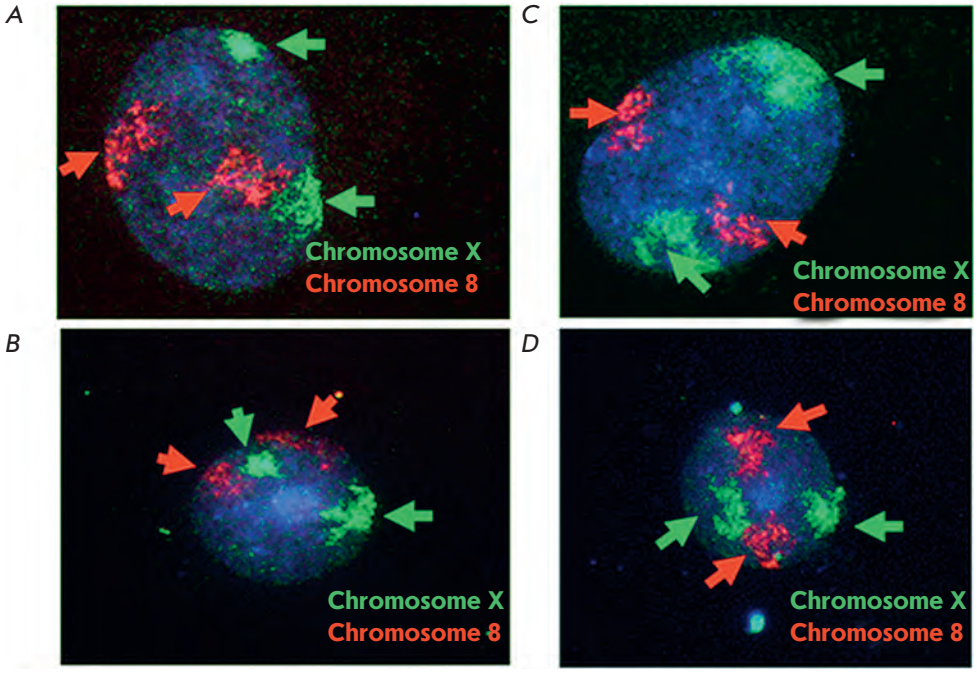


Fig. 2. X chromosome territories in human pluripotent stem cells. A, B – The nucleus of hESM04 has one dispersed and one compact X chromosome territory; C, D – The nucleus of iPS MA02 has two dispersed X chromosome territories. Chromosomes X are green; chromosomes 8 are red; DNA is blue (DAPI); green arrows indicate the X chromosomes, red arrows indicate chromosomes 8

chromosome. The degrees of compactness of the X chromosome territories inside one nucleus were compared using the nuclei on which two non-overlapping zones of DNA probe hybridization to the X chromosome were well-pronounced. The nuclei of pluripotent cells were subdivided into two types without any intermediate forms: (1) nuclei with one dispersed X chromosome territory with a low staining density and one compact territory with a high staining density (*Figs. 2A,B*); (2) nuclei with two dispersed X chromosome territories with the same staining density (*Figs. 2C,D*). The results of the analysis of the distribution of pluripotent cell nuclei on the degree of compactness of the X chromosomes are shown in *Fig. 3*. The frequencies (%) of different types of nuclear organization in all cell lines are shown. The chromosome territories in pluripotent cells with two Xa were dispersed in the overwhelming majority of the analyzed nuclei (over 90%).

In cell lines where the reprogramming process was not complete (i.e., in clones not truly pluripotent: iPS-6 and iPS-7), the third type of chromatin condensation status was also observed. In this case, two X chromosome territories insignificantly differ in size and are characterized by a high degree of chromatin condensation, similarly to the autosomes in these cells and in most HUVEC cells, which originally were an object of reprogramming (*Fig. 3*). It should be mentioned that no differences in the volumes of the territories of the Xa and Xi have been detected in some studies devoted to the investigation of the arrangement of X chromosome territories in somatic cells [2, 20, 21]. The difference between the somatic and pluripotent cells can be attributed to the fact that nuclear chromatin in pluripotent cells is characterized by a considerably higher plasticity as compared with the more compact chromatin in somatic cells [1, 22]. It is interesting to mention that as the iPSC clone iPS-6 was being cultured (passages 5 through 12), the cells lost the chromosome territory of "somatic" type and most cells had one compact and one dispersed X chromosome territory in the later passage.

The results of the comparison of the degree of compactness of the X chromosomes and their replication timing are listed in *Table 2*.

Most nuclei in the hESM04 cell line had one compact and one dispersed X chromosome territory (*Fig. 2*), which was accompanied by a well-pronounced asynchronous replication of the X chromosomes. Nevertheless, no correlation between the late replication of Xi and compactness of the Xi chromosome territory has been observed in the other pluripotent cell lines.

As can be seen in *Table 2*, despite the fact that both X chromosome territories in the ESC hESM01 line were dispersed, replication of the X chromosomes in this cell line occurred asynchronously, with one of the X chromosomes replicating in the late S-phase. In the iPSC clone iPS-12, half of the nuclei had two identical, dispersed X chromosome territories; nevertheless, despite the partial cytological signs of reactivation (the pres-

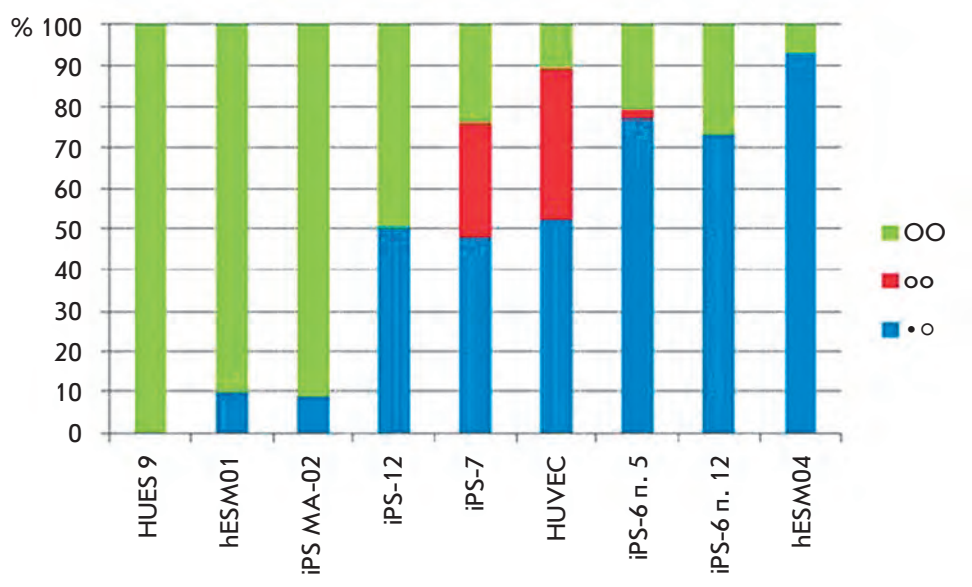


Fig. 3. Frequencies of nuclei with different organizations of X chromosome territories in human pluripotent stem cells and in HUVEC. •O – nuclei with one dispersed and one compact X chromosome territory (blue); OO – nuclei with two dispersed X chromosome territories (green); oo – nuclei with two small compact X chromosome territories ("somatic" type) (red)

Table 2. Comparison of chromosome territory folding and replication patterns of Xs

Cell lines	•O, %*	oo, %*	OO, %*	Replication pattern
HUES9	0		100	Synchronous, early
hESM04	93		7	Asynchronous, late replication of Xi
hESM01	10		90	Asynchronous, late replication of Xi
HUVEC	52	37	11	Asynchronous, late replication of Xi
IPS-7	48	28	24	Asynchronous, late replication of Xi
IPS-6 p.5	77	2	21	Asynchronous, late replication of Xi
IPS-6 p.12	73		27	Asynchronous, late replication of Xi
IPS-12	50		50	Asynchronous, late replication of Xi
iPS MA02	3		97	Synchronous, early

*•O – nuclei with one dispersed and one compact X chromosome territory; oo – nuclei with two small compact X chromosome territories ("somatic" type); OO – nuclei with two dispersed X chromosome territories.

ence of H3K4me2, *Table 1*), one of the X chromosomes replicated in the late S-phase in all cells analyzed ($N > 20$) and possessed the Xi status.

The ESC HUES 9 and iPSC MA-02 lines, in which the homologous X chromosomes replicated almost synchronously and neither of the X chromosome was characterized by late replication, had two dispersed X chromosome territories (in all nuclei for HUES 9 and in most nuclei for MA-02) (*Table 2*).

Thus, summarizing the data obtained for all the cell lines used in this study, one can assume that the Xi status can be retained in pluripotent cells without the formation of the conventional "Barr body" (i.e., the compactly packed chromosome territory). Nevertheless, the active status of both X chromosomes and the transition to synchronous replication during re-programming require a dispersed status of both X chromosomes in the interphase nucleus.

During the reprogramming to a pluripotent state in the cases when partial or complete Xi reactivation occurs, dispersion of the chromosome territory is likely to take place before the major heterochromatin marks

H3K27me3 and H3K9me3 disappear (as it can be seen by the example of the iPSC clone iPS-12).

It has been demonstrated that long-term cultivation of partially reprogrammed iPSCs frequently promotes the completion of reprogramming, the acquisition of a pluripotent state, and a loss of characteristics by somatic cells. The change in the compactness of X chromosome territories, which occurred during the cultivation of the iPSC clone iPS-6, demonstrates that this parameter can be an additional marker of the reprogramming of somatic cells to a pluripotent state.

CONCLUSIONS

It has been demonstrated in our studies that human PSCs with Xi can possess either dispersed or compact chromosome territory of Xi. ESC lines with both X chromosomes active or iPSC lines in which the X chromosome has been reactivated during reprogramming have dispersed both of the X chromosome territories in the interphase nucleus.

Thus, a conclusion can be drawn that human PSCs have a mechanism that allows them to maintain an inactivated status of the X chromosome, which does not depend directly on the degree of compactness of its chromosome territory in the interphase nucleus. Late replication of the inactivated X chromosome is also independent of its degree of compactness.

On the other hand, the X chromosome territory has to be dispersed for the X chromosome to become active during its reactivation and synchronous replication. ●

This work was supported by the Program of the Moscow City Government and the Russian Foundation for Basic Research (grants № 11-04-01212-a and 10-04-01736-a).

REFERENCES

1. Bártová E., Galiová G., Krejcí J., Harnicarová A., Strasák L., Kozubek S. // Dev.Dyn. 2008. V. 237. P. 3690–702.
2. Visser A. E., Eils R., Jauch A., Little G., Bakker P. J., Cremer T., Aten J. A. // Exp. Cell Res. 1998. V. 243. P. 398–407
3. Ryba T., Hiratani I., Lu J., Itoh M., Kulik M., Zhang J., Schulz T. C., Robins A.J., Dalton S., Gilbert D.M. // Genome Res. 2010. V. 20. P. 761–770.
4. Plath K., Mlynarczyk-Evans S., Nusinow D.A., Panning B. // Annu.Rev.Genet. 2002. V. 36. P. 233–278.
5. Hoffman L. M., Hall L., Batten J. L., Young H., Pardasani D., Baetge E. E., Lawrence J., Carpenter M.K. // Stem Cells. 2005. V. 23. P. 1468–1478.
6. Silva S.S., Rowntree R.K., Mekhoubad S., Lee J.T. // Proc. Natl.Acad.Sci. 2008. V. 105. P. 4820–4825.
7. Shen Y., Matsuno Y., Fouse S. D., Rao N., Root S., Xu R., Pellegrini M., Riggs A. D., Fan G. // Proc.Natl.Acad.Sci. 2008. V. 105. P. 4709–4714.
8. Lagarkova M.A., Shutova M.V, Bogomazova A.N., Vassina E.M., Glazov E.A., Zhang P., Rizvanov A.A., Chestkov I.V., Kiselev S.L. // Cell Cycle. 2010. V. 9. P. 937–946.
9. Tchieu J., Kuoy E., Chin M.H., Trinh H., Patterson M., Sherman S.P., Aimiuwu O., Lindgren A., Hakimian S., Zack J.A, et al. // Cell Stem Cell. 2010. V. 7. P. 329–342.
10. Marchetto M.C., Carromeu C., Acab A., Yu D., Yeo G.W., Mu Y., Chen G., Gage F.H., Muotri A.R.. // Cell. 2010. V. 143. P. 527–539.
11. Lagarkova M.A., Eremeev A.V., Svetlakov A.V., Rubtsov N.B., Kiselev S.L. // In Vitro Cell Dev. Biol. Anim. 2010. V. 46. P. 284–293.
12. Cowan C.A., Klimanskaya I., McMahon J., Atienza J., Witmyer J., Zucker J.P., Wang S., Morton C.C., McMahon A.P., Powers D. et al. // N. Engl. J. Med. 2004. V. 350. P. 1353–1356.
13. Baudin B., Bruneel A., Bosselut N., Vaubourdolle M. // Nat.Protoc. 2007. V. 2. P. 481–485.
14. Takahashi K., Yamanaka S. // Cell. 2006. V. 126. P. 663–676.
15. Shutova M.V., Chestkov I.V., Bogomazova A.N., Lagarkova M.A., Kiselev S.L. // Springer Protocols Handbook series 2012. P.133–149.
16. Shutova M.V., Bogomazova A.N., Lagarkova M.A., Kiselev S.L. // Acta Naturae. 2009. V. 1. P. 91–92.
17. Bacher C. P., Guggiari M., Brors B., Augui S., Clerc P., Avner P., Eils R., Heard E.. // Nat.Cell Biol. 2006. V. 8. P. 293–299.
18. Bruck T., Benvenisty N. // Stem Cell Res. 2011. V. 6. P. 187–93.
19. Federico, C., Cantarella, C. D., Di Mare, Tosi S., Saccone S. // Chromosoma. 2008. V. 117. P. 399–410
20. Eils R, Dietzel S, Bertin E, Schröck E., Speicher M.R., Ried T., Robert-Nicoud M., Cremer C., Cremer T. // Cell Biol. 1996. V.135. P.1427–1440.
21. Meshorer, E., Yellajoshula, D., George, E., Scambler P.J., Brown D.T., Misteli T. // Developmental cell. 2006. V. 10. P. 105–116

Mycobacterium tuberculosis Transcriptome Profiling in Mice with Genetically Different Susceptibility to Tuberculosis

T.A. Skvortsov[1*], D.V. Ignatov[1], K.B. Majorov[2], A.S. Apt[2], T.L.Azhikina[1]
[1]Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Str., 16/10, Moscow, Russia, 117997
[2]Central Institute for Tuberculosis, Yauza Alley, 2, Moscow, Russia, 107564
*E-mail: timofey@ibch.ru
Received 18.12.2012

ABSTRACT **Whole transcriptome profiling is now almost routinely used in various fields of biology, including microbiology.** *In vivo* **transcriptome studies usually provide relevant information about the biological processes in the organism and thus are indispensable for the formulation of hypotheses, testing, and correcting. In this study, we describe the results of genome-wide transcriptional profiling of the major human bacterial pathogen** *M. tuberculosis* **during its persistence in lungs. Two mouse strains differing in their susceptibility to tuberculosis were used for experimental infection with** *M. tuberculosis***. Mycobacterial transcriptomes obtained from the infected tissues of the mice at two different time points were analyzed by deep sequencing and compared. It was hypothesized that the changes in the** *M. tuberculosis* **transcriptome may attest to the activation of the metabolism of lipids and amino acids, transition to anaerobic respiration, and increased expression of the factors modulating the immune response. A total of 209 genes were determined whose expression increased with disease progression in both host strains (commonly upregulated genes, CUG). Among them, the genes related to the functional categories of lipid metabolism, cell wall, and cell processes are of great interest. It was assumed that the products of these genes are involved in** *M. tuberculosis* **adaptation to the host immune system defense, thus being potential targets for drug development.**
KEYWORDS *Mycobacterium tuberculosis*; **transcriptome** *in vivo*; **next generation sequencing; RNA-seq; tuberculosis.**
ABBREVIATIONS **PDIM – phthiocerol dimycocerosate; CC – coincidence cloning; CUG – commonly upregulated genes.**

INTRODUCTION

Genome-wide expression studies generate enormous amounts of data and have a virtually boundless potential for application. For instance, the data obtained in such experiments can be used to assess the effectiveness of antibiotics or to study changes in the bacterial metabolism during the course of an infection. Among the infections caused by pathogenic bacteria, tuberculosis ranks first in terms of mortality rate and is responsible for 1.5 million deaths annually. It is not surprising that the first study of the transcriptome of its causative agent, *Mycobacterium tuberculosis*, was conducted within a year after the publication of the whole-genome sequencing data for this bacterium [1, 2]. Five years later, a large number of studies describing the results of using microarrays for the transcriptome analysis of mycobacteria under various conditions were published [3, 4]. Nevertheless, although microarrays and massively parallel sequencing have been widely used for microbiological studies and adapted to the study of whole-genome expression in mycobacteria, the majority of such studies have been carried out *in vitro*. Meanwhile, it is rather difficult to carry out an analysis of gene expression in mycobacteria during the progression of tuberculosis *in vivo*, which is of the greatest research interest [5, 6]. The results of a study of gene expression in *M. tuberculosis* during experimental infections in two strains of mice differing in susceptibility to tuberculosis are presented in this work. Since the genotypic differences of animals directly affect gene expression of the pathogen, an increase in the expression of a number of bacterial genes under more unfavorable conditions (in the organism of

a host resistant to the infection) attests to the fact that these genes significantly contribute to the adaptation of *M. tuberculosis* to the host's defense mechanisms.

EXPERIMENTAL PROCEDURES

Experimental infection and RNA extraction

I/StSnEgYCit (I/St) and C57BL/6YCit (B6) mice were kept under standard conditions in the vivarium of the Central Research Institute of Tuberculosis, Russian Academy of Medical Science, in accordance with Order of the USSR Ministry of Health № 755 dated August 12, 1977, and the NIH Office of Laboratory Animal Welfare Assurance № A5502-11. The mice had *ad libitum* access to food and water. All the experimental procedures were approved by the Bioethics Committee of the Central Research Institute of Tuberculosis.

Female mice of both strains aged 2.5–3.0 months were infected with the virulent *M. tuberculosis* strain H37Rv using an inhalation exposure system (Glas-Col, Terre Haute, USA) at a dose of 100–200 CFU/mouse. The infected animals were euthanized 4 and 6 weeks after their infection. Their lung tissues were immediately used to extract RNA. Total RNA was extracted using the SV Total RNA Isolation System (Promega, USA). The RNA samples were then treated with DNAse I (MBI Fermentas, Lithuania) to remove trace amounts of DNA.

Synthesis of cDNA

cDNA was synthesized using the template-switching effect (Clontech, USA) according to the procedure described in [7]. cDNA was synthesized using PowerScript II reverse transcriptase (Clontech, USA) following the manufacturer's protocol. The oligonucleotide primers BR (5'AAGCAGTGGTATC AACGCAGAGTAC(N)9) and SMART (5'AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG) were added into 2 μg of total RNA from each sample in 11 μl of a buffer solution. The resulting mixture was incubated for 2 min at 70°C and placed on ice for 10 min. The mixture kept in the ice bath was supplemented with 11 μl of a solution containing 4 μl of a 5× reverse transcriptase buffer, 2 μl of a solution of 100 mM DDT, 2 μl of a solution of 10 mM of each dNTP, and 1 μl (200 AU) of PowerScript reverse transcriptase (Clontech, USA). The control reaction (RT–) with no reverse transcriptase added was conducted simultaneously with the reverse transcriptase reaction (RT+). The reaction mixtures RT+ and RT– were incubated at 37°C for 10 min and for an additional 40 min at 42°C. 30 cycles of PCR (95°C 20 s; 64°C, 20 s; 72°C, 2 min) with the 5S primer (5'GTGGTATCAACGCAGAGT) were used to amplify the synthesized cDNA. The amplified cDNA was purified using the QIAquick PCR Purification kit (Qiagen, USA).

Coincidence cloning

Coincidence cloning was performed according to [8]. The genomic DNA of *M. tuberculosis* H37Rv and the total cDNA samples (i.e., cDNA synthesized using total RNA as a template) were digested with the restriction endonucleases RsaI and AluI (MBI Fermentas, Lithuania). The suppression adapters were ligated to the resulting fragments of the genomic DNA and cDNA (adapter structures were given in [7]). The mixture containing 100 ng of the sample of genomic DNA with adapters and 100 ng of the corresponding sample of cDNA fragments with adapters in 2 μl of the hybridization buffer HB (50 mM HEPES, pH 8.3; 0.5 M NaCl; 0.02 mM EDTA, pH 8.0) was incubated at 99°C for 5 min (denaturation) and subsequently at 68°C for 18 h (re-naturation). Next, the reaction mixture was supplemented with 100 μl of the hybridization buffer HB preliminarily heated to 68°C. 1 μl of the resulting mixture was used as a template in a two-step PCR. The first PCR step was performed in a reaction volume of 25 μl containing 10 pmol of the outer primer T7. After the mixture was preincubated at 95°C for 5 min, 20 amplification cycles were performed (94°C, 30 s; 66°C, 30 s; 72°C, 90 s). The second amplification step was carried out using inner primers (94°C, 30 s; 68°C, 30 s; 72°C, 90 s; 25 cycles) [7]. The tenfold diluted amplicon obtained in step 1 was used as a template for the second amplification step. The amplicon obtained in step 2 was purified using the QIAquick PCR Purification kit (Qiagen, USA) and used for massively parallel sequencing.

Massively parallel sequencing

Prior to sequencing, equal amounts of each of the coincidence cloning products (500 ng of each amplicon) were combined to obtain a single sample. Massively parallel pyrosequencing of the sample was performed on the GS FLX genetic instrument (454 Roche, Germany). The resulting sequences (reads) were tested for the presence of adapter sequences. The reads with truncated, incorrect, or missing adapter sequences were eliminated from further analysis. The remaining sequences (190, 031 reads) were divided into three groups (libraries) depending on the nucleotide sequence of their adapters; CC6(SUS) represented transcriptomes of *M. tuberculosis* from the lung tissues of I/St mice on week 6 post-infection; (CC4(RES) and CC6(RES) represented transcriptomes of *M. tuberculosis* from lung tissues of B6 mice on weeks 4 and 6 post-infection, respectively. A file in FASTA format, which contained nucleotide sequences from all three groups, was used for further analysis.

Table 1. Results of sequencing and library mapping for CC4(RES), CC6(RES), and CC6(SUS)

Library	CC4(RES)	CC6(SUS)	CC6(RES)
Total reads	73410	75655	40966
Mtb-specific reads (unique)	14990	43618	34234
Mtb-specific reads (unique), % of the total number	20.4	57.7	83.6
Genes expressed (number of reads > 0)	1012	1353	1940
Genes expressed, % of the total number of genes	25.2	33.7	48.3
IGR expressed (number of reads > 0)	164	221	376
IGR expressed, % of the total number of genes	5.3	7.2	12.3

The authors can provide the FASTA files upon request.

Sequence mapping and statistical analysis

The resulting sequences were mapped onto the genomic sequence of *M. tuberculosis* H37Rv (assembly GenBank AL123456.2) using the blastn tool from the NCBI BLAST+ software package with the parameters set as follows: -perc identity 95 and -evalue 0.01. The sequences with at least 95% identity to the genome sequence of *M. tuberculosis* H37Rv for the entire length were eventually selected. All the sequences shorter than 40 nucleotides were eliminated from processing. The sequences mapped onto the *M. tuberculosis* genome in a non-unique manner (at two sites and more) were also eliminated from further analysis. The reads mapped onto the intergenic sequences were not eliminated from further analysis, since it would have changed the library size and caused errors during the subsequent statistical analysis. Next, the number of reads in each library was determined for each gene and intergenic region, and the comparative analysis of the abundance of cDNA fragments corresponding to the bacterial genes and intergenic regions was conducted using the Audic-Claverie algorithm [9]. The differences in the expression of the genes (intergenic regions) were considered to be significant if the number of reads mapped onto the gene (intergenic region) in at least one of the two libraries being compared was no less than 20 and $p < 0.05$.

RESULTS AND DISCUSSION

In order to reveal the features of the expression profile of *M. tuberculosis* that correlate with infection progression, comparative quantitative and qualitative analyses of the sequences transcribed in infected mice, genetically susceptible (inefficient immune response), and resistant (efficient response) to these bacteria were conducted at different stages of the infectious process.

We have compared the transcriptomes of *M. tuberculosis* H37Rv in infected mice of two strains, I/StSn-EgYCit (I/St) and C57BL/6YCit (B6). These mouse strains had been thoroughly described earlier [10]. In B6 mice, resistance to *M. tuberculosis* is higher as compared to I/St mice, which manifests itself in the less aggressive course of the infectious process in B6 mice and the longer survival of the infected animals.

Female mice of both strains were euthanized 4 and 6 weeks after they had been aerogenically infected with *M. tuberculosis* bacteria to isolate the total RNA from their lungs. The total RNA samples from the lung tissues in I/St and B6 mice were used to synthesize cDNA subsequently enriched in bacterial cDNA fragments via coincidence cloning [8]. A total of three sequence libraries representing *M. tuberculosis* transcriptomes were obtained from the tissues of I/St mice on week 6 after the infection (CC6(SUS)) and the tissues of B6 mice on weeks 4 and 6 after they had been infected (CC4(RES) and CC6(RES), respectively).

The nucleotide sequences of the cDNA fragments of these libraries were determined using 454 massively parallel pyrosequencing. The general scheme of the experiment is shown in *Figure*; the general characteristics of the analyzed libraries are listed in *Table 1*. The nucleotide sequences of 190, 031 cDNA fragments were identified. Among them, 73, 410 sequences were from the CC4(RES) library; 75, 655 were from the CC6(SUS) library; and 40, 966, from CC6(RES). The resulting sequences were mapped onto the genomic sequence of *M. tuberculosis* H37Rv (Assembly GenBank AL123456.2) using the blastn command line tool from the NCBI BLAST+ software package.

The mapping revealed that the CC4(RES), CC6(SUS) and CC6(RES) libraries contained 14, 990 (20.42%),

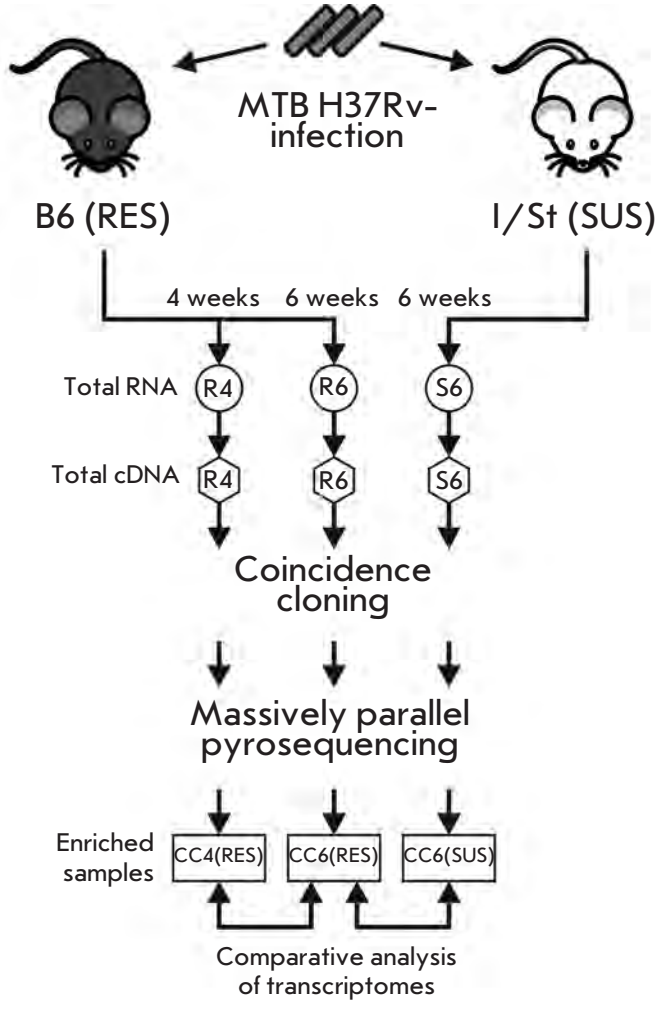


Fig. 1. Scheme of transcriptome comparison. RES – genetically resistant mice, SUS – genetically susceptible mice, CC – library enriched in bacterial cDNA

43, 618 (57.65%) and 34, 234 (83.57%) *M. tuberculosis* sequences, respectively. These data attest to the fact that a significant enrichment of cDNA in the bacterial sequences was attained.

Among 4,012 genes and seven pseudogenes of *M. tuberculosis*, 1,012 (25.2% of the total number of genes), 1,353 (33.7%), and 1,940 genes (48.3%) were expressed in the CC4(RES), CC6(SUS), and CC6(RES), respectively. 1,428 (35.5%) genes were expressed in none of the samples, while 469 (11.7%) genes were expressed in each sample.

Noncoding RNAs

The recent studies employing the methods of massively parallel sequencing have demonstrated that the bacterial transcriptome is much more complex than it was thought to be. The noncoding RNAs have been shown to be an important part of the transcriptome and to regulate various cell functions (replication, energy metabolism, and regulation of the expression of virulence factors in a number of pathogenic bacteria) [11].

We have searched for transcripts from the loci lying in the intergenic regions, since such localization attests to the fact that these transcripts have the potential to belong to the class of noncoding RNAs. Based on data on the structural organization of the genome, we have identified 3,069 intergenic regions. The analysis of pyrosequencing data showed that the transcripts of 164 (5.3%), 221 (7.2%), and 356 (12.3%) intergenic regions were presented in the CC4(RES), CC6(SUS), and CC6(RES) samples, respectively. . It should be mentioned that transcripts from 27 (0.9%) of the intergenic loci were present in each of the three samples, while transcripts from 2,490 (81.1%) loci have not been observed.

The experimental data on the expression of intergenic regions were compared with the available literature and database information on the localization of small RNA genes.

A significant expression of sequences from a number of intergenic regions (in particular, from IGR3987 and IGR0629 in sample CC6(SUS) and from IGR1186 in sample CC6(RES)) was detected. The experimental results confirmed the presence of expression from the intergenic regions IGR3987, IGR0629, and IG1136, which correlates with the data obtained by Arnvig *et al.* [12] (according to the denotations used in [12], the intergenic regions IGR2975, IGR0479, and IGR0858, respectively). We believe that the detected expression of intergenic sequences can attest to the potential localization of small RNA genes in these loci. With allowance for their differential expression, it could be an indication of the compensatory response of a pathogen to external factors.

Genes whose expression increases with infection progression

We compared the transcriptomes of *M. tuberculosis* during the infection progression in the genetically resistant to tuberculosis mouse strain (CC6(RES) and CC4(RES)) and at a single time point in the genetically different mouse strains (CC6(RES) and CC6(SUS)). The comparison was aimed at searching for genes whose expression increases with infection progression (i.e., in B6 mice on week 6 after the infection) as compared to the other conditions. The comparison of CC6(RES) and CC4(RES) revealed 226 genes whose expression increased with infection progression in the tissues of B6 mice. The comparison of CC6(RES) and CC6(SUS) revealed 253 genes with increased expression in the CC6(RES) sample.

The comparison of CC6(RES) and CC4(RES) revealed only 17 genes whose expression in CC6(RES) was higher than that in CC4(RES), while the expression of 44 genes in CC6(RES) was higher as compared to CC6(SUS). These results presumably demonstrate that the first comparison characterizes the dynamics of the changes in pathogen gene expression with time within a single micro-environment. In the latter case, the differences between two different micro-environments are revealed, which affects a larger number of genes whose expression increases in CC6(RES).

The genes whose expression in the CC6(RES) sample is increased only as compared to CC4(RES) mostly fall into the following categories: cell wall and cell processes, intermediary metabolism and respiration, and lipid metabolism. The protein products of 12 out of 17 genes were previously detected in the fraction containing the cellular membrane and/or cell wall, where they predominantly play roles in transport and protection. For instance, the *embA* gene encodes indolylacetylinositol arabinosyltransferase EmbA, which participates in arabinan synthesis. Mutations in this gene cause resistance to ethambutol. The *Rv3273* gene encodes carbonate dehydratase, which is involved in sulfate transport (TubercuList). An analysis using the KEGG Pathways (http://www.genome.jp/kegg/pathway.html) and TBCYC (http://tbcyc.tbdb.org/) databases has revealed no metabolic pathways that would be activated at the later stages of infection progression. This fact can either result from random fluctuations in pathogen gene expression, its response to random changes in the properties of the micro-environment, or it can be an indication of small, but significant changes in the functional activity of *M. tuberculosis* at different time points.

The comparison of the CC6(RES) and CC6(SUS) samples has revealed that CC6(RES) contains a larger number of genes whose expression is higher in this sample as compared to that in CC6(SUS). The increased energy metabolism manifested itself as increased expression of the genes encoding three NADH-dehydrogenase subunits (*nuoH, nuoI, nuoL*); a higher activity of the tricarboxylic acid cycle (*acn*); and as increased expression of the *Rv1916* gene. *Rv1916* is the second component of the *aceA* (*icl2*) gene, which is divided in the genome of *M. tuberculosis* H37Rv into two individually expressed modules, *Rv1915* and *Rv1916* (*aceAa* and *aceAb*). The other important differences include the increased expression of the genes whose products are responsible for the metabolism and catabolism of lipids and amino acids (*lipV, lipF, Rv2531c*), as well as the enzymes that participate in DNA repair (*recO, recB*). This is a rather predictable pattern, since the microenvironment of the resistant host is a hostile habitat, which explains the demand for higher activity from the repair systems. The increased expression of lipolytic enzymes (*lipF, lipV, plcA*), enzymes of the tricarboxylic acid cycle, and *aceAb* may attest to the fact that a greater amount of lipids is used as a source of energy and carbon.

We focused on searching for *M. tuberculosis* genes whose increased expression does not depend on the genetic features of the host organism. These genes form a certain basic set, whose genes are responsible for the universal compensatory response of the pathogen to adverse environmental conditions. These genes are referred to as commonly upregulated genes (CUG): a total of 209 genes whose expression was increased in both comparisons (*Table 2*). According to the results of transposon mutagenesis, 44 genes of *M. tuberculosis* H37Rv are essential [13]. Rv3569c, Rv3537, and Rv3563 were earlier shown to be essential for survival of *M. tuberculosis* in mouse macrophages (TubercuList, http://tuberculist.epfl.ch).

We have grouped the GUG genes into functional categories (TubercuList) and compared their distribution to that of all the *M. tuberculosis* genes. These distributions show a general similarity, with the exception of the genes that belong to the lipid metabolism category, which can further demonstrate their importance for the processes of bacterial adaptation.

Two categories (conserved hypotheticals (59 genes) and unknown (2 genes) contained slightly less than a third of all the genes. Despite the absence of any known functions, the genes in this category are potential therapeutic targets, since their low degree of homology with the genes of other microorganisms means that they are typical of mycobacteria (in particular, of *M. tuberculosis*); thus, they presumably determine their virulence properties.

The high expression level of the genes of various nutrient uptake and accumulation systems (e.g., phosphate (*pstS1*) and iron (*irtA, mbtC, mbtE, mbtF*)) attests to the fact that mycobacteria exist under conditions of insufficient nutrient supply. Phosphorus deficiency is also indicated by the increased expression of the *senX3* gene, the sensor component of the two-component regulatory system *senX3/regX3*, which regulates the so-called stringent response under conditions of phosphorus deficiency. The transition to the use of lipids as the main source of energy and carbon is indicated by the expression of lipid metabolism genes (*fadD, fadE, lipU, lipJ*). Another feature of the genes that belong to the CUG set consists in increased expression of genes whose products are somehow associated with amino acid metabolism (*aspC, hisB, thrB, thrS*). The reason behind this phenomenon remains unclear, since stimulation of enzyme expression can be caused by both the absence of the required amino acids (and, hence, the

Table 2. GUC gene family

Gene	Functional category (according to TubercuList)
Rv0028, Rv0074, Rv0269c, Rv0274, Rv0281, Rv0421c, Rv0428c, Rv0433, Rv0448c, Rv0455c, Rv0492A, Rv0525, Rv0597c, Rv0695, Rv1179c, Rv1186c, Rv1203c, Rv1232c, Rv1419, Rv1428c, Rv1828, Rv1835c, Rv1868, Rv1998c, yfiH, Rv2974c, Rv3030, Rv3031, Rv3205c, Rv3272, Rv3519, Rv3627c, Rv3651, Rv3662c, Rv3703c, Rv3753c, Rv0026, Rv0061, Rv0140, Rv0141c, Rv0145, Rv0332, Rv0712, Rv0785, Rv0998, Rv1514c, Rv1515c, wbbL2, Rv1760, Rv2077A, Rv2135c, Rv2466c, Rv2699c, Rv2751, Rv2823c, Rv3067, Rv3090, Rv3094c, Rv3510c	CH – conserved hypotheticals
Rv0051, Rv0309, lprL, Rv0621, Rv0876c, lytB2, irtA, Rv1687c, secA2, Rv2209, Rv2265, mmpL7, Rv3194c, Rv3658c, embC, espE, ponA1, Rv0072, narK3, iniA, cpsY, lpqR, pstS1, Rv0996, kdpC, Rv1097c, sugB, Rv1431, Rv1667c, Rv2136c, Rv2203, efpA, rip, Rv2963, lpqF	CWaCP – cell wall and cell processes
Rv0161, ndhA, Rv0526, menH, Rv0805, lipU, glyA1, dapE, atpF, atpH, Rv1432, frdB, cmk, plcD, lipJ, cobK, cobS, cysK1, cysE, gdh, gabT, miaA, ilvC, guaB2, cyp142, hsaD, Rv0089, Rv0331, aspC, hemA, Rv0567, atsA, gltA2, Rv0943c, Rv1096, Rv1106c, narH, thrB, hisB, ilvG, rocD1, plcB, phoH1, ggtB, lepA, Rv2499c, dapF, purU, kstD, folP1	IMaR – intermediary metabolism and respiration
end, fusA1, polA, lysX, helZ, spoU, ppiB, thrS, Rv3201c	IP – information pathway
Rv0095c, Rv0920c, Rv2791c	ISaP – insertion sequences and phages
fadD10, nrp, fadD7, fadE4, fadD2, fadD12, pks17, pks12, mbtF, mbtE, mbtC, TB7.3, accA3, fadE27, fadD17, accD4, mmaA3, mmaA1, fadE19, Rv2613c, fadD26, ppsC, ppsD, fadD19, fadE31, fadE32, pks13	LM – lipid metabolism
PPE8, PE_PGRS19, PE16, PPE34, PPE50, PE2, PPE64	PE/PPE – PE/PPE protein families
pknA, senX3, trcR, Rv1359, fhaA, Rv0465c, Rv3066, Rv3736	RP – regulatory proteins
Rv2645, Rv2818c	U – unknown
treS, mce2C, Rv1026, ephB, vapB16, Rv2581c, vapC3, cinA, virS	VDA – virulence, detoxification, adaptation

demand for their synthesis) and by their presence (and the possibility to use them by bacteria).

The transition to anaerobic nitrate respiration, which is typical of latent infection, is attested by the increased expression of the *narH* and *narK3* genes [14]. We have also classified the *atpF* and *atpH* genes as CUG genes, although decreased expression of these genes during infection progression is reported, since the energy demand of a pathogen decreases as it acquires the state of latent infection [15, 16].

The function of PE/PPE proteins is not clear. They are considered to be required to ensure antigenic variability in mycobacteria [17]. Nevertheless, the expression levels of the *Rv0152c* and *Rv0355c* genes in the CC6(RES) sample are high, and their expression has also been detected in CC4(RES) and CC6(SUS). The *Rv3135* gene belongs to the group of essential genes in *M. tuberculosis* H37Rv. All these observations suggest that PE/PPE genes have some other functions in addition to ensuring antigenic variability.

Finally, the *secA2*gene is worth mentioning. This gene encodes translocase SecA2, the component of the *M. tuberculosis* Sec transport system that enables the secretion of superoxide dismutase SodA and catalase KatG. A live vaccine based on the *ΔsecA2* mutant of *M. tuberculosis* has demonstrated high efficiency and safety in tests on animals [18].

A microarray-based comparative analysis of the expression profiles of 17 members of the *M. tuberculosis* complex in activated and nonactivated murine macrophages has recently been conducted[19]. As a result, 280 genes (168 with universally increased and 112 genes with universally decreased expression) were identified whose changes in expression were independent of the strain and the activation status of a macrophage. We compared the CUG genes with 168 genes from [19] characterized as having universally increased expression and selected eight genes *(Rv0140, Rv0145, atsA, Rv2466c, fadD26, ilvC, Rv3067,* and *kstD*) that were featured in both lists. Such an insignificant coincidence can be attributed to the fact that a) the infection of cultured macrophages is a relatively simplified model as compared to the complex relationships between the host cells and the pathogen during the infectious process in the host organism; b) the data obtained by a microarray analysis of expression can significantly differ from those obtained by massively parallel sequencing. In particular, Ward *et al.* [20] reported a divergence of the results obtained using these two methods.

Nevertheless, the results obtained by us and Homolka *et al*. [19] are not widely divergent. A functional analysis of the genes with increased expression has demonstrated that they are associated with intracellular stress factors, such as hypoxia and reactive oxygen and nitrogen species, cell wall remodeling, and fatty acid metabolism. The genes associated with iron deficiency (genes in the *mtbA-F* cluster) and those involved in the biosynthesis of valine and isoleucine (*ilvB–ilvN–ilvC*) and phthiocerol dimycocerosates (PDIM) of the cell wall (*ppsA-D*) can be mentioned as an example.

Products of the CUG family genes as potential therapeutic targets

Six genes which could potentially be used as therapeutic targets (or have already been proposed as such) were selected after a review of the available literature and databases. The products of these genes (*hisB, aspC, PPE50, Rv1026, ilvC*, and *Rv1186c*) have been mentioned as attractive therapeutic targets, since the disturbances in their functional activity has the maximum destabilizing effect on the metabolism of *M. tuberculosis*. Thus, for example, aspartate aminotransferase AspC encoded by the *aspC (Rv0337c)* gene was identified in the metabolic network of a mycobacterial cell as an enzyme whose inactivation affects a large number of other *M. tuberculosis* proteins, thus efficiently disintegrating a large number of biochemical cycles [21]. The protein products of two other genes, *Rv1186c* and PPE50 (*Rv3135)*, have been included in the list of the most promising therapeutic targets, which was compiled based on data on the expression, participation in various metabolic pathways, and structural homology with other bacterial and human proteins [22]. The *Rv1026, hisB (Rv1601)*, and *Rv3001c* genes are believed [23] to encode products suitable for designing specific inhibitors. The protein encoded by the *Rv1601* (*hisB*) gene is independently considered to be a promising therapeutic target [24]. It should be mentioned that the expression level of the *Rv1026* gene encoding pyrophosphatase is increased in the macrophages and lungs of infected mice [25, 26], as well as under conditions of inhibited translation in mycobacteria [27]. It has recently been demonstrated that polyphosphate deficiency due to the hydrolytic activity of Rv1026 can change the fatty acid content in the cell wall of *M. smegmatis*, thus affecting sliding motility and biofilm formation [28]. Hence, it can be expected that the products of the remaining CUG genes could be used to design therapeuticals or for diagnosing tuberculosis.

CONCLUSIONS

Infectious diseases caused by intracellular pathogenic bacteria pose a serious medical problem. The progression of the infection depends not only on host defense mechanisms, but also on the specific expression of bacterial genes. Changes in the gene expression in response to various reactions of the host immune system are necessary for the survival and reproduction of pathogenic bacteria. The investigation of the changes in the transcription profile of *M. tuberculosis* in response to various stimuli and external factors allows one to describe the adaptative mechanisms required for an efficient infection of a host organism by a bacterium.

The investigation of the transcription profiles of *M. tuberculosis* under various conditions made it possible to reveal the gene set (CUG) whose expression increases with infection progression and is independent of the genetic features of the host organism. The expression of genes from this core set can be regarded as the universal response of mycobacteria to various environmental stress factors. Further accumulation and analysis of *M. tuberculosis* gene expression data will considerably simplify the development of efficient approaches to the diagnosis and treatment of tuberculosis. ●

This work was supported by the Ministry of Education and Science of the Russian Federation (Project № 8308); the Russian Foundation for Basic Research (grant № 11-04-01325); the Program of the state support of the leading scientific schools in Russia (project NSh-1674.2012.4); and the RAS Presidium Program "Molecular and Cell Biology".

REFERENCES

1. Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E., 3rd., et al. // Nature. 1998. V. 393. № 6685. P. 537–544.
2. Wilson M., DeRisi J., Kristensen H.H., Imboden P., Rane S., Brown P.O., Schoolnik G.K. // Proc. Natl. Acad. Sci. USA. 1999. V. 96. № 22. P. 12833–12838.
3. Butcher P.D. // Tuberculosis (Edinb.). 2004. V. 84. № 3–4. P. 131–137.
4. Kendall S.L., Rison S.C., Movahedzadeh F., Frita R., Stoker N.G. // Trends Microbiol. 2004. V. 12. № 12. P. 537–544.
5. Skvortsov T.A., Azhikina T.L. // Russian Journal of Bioorganic Chemistry. 2010. V. 36. № 5. P. 550–559.
6. Skvortsov T.A., Azhikina T.L. // Russian Journal of Bioorganic Chemistry. 2012. V. 38. № 4. P. 391–405.
7. Ignatov D.V., Skvortsov T.A., Majorov K.B., Apt A.S., Azhikina T.L. // Acta Naturae. 2010. V. 2. № 3. P. 78–84.
8. Azhikina T.L., Skvortsov T.A., Radaeva T.V., Mardanov A.V., Ravin N.V., Apt A.S., Sverdlov E.D. // Biotechniques. 2010. V. 48. № 2. P. 139–144.
9. Audic S., Claverie J.M. // Genome Res. 1997. V. 7. № 10. P. 986–995.

10. Kondratieva E., Logunova N., Majorov K., Averbakh M., Apt A. // PLoS One. 2010. V. 5. № 5. P. e10515.
11. Arnvig K., Young D. // RNA Biology. 2012. V. 9. № 4. P. 427–436.
12. Arnvig K.B., Comas I., Thomson N.R., Houghton J., Boshoff H.I., Croucher N.J., Rose G., Perkins T.T., Parkhill J., Dougan G., et al. // PLoS Pathog. 2011. V. 7. № 11. P. e1002342.
13. Sassetti C.M., Boyd D.H., Rubin E.J. // Mol. Microbiol. 2003. V. 48. № 1. P. 77–84.
14. Shi L., Sohaskey C.D., Kana B.D., Dawes S., North R.J., Mizrahi V., Gennaro M.L. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. № 43. P. 15629–15634.
15. Stokes R.W., Waddell S.J. // Future Microbiol. 2009. V. 4. № 10. P. 1317–1335.
16. Waddell S.J. // Drug Discov. Today: Dis. Mech. 2010. V. 7. № 1. P. e67–e73.
17. Karboul A., Mazza A., Gey van Pittius N.C., Ho J.L., Brousseau R., Mardassi H. // J. Bacteriol. 2008. V. 190. № 23. P. 7838–7846.
18. Hinchey J., Jeon B.Y., Alley H., Chen B., Goldberg M., Derrick S., Morris S., Jacobs W.R., Jr., Porcelli S.A., Lee S. // PLoS One. 2011. V. 6. № 1. P. e15857.
19. Homolka S., Niemann S., Russell D.G., Rohde K.H. // PLoS Pathog. 2010. V. 6. № 7. P. e1000988.
20. Ward S.K., Abomoelak B., Marcus S., Talaat A.M. // Front. Microbiol. 2010. V. 1. P. 121.
21. Raman K., Vashisht R., Chandra N. // Mol. BioSystems. 2009. V. 5. № 12. P. 1740–1751.
22. Raman K., Yeturu K., Chandra N. // BMC Systems Biol. 2008. V. 2. № 1. P. 109.
23. Kalapanulak S. High quality genome-scale metabolic network reconstruction of mycobacterium tuberculosis and comparison with human metabolic network: application for drug targets identification. Edinburgh: Univ. of Edinburgh, 2009.
24. Nisa S.Y. ParA: a novel target for anti-tubercular drug discovery. Wellington: Victoria Univ. of Wellington, 2010.
25. Srivastava V., Jain A., Srivastava B.S., Srivastava R. // Tuberculosis. 2008. V. 88. № 3. P. 171–177.
26. Srivastava V., Rouanet C., Srivastava R., Ramalingam B., Locht C., Srivastava B.S. // Microbiology. 2007. V. 153. № 3. P. 659–666.
27. Boshoff H.I.M., Myers T.G., Copp B.R., McNeil M.R., Wilson M.A., Barry C.E. // J. Biol. Chem. 2004. V. 279. № 38. P. 40174–40184.
28. Shi T., Fu T., Xie J. // Curr. Microbiology. 2011. V. 63. № 5. P. 470–476.

Peculiarities of the Regulation of Gene Expression in the Ecl18kI Restriction–Modification System

O. Yu. Burenina[1], E. A. Fedotova[1], A. Yu. Ryazanova[2], A. S. Protsenko[3], M. V. Zakharova[3], A. S. Karyagina[2,4,5], A. S. Solonin[3], T. S. Oretskaya[1,2], E. A. Kubareva[2*]

[1]Chemistry Department, Lomonosov Moscow State University, Leninskie Gory, 1, bld. 3, Moscow, Russia, 119991

[2]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Leninskie Gory, 1, bld. 40, Moscow, Russia, 119991

[3]Skryabin Institute of Biochemistry and Physiology of Microorganisms, pr. Nauki, 5, Pushchino, Moscow Region, Russia, 142290

[4]Gamaleya Research Institute of Epidemiology and Microbiology, Gamaleya Str., 18, Moscow, Russia, 123098

[5]Institute of Agricultural Biotechnology, Timiryazevskaya Str. 42, Moscow, Russia, 127550

*E-mail: kubareva@belozersky.msu.ru

Received 29.11.2012

ABSTRACT **Transcription regulation in bacterial restriction–modification (R–M) systems is an important process, which provides coordinated expression levels of tandem enzymes, DNA methyltransferase (MTase) and restriction endonuclease (RE) protecting cells against penetration of alien DNA. The present study focuses on (cytosine-5)-DNA methyltransferase Ecl18kI (M.Ecl18kI), which is almost identical to DNA methyltransferase SsoII (M.SsoII) in terms of its structure and properties. Each of these enzymes inhibits expression of the intrinsic gene and activates expression of the corresponding RE gene via binding to the regulatory site in the promoter region of these genes. In the present work, complex formation of M.Ecl18kI and RNA polymerase from *Escherichia coli* with the promoter regions of the MTase and RE genes is studied. The mechanism of regulation of gene expression in the Ecl18kI R–M system is thoroughly investigated. M.Ecl18kI and RNA polymerase are shown to compete for binding to the promoter region. However, no direct contacts between M.Ecl18kI and RNA polymerase are detected. The properties of M.Ecl18kI and M.SsoII mutants are studied. Amino acid substitutions in the N-terminal region of M.Ecl18kI, which performs the regulatory function, are shown to influence not only M.Ecl18kI capability to interact with the regulatory site and to act as a transcription factor, but also its ability to bind and methylate the substrate DNA. The loss of methylation activity does not prevent MTase from performing its regulatory function and even increases its affinity to the regulatory site. However, the presence of the domain responsible for methylation in the M.Ecl18kI molecule is necessary for M.Ecl18kI to perform its regulatory function.**

KEYWORDS **restriction–modification systems; (cytosine-5)-DNA methyltransferase; DNA–protein interactions; transcriptional regulation.**

ABBREVIATIONS **MTase – DNA methyltransferase; PAGE – polyacrylamide gel electrophoresis; RNAP – RNA polymerase; R–M system – restriction–modification system; RE – restriction endonuclease; AdoMet – *S*-adenosyl-*L*-methionine; M.Ecl18kI – DNA methyltransferase Ecl18kI; M.SsoII – DNA methyltransferase SsoII; R.Ecl18kI – restriction endonuclease Ecl18kI. Prefix "d" for designating deoxyribonucleosides, oligodeoxyribonucleotides, and DNA duplexes is omitted.**

INTRODUCTION

Restriction–modification (R–M) systems are abundant in bacterial cells; they contain genes that encode restriction endonucleases (RE) and DNA methyltransferases (MTases). RE hydrolyzes a certain sequence in a double-stranded DNA (dsDNA), while MTase methylates the same sequence at a strictly determined position, thus preventing its cleavage by RE. The R–M functions as a primitive immune system that protects a host bacterium from penetration by alien DNA: RE hydrolyses the intruding DNA that is not methylated by the corresponding MTase [1]. The activity levels of

the RE and the MTase in the cell are to be strictly coordinated. An extremely low level of the MTase gene expression as compared with the RE gene may cause cell death via hydrolysis of cellular DNA, whereas its excessively high level cannot protect the cell against penetration by an alien DNA.

Although there is no doubt that gene expression in the R–M systems is regulated, the mechanisms underlying this process are poorly studied. It has been demonstrated by recent research that coordinated gene expression in R–M systems is presumably determined by regulation at the transcriptional level. Three major types of regulation can be distinguished: via C-proteins, via methylation of the promoter region of the R–M system by the MTase, and via the interaction between the MTase and the regulatory sites in DNA, which differ from the methylation site [2]. This study focuses on the latter type of regulation, which is typical of (cytosine-5)-DNA MTases (enzymes that methylate the cytosine residue at position 5) belonging to the type II R–M systems. Over 300 (cytosine-5)-DNA MTases have been recently characterized; however, the existence of the regulatory function has been experimentally confirmed only for six of them (M.MspI, M.EcoRII, M.ScrFIA, M1.LlaJI, M.SsoII, and M.Ecl18kI) [2].

The type II R–M system SsoII has been most thoroughly studied. The genes of this system are located in natural plasmid P4 (4250 bp) from the *Shigella sonnei* 47 strain; they are divergently oriented; the intergenic region consists of 109 bp [3]. The other four SsoII-like R–M systems isolated from various bacterial strains have been described; their MTases are either identical to M.SsoII in terms of the amino acid sequence (M.Kpn2kI from *Klebsiella pneumoniae* 2k) or differ insignificantly. Thus, MTases Ecl18kI from *Enterobacter cloacae* 18k and StyD4I from *Salmonella typhi* D4 carry Met instead of Ile at position 56, while MTase SenPI from *Salmonella enteritidis* P1 contains Ile56 and, in addition, Gly instead of Glu at position 11 [4–7]. The nucleotide sequences of the corresponding genes share 99–100% identity; those of the intergenic regions are absolutely identical. Hence, the data on the functioning of the enzymes from one of these systems can be extrapolated to the other systems as well.

All SsoII-like R–M systems recognize sequence 5'-CCNGG-3'/3'-GGNCC-5' (N = A, G, C or T) in dsDNA and methylate the inner C residue in this sequence in the presence of the cofactor *S*-adenosyl-*L*-methionine (AdoMet) forming 5-methyl-2'-deoxycytidine [4]. The promoter elements of the genes encoding the RE and MTase of the SsoII-like R–M systems have been determined using the Ecl18kI system as an example; the *in vitro* transcriptional regulation of these genes by M.Ecl18kI has been also shown. In order to regulate transcription, M.Ecl18kI binds to the so-called regulatory site, the 15-mer inverted repeat 5'-GGACAA**A**TTGTCCT-3'/3'-CCTGTT**T**AACAGGA-5', which is localized inside the promoter region of the genes of the Ecl18kI R–M system [9]. The nucleotides that participate in the formation of specific DNA–protein contacts with the MTase are located inside the regulatory site (*Fig. 1*) [10, 11]. All SsoII-like MTases are two-domain proteins whose N-terminal region (residues 1–71) provides transcriptional regulation, while the region of 72–379 residues is responsible for DNA methylation. The N-terminal region of M.SsoII has been shown to have a strongly pronounced secondary structure [12] in which the "helix–turn–helix" (HTH) motif is predicted with a high probability. Two M.SsoII molecules (which are monomeric in the apo-form) interact with the regulatory site [12]. The data [13] regarding the putative contacts in the complex between the M.SsoII N-terminal region and the regulatory site are summarized in *Fig. 1*.

In order to refine the mechanism of gene transcription regulation in the SsoII-like R–M systems, the efficiency of complex formation of M.Ecl18kI and *E. coli* RNA polymerase (RNAP) with DNA fragments containing the regulatory elements of the genes of the Ecl18kI R–M system is assessed in this study. All known SsoII-like R–M systems have been isolated from various enterobacterial strains (*E. coli* belonging to them as well); thus, the use of *E. coli* RNAP is reasonable. The role of residues Lys21, Lys31, Lys46, and Lys53 in the M.Ecl18kI N-terminal region for the binding of this protein to the regulatory site, as well as their effect on

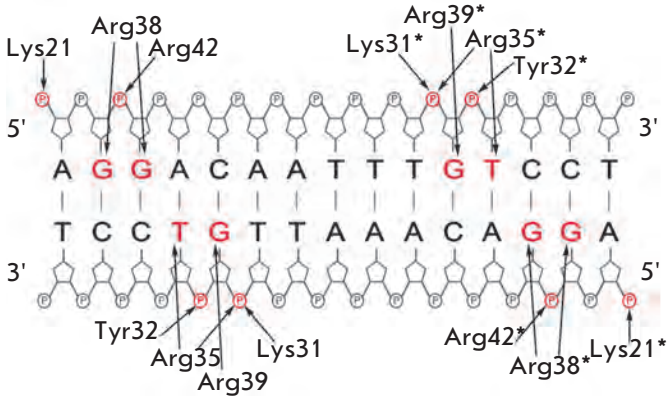


Fig. 1. Scheme of the contacts between the amino acid residues in the N-terminal domain of M.SsoII dimer and the regulatory site in DNA. Heterocyclic bases and phosphate groups which interact with M.SsoII (identified by footprinting technique) are shown in red. Amino acid residues of the second M.SsoII subunit are marked with asterisks

the MTase ability to act as a transcription factor and on the interaction between the enzyme and the methylation site are being studied for the first time.

METHODS

Protein purification

MTase Ecl18kI and its mutant forms were purified by affinity chromatography on Ni-NTA agarose [4]. *E. coli* RNA polymerase was sequentially purified by Ni-NTA-agarose and heparin-sepharose affinity chromatography, followed by DEAE cellulose ion-exchange chromatography [14].

Synthesis of DNA fragments I–III

Fragments **I–III** were synthesized by polymerase chain reaction (PCR) on an Eppendorf Mastercycler personal thermal cycler (Eppendorf North America, USA). The DNA fragment **I** was obtained using the primers 5'-TTGAGTCATATGAAGTCTTTCTC-3' and 5'-AGCAAATGGCGTAATAAAATGC-3'; the DNA fragment **II**, using 5'-TCATGCATGTCTACCAGAA-3' and 5'-CATAAAAAATAACCTTTTATACT-3'; the DNA fragment **III**, using 5'-TTGAGTCATATGAAGTC-3' and 5'-CCAACACTTAATTCTGG-3'. Hybridization (annealing) temperature for each pair of primers was 62, 54, and 46°C, respectively. The PCR cycle (90°C – 60 s, primers annealing – 60 s, 72°C – 40 s) was repeated 25 times. After the PCR, DNA was precipitated with ethanol (2.5 vol.) in the presence of 1 M NaCl. The target DNA was isolated from agarose gel using microcentrifuge tubes Spin-X Centrifuge Tube Filters (Costar, USA).

Equilibrium binding of the proteins to the DNA ligands

The 5'-ends of the oligonucleotides were radioactively labeled using T4 polynucleotide kinase (10 units, Fermentas, Lithuania) and [γ-^{32}P]ATP. The complex formation between the M.Ecl18kI and DNA fragments **I–II**, as well as between the RNAP and DNA fragments **I–III**, was conducted in 10 μl of the binding buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 5 mM β-mercaptoethanol) in the presence of heparin (equimolar amount to the protein) for 40 min at 37°C. In the case of M.Ecl18kI, the reaction mixture contained 1 mM AdoMet. The DNA–protein complex and the unbound DNA duplex were separated by gel electrophoresis in 1% agarose gel. After the electrophoresis, the agarose gels were dried on a supporting plate at 90°C in a hot air flow. The dissociation constants (K_d) of the DNA–protein complexes were determined by the Scatchard technique [15]. The concentrations of M.Ecl18kI and RNAP were 60 and 30 nM, respectively. The concentrations of the DNA duplex **II** were varied within a range from 5 to 120 nM. The complex formation of the mutants M.Ecl18kI(K46A), M.Ecl18kI(K53A), and M.Ecl18kI(K21A) with the DNA fragments **IV** and **V** was conducted in 20 μl of the binding buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl, 5 mM DTT, 50 ng/μl poly(dI·dC)) for 20 min at 37°C. The concentrations of the DNA duplexes **IV** and **V** were varied within a range from 20 to 100 nM. The concentrations of M.Ecl18kI(K46A), M.Ecl18kI(K53A), and M.Ecl18kI(K21A) were equal to 560, 400, and 400 nM, respectively, when binding to the DNA fragment **IV** and were equal to 200, 1600, and 5600 nM, respectively, when binding to the DNA fragment **V**.

Determination of the initial rate of the substrate DNA methylation

The initial rate of the substrate DNA methylation by MTases Ecl18kI, SsoII, and their mutant forms was determined as previously described [9], on the basis of the degree of the duplex **V** "protection" against hydrolysis by RE Ecl18kI (R.Ecl18kI). For this purpose, 350 nM of the radiolabeled DNA duplex **V** was incubated with MTase in the binding buffer containing 1 mM AdoMet for 0.5–60 min at 37°C. The reaction mixture was then kept at 65°C for 10 min to inactivate the enzyme, and cooled to 25°C. Next, $MgCl_2$ (up to 10 mM) and R.Ecl18kI (up to 240 nM) were added and the reaction mixture was incubated at 37°C for 1 h. The initial active concentrations of the MTases were identical (14 nM). The degree of hydrolysis of the unmethylated DNA duplex **V** by R.Ecl18kI was taken as 100%. The degree of methylation of the DNA duplex **V** by the MTases was calculated with respect to this value, and the kinetic curves were plotted. The initial methylation rate (v_0) of the DNA duplex **V** by MTase was calculated as an angular coefficient (slope ratio) of the initial linear region on the kinetic curve.

In vitro transcription

The purified DNA fragment **I** (0.25 μg) was incubated with RNAP (3 pmol) in the transcription buffer (40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 10 mM DTT, 10 mM NaCl, 2 mM spermidine) in 8 μl for 10 min at 37°C. Next, the mixture was supplemented with 2 μl of an aqueous heparin solution (0.25 μg/μl) and incubated at 37°C for an additional 10 min. 10 μl of a ribonucleoside triphosphates mixture (UTP (12 μM), ATP, GTP, and CTP (500 μM each)) containing 0.5 μCi [α-^{32}P]UTP and 24 units of the RNase inhibitor RiboLock (Fermentas, Lithuania) were then added. After incubation at 37°C for 1 h, the reaction mixture was mixed with 10 μl of RNA Loading Dye (Fermentas, Lithuania) and loaded onto a polyacrylamide gel.

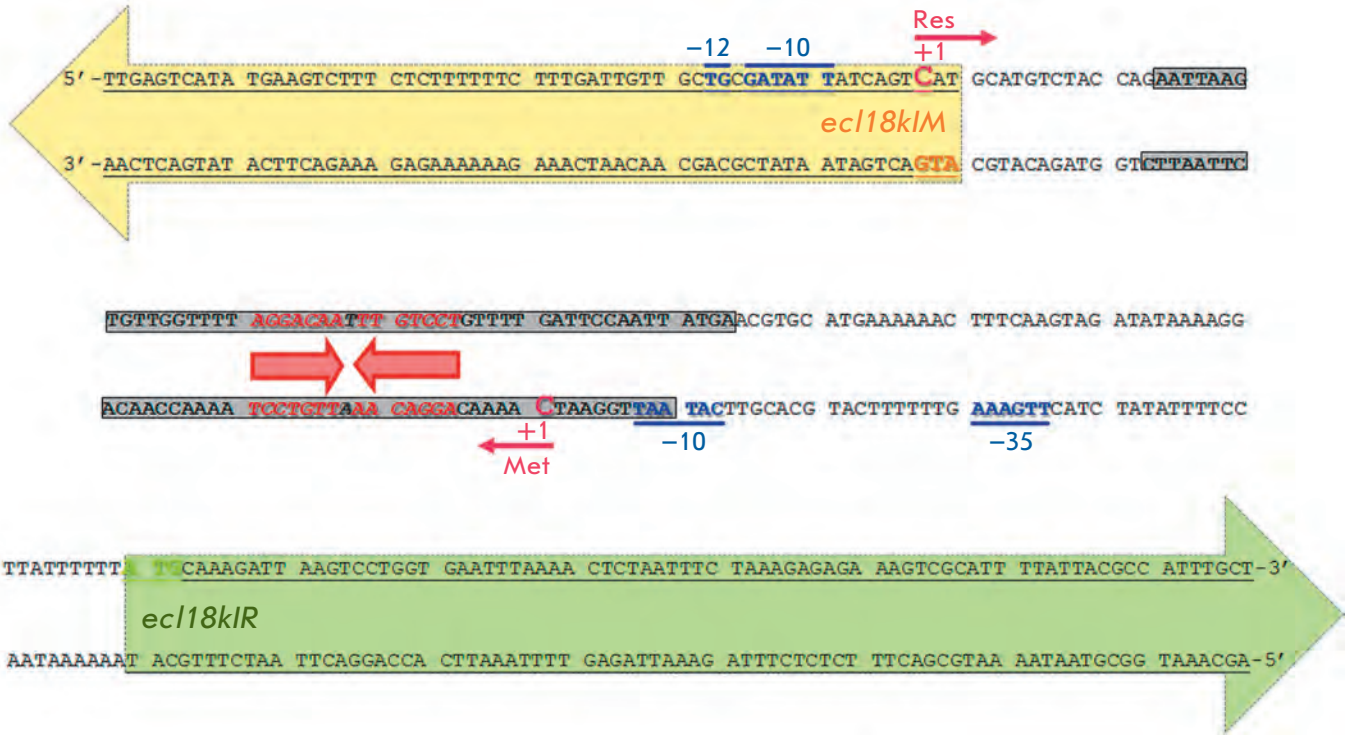


Fig. 2. Genetic arrangement of the Ecl18kI restriction–modification system promoter region, the DNA fragment **I**. The directions of the MTase and the RE genes are shown with yellow and green arrows, respectively. The start codons are also marked with the corresponding colors (yellow or green). The region protected by MTase against DNase hydrolysis is shown in grey. The regulatory site of M.Ecl8kI (inverted repeat) is marked with red letters and red arrows. The transcription initiation points and the promoter elements are shown in pink and blue, respectively

Characterization of the regulatory activity of the methyltransferases

The regulatory activity of the mutant forms of M.Ecl18kI and M.SsoII was assessed via *in vitro* transcription from the DNA fragment **I** in the presence of the corresponding proteins. The wild-type M.Ecl18kI or M.SsoII were used in the control experiments. The reaction mixtures were analyzed by 5% polyacrylamide gel electrophoresis (PAGE; the gel contained 7 M urea) at a field intensity of 5 V/cm in TBE buffer. Only the resulting RNA transcripts contained the radiolabel. In the presence of the SsoII-like MTases capable of acting as regulatory proteins, the following changes were observed: an increase in the radioactivity of the region corresponding to the RNA transcript from the RE gene promoter and a decrease in the radioactivity of the region corresponding to the RNA transcript from the MTase gene promoter. The fraction (%) of the RE gene transcript in the total radioactivity of the resulting transcripts (taken as 100%) at various MTase concentrations was determined. Identical active concentrations of the MTases were used to ensure a correct comparison of the yields of the transcription products in the reaction. They were obtained from the Scatchard plots used to determine the K_d values for the complexes between the proteins and the duplex **IV** containing the regulatory site [15]. The fraction of the transcript from the RE gene promoter was plotted as a function of the MTase active concentration. The relative yield of this transcript per unit of the MTase active concentration was then determined. For this purpose, the ratio between the angular coefficients (slope ratios) of the initial linear region on the curves of the mutant MTase and the wild-type M.Ecl18kI (or M.SsoII) was calculated.

RESULTS AND DISCUSSION

Complex formation of RNA polymerase and M.Ecl18kI with the DNA fragments containing the intergenic region of the Ecl18kI R–M system

Figure 2 shows the genetic arrangement of the Ecl18kI R–M system (based on the data [8, 11]) by the example of the 247-bp DNA fragment **I**. The MTase gene pro-

moter is localized directly before the regulatory site and partially overlaps the region which is protected by M.Ecl18kI from DNAse I cleavage. We had assumed that the mechanism of negative regulation of the MTase gene expression may consist in physical blocking of the RNAP access to the MTase gene promoter as M.Ecl18kI binds to the regulatory site.

To verify this hypothesis, complex formation of both proteins with the 116-bp DNA fragment **II** containing the intergenic sequence of the Ecl18kI R–M system (the regulatory site, the transcription initiation point, and the promoter elements of the MTase gene *ecl18kIM*) but lacking the promoter elements of the RE gene *ecl18kIR* was studied (*Fig. 3A*). After RNAP was added to the MTase–DNA mixture, no other complexes but MTase–DNA and RNAP–DNA emerged in the reaction mixture. This fact eliminates the possibility of direct contact between M.Ecl18kI and RNAP. Moreover, the 5-fold excess of M.Ecl18kI (with respect to RNAP) resulted in virtually complete disappearance of the RNAP–DNA complex. Therefore, MTase binding to the regulatory site does impede the interaction between RNAP and the promoter region of the SsoII R–M system genes (*Fig. 3B*).

The efficiency of RNAP and M.Ecl18kI binding to the MTase promoter and to the regulatory site was assessed by determining the K_d values of the DNA–protein complexes. K_d = 12 ± 1 nM for the M.Ecl18kI complex with the DNA fragment **II**, while K_d = 25 ± 1 nM for the RNAP complex with the same fragment. Thus, the control over the MTase expression level can be attributed to the competition between RNAP and M.Ecl18kI for the binding site. The insignificant (2-fold) difference in the MTase and RNAP affinity to this DNA region allows preventing premature inhibition of M.Ecl18kI synthesis, i.e. controlling the expression level of the MTase gene more accurately. Thus, the level of M.Ecl18kI synthesis does not fall below the minimal value that ensures maintenance of the specific methylation of cellular DNA.

Since the MTase is localized near the transcription initiation point of the RE gene (*Fig. 2*), it seems quite possible that M.Ecl18kI has a negative effect on the *ecl18kIR* gene transcription. However, an opposite effect is observed. We had assumed that RNAP and M.Ecl18kI could be bound simultaneously to the same DNA fragment, RNAP interacting with the RE gene promoter, while the MTase interacts with its regulatory site. This assumption is verified experimentally (*Fig. 3C,D*): sequential addition of RNAP and M.Ecl18kI to the 247-bp DNA fragment **I** leads to a ternary complex formation (supposedly RNAP–M.Ecl18kI–DNA) which has a lower electrophoretic mobility as compared with the RNAP–DNA and M.Ecl18kI–DNA complexes. Since two M.SsoII molecules bind to a single regulatory site [12], it is highly probable that each complex (RNAP–M.Ecl18kI–DNA and M.Ecl18kI–DNA) contains two M.Ecl18kI molecules.

Complex formation between RNAP and the two different promoters shows that the degree of RNAP binding to the DNA fragment **III** (*Fig. 3E,F*), which contains the transcription initiation point and the promoter regions of the *ecl18kIR* gene only, is 4-fold lower than the degree of RNAP binding to the DNA fragment **II**, which contains the transcription initiation point and promoter regions of the *ecl18kIM* gene only. Thus, the *ecl18kIM* gene promoter is stronger than the *ecl18kIR* gene promoter, and transcription primarily occurs from the MTase gene promoter in the absence of M.Ecl18kI. This phenomenon can also be stipulated by the "sitting duck" mechanism of transcriptional interference [16] when the rates of the open RNAP complex transition into the elongation form for two closely spaced promoters differ considerably and the activity of the weaker promoter is suppressed due to the intensive transcription of the stronger one.

Analysis of the ability of the M.Ecl18kI N-terminal region to regulate gene transcription in the restriction–modification system *in vitro*

The experiments with deletion mutants have demonstrated that the M.SsoII ability to act as a transcription factor can be attributed to the N-terminal region of this protein, which consists of 71 residues [3]. The amino acid sequences of the N-terminal regions of M.Ecl18kI and M.SsoII significantly resemble C-proteins. When comparing the M.SsoII regulatory site with the idealized sequence of C-boxes (5'-GACT...AGTC-3') [17], 6 out of 8 nucleotides coincide. Considering the significant variability among the sequences of the C-boxes, the regulatory site recognized by M.Ecl18kI can also be classified as a C-box. The deletion mutant Δ(72–379) M.Ecl18kI, which is the N-terminal region of M.Ecl18kI, retains its strongly pronounced secondary structure and is capable of specific binding to the DNA containing the regulatory site; however, the efficiency of such binding is an order of magnitude lower than that of the full-length protein [12].

The effect of Δ(72–379)M.Ecl18kI on the *in vitro* transcription of the *ecl18kIR* and *ecl18kIM* genes has been studied. The full-length M.Ecl18kI was used in the control experiment (*Fig. 4*). Transcription from the 247-bp DNA fragment **I** resulted in two products corresponding to the transcripts from the RE gene promoter (~190 nucleotides) and from the MTase gene promoter (~110 nucleotides). When the reaction mixture was titrated with increasing amounts of M.Ecl18kI, the fraction of the MTase gene transcript decreased considerably,

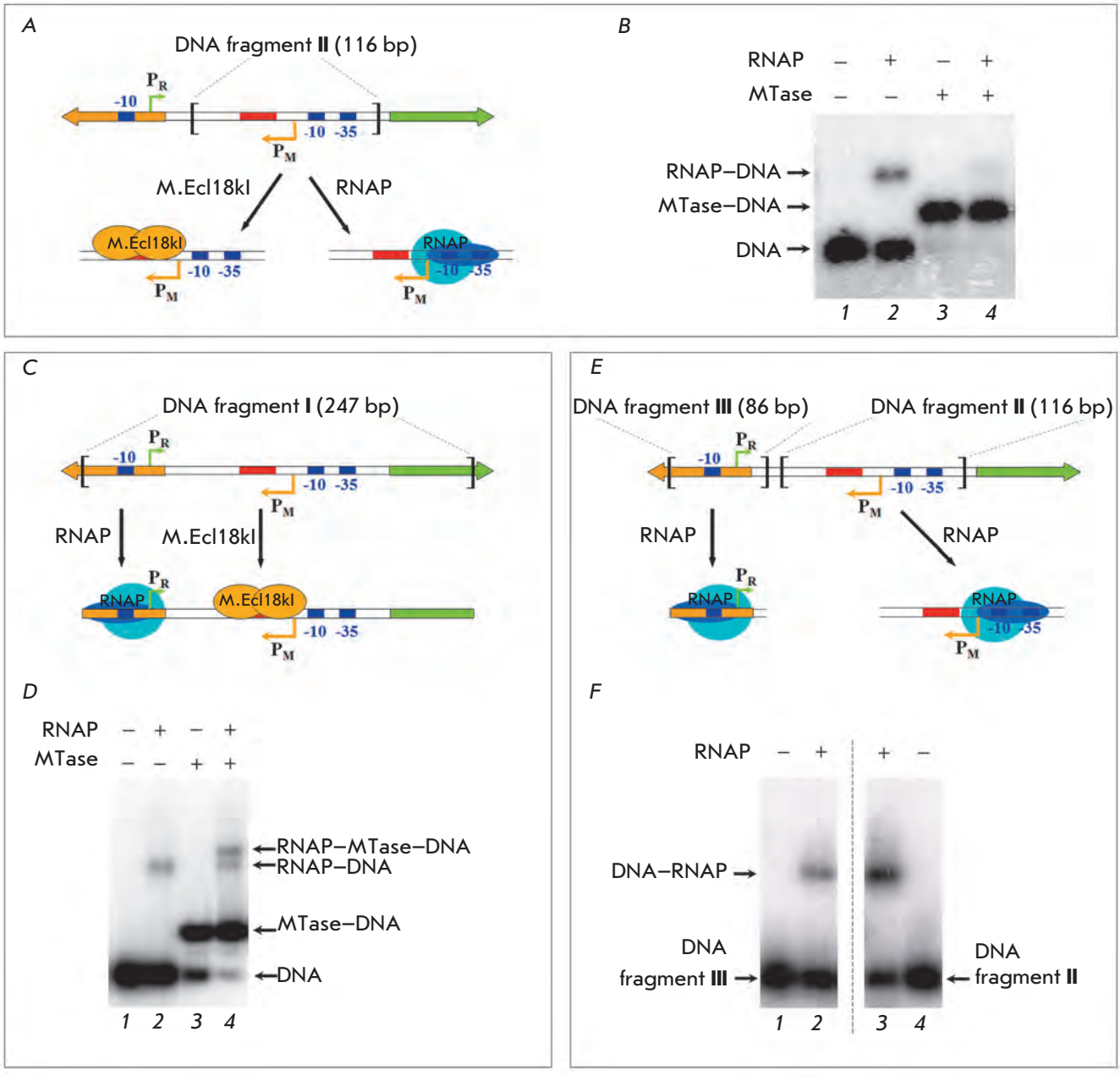


Fig. 3. Complex formation of M.Ecl18kI and RNA polymerase with the DNA fragments that contain different elements of the intergenic region of the Ecl18kI restriction–modification system. **A**, **C**, **E** – schematic representations of the DNA–protein complexes formation. The directions of the MTase and the RE genes are shown with yellow and green arrows, respectively. P_R, P_M – transcription initiation points of the RE and MTase genes, respectively (also marked with thin arrows). The promoter elements are shown in blue, the regulatory site is shown in red. **B, D** – complex formation of RNA polymerase (30 nM) with the DNA fragments **II** or **I**, respectively (15 nM) in the presence or absence of M.Ecl18kI excess (150 nM) under specific binding conditions (with 300 nM heparin). **F** – complex formation between RNA polymerase (190 nM) and the DNA fragments **III** or **II** (30 nM). Radioautographs of 1% agarose gels

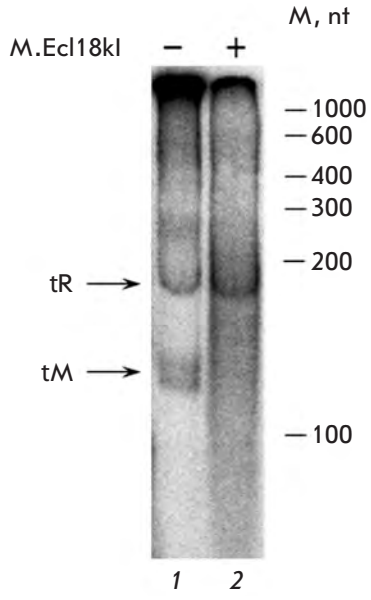


Fig. 4. Analysis of the transcripts from the DNA fragment I by electrophoresis in 5% polyacrylamide gel under denaturing conditions. Radioautography: 1 – the transcripts in the absence of M.Ecl18kI, 2 – the transcripts in the presence of M.Ecl18kI 4-fold excess (considering active concentrations of the enzymes). The positions of RNA length markers are shown on the right side.
tR – the transcript from the *ecl18kIR* gene promoter;
tM – the transcript from the *ecl18kIM* gene promoter;
nt – nucleotides

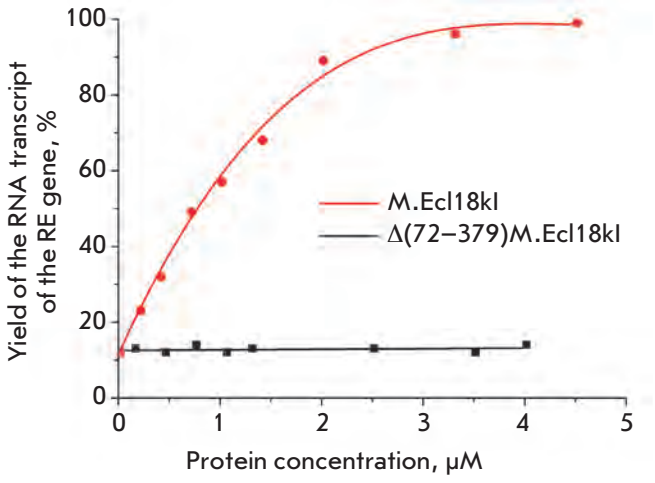


Fig. 5. Dependence of the transcript yield from the *ecl-18kIR* gene promoter on the concentration of M.Ecl18kI or its deletion mutant Δ(72–379)M.Ecl18kI

while that of the RE gene transcript increased (*Figs. 4,5*). Meanwhile, the addition of Δ(72–379)M.Ecl18kI to the reaction mixture caused no changes in the ratio between the yields of the two transcripts; i.e., this deletion mutant could not function as a transcription factor (*Fig. 5*). This fact is probably due to the low affinity of Δ(72–379)M.Ecl18kI to the DNA carrying the regulatory site [12]: such a protein cannot efficiently compete with RNAP for binding to the promoter region. It is also possible that the deletion mutant covers a considerably smaller DNA fragment as compared with the full-length M.Ecl18kI and therefore is not a steric impediment for RNAP. Thus, the region responsible for methylation is necessary to maintain the regulatory function of M.Ecl18kI. This result agrees with the recently proposed structural model of the complex between the SsoII-like MTases and the regulatory site within the intergenic region of the R–M system: the N-terminal regions of both MTase molecules specifically interact with the regulatory site, while the regions responsible for methylation are nonspecifically bound to the DNA flanking the regulatory site [18].

Model of gene transcription regulation in the Ecl18kI restriction–modification system

After the R–M system penetrates into a cell, the MTase is actively synthesized from the stronger promoter, which is required to protect cellular DNA against the hydrolysis by RE. A certain amount of MTase, which can efficiently protect the cell against bacteriophage infection, is produced with time. Then, two MTase molecules bind to the regulatory site and block the RNAP access to the promoter of the MTase gene (*Fig. 6*). No complex formation between MTase and RNAP occurs in this case; i.e., the mechanism of transcription suppression of the MTase gene is based exclusively on the competition between the MTase and RNAP for binding to the intergenic region of the Ecl18kI R–M system. The close K_d values attest to the fact that even small changes in the MTase concentration are expected to affect the efficiency of the MTase gene transcription.

The interaction between the M.SsoII regions responsible for methylation with the DNA flanking the regulatory site described in [18] seems to confer additional strength to the DNA–protein complex. This circumstance allows SsoII-like MTase to successfully compete with RNAP for binding to the promoter region, resulting in the suppression of the MTase gene transcription and stabilization of the MTase concentration in the cell. It can be assumed that binding of the enzyme region responsible for methylation to the DNA flanking the regulatory site is a compensatory mechanism which is required to make the effect on transcription of a MTase dimer bound to the regulatory site as efficient as that

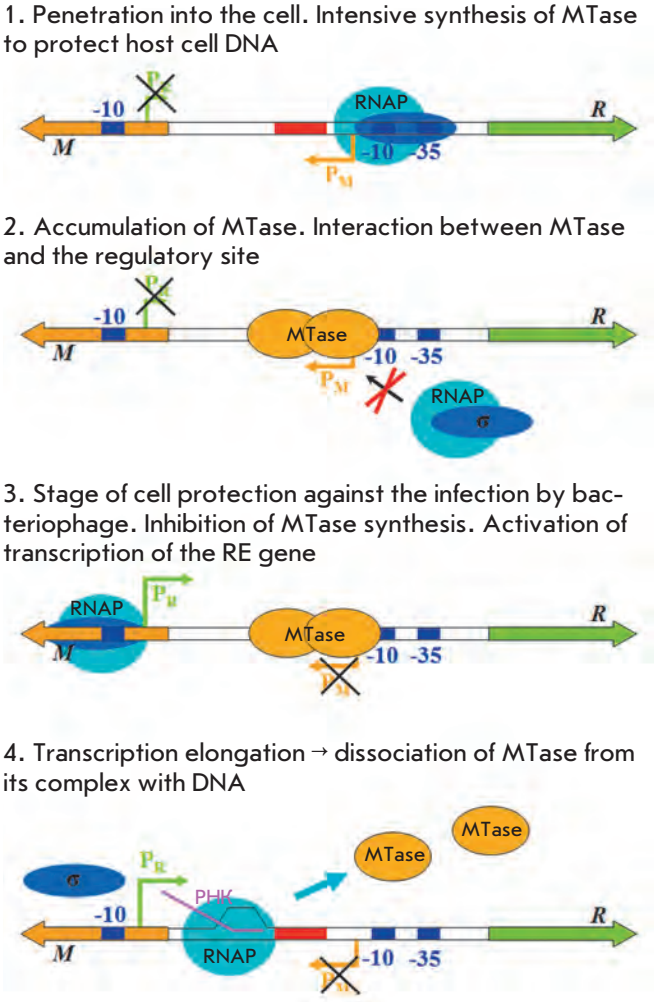


Fig. 6. The supposed mechanism of regulation of gene transcription in SsoII-like restriction–modification systems. The designations are similar to those given in the legends to Figs. 2 and 3

of two C-protein dimers bound to two palindromic sites in DNA. The fact that a deletion mutant, which is the N-terminal region of M.Ecl18kI, does not have this "additional" interaction explains the low stability of its complex with the DNA and its inability to control transcription in the Ecl18kI R–M system.

Binding between M.Ecl18kI and the regulatory site results in indirect activation of the RE gene promoter by preventing RNAP from binding to the MTase gene promoter. During transcription from the RE gene promoter, RNAP runs against the MTase region, which is responsible for methylation and nonspecifically interacts with the DNA region flanking the regulatory site [18]. These nonspecific DNA–protein contacts can be relatively easily destroyed by RNAP, which melts DNA in the elongation complex. It is possible that both MTase subunits are pushed away from the DNA, which can be caused by the reduced affinity of the enzyme to the DNA melted during the elongation process.

Effect of single amino acid substitutions on the regulatory activity of the SsoII-like methyltransferases

The mutant form of M.SsoII containing Cys142 substitution in the region responsible for methylation. Cys142 in the M.Ecl18kI (M.SsoII) molecule plays the key role in catalyzing the methyl group transfer from the reaction cofactor AdoMet to the substrate DNA [19]. Replacement of Cys142 by Ala results in loss of M.SsoII enzymatic activity. The efficiency of the mutant protein binding to the methylation site decreases; however, the mutant has a considerably higher affinity to the regulatory site (*Table*) [9]. The M.SsoII(C142A) ability to regulate *in vitro* transcription of the genes in the Ecl18kI R–M system was tested in this study.

The yields of the transcripts of the *ecl18kIR* gene in the presence of M.SsoII, M.Ecl18kI, or the mutant protein M.SsoII(C142A) are almost identical (*Fig. 7*, *Table*). Hence, loss of the methylation function does not affect MTase's ability to function as a transcription factor.

The mutant forms of M.Ecl18kI containing substitutions in the region responsible for the regulatory function. Based on the model of the complex between the M.SsoII N-terminal region and the regulatory site [13], a hypothesis has been proposed that residues Lys21, Lys31, Arg35, Arg38, Arg39, and Arg42 interact with DNA (*Fig. 1*). We studied the regulatory properties of the M.Ecl18kI mutants, where one of the above-mentioned residues was replaced by Ala (*Table*). The M.Ecl18kI mutants with one of the residues (Arg15, Lys46, or Lys53) replaced by Ala were used as a control. The regulatory activity of all the M.Ecl18kI mutant forms was tested by conducting *in vitro* transcription in the presence of these proteins. Wild-type M.Ecl18kI was used in the control experiment.

It is shown for the first time that the amino acid substitutions in the N-terminal region affect the MTase ability to regulate transcription in the R–M system (*Table*). An interesting and unexpected result of the study is the dynamics of the yield changes of the transcripts from the RE gene promoter, which differed among different M.Ecl18kI mutants at the same active concentrations. The mutants exhibiting high affinity to the regulatory site had been expected to regulate transcription more efficiently, whereas the regulation

Characterization of the DNA-binding, regulatory, and methylating activities of the MTases Ecl18kI, SsoII, and their mutant forms[1]

MTases	Relative yield of the RE gene transcript per unit of active concentration of MTase	K_d of the complex between MTase and the regulatory site, nM[1,2]	K_d of the complex between MTase and the methylation site, nM[1,3]	Relative initial methylation rate[1,3]
Ecl18kI	1.0	224 ± 24	87 ± 12	1
SsoII	1.0	248 ± 33	144 ± 14	1
SsoII(C142A)	1.0	35 ± 3	172 ± 10	–
Ecl18kI(R15A)	0.4	56 ± 13	103 ± 24	< 1
Ecl18kI(K21A)	3.9	48 ± 9	87 ± 3	38
Ecl18kI(K31A)	1.0	198 ± 29	26 ± 3	29
Ecl18kI(R35A)	–	> 4000	140 ± 12	2
Ecl18kI(R38A)	–	> 4000	96 ± 13	11
Ecl18kI(R39A)	0.4	93 ± 14	266 ± 4	22
Ecl18kI(R42A)	2.5	32 ± 2	256 ± 4	< 1
Ecl18kI(K46A)	13.5	250 ± 32	> 4000	–
Ecl18kI(K53A)	1.8	206 ± 7	> 4000	–

[1] Data for M.Ecl18kI, M.SsoII, M.SsoII(C142A), M.Ecl18kI(R15A), M.Ecl18kI(R35A), M.Ecl18kI(R38A), M.Ecl18kI(R39A), and M.Ecl18kI(R42A) have been published earlier [9].
[2] The complex formation was studied using the 31-bp DNA duplex **IV** containing the regulatory site:
5′-TTGGTTTT**AGGACAA*T*TTGTCCT**GTTTTGAT-3′
3′-AACCAAAA**TCCTGTT*A*AACAGGA**CAAAACGA-5′ (DNA duplex **IV**).
[3] The complex formation and methylation activity were studied using the 30-bp DNA duplex **V** containing the methylation site:
5′-GATGCTGCCAA**CCTGG**CTCTAGGTTCATAC-3′
3′-CTACGACGGTT**GGACC**GAGATCGAAGTATG-5′ (DNA duplex **V**).

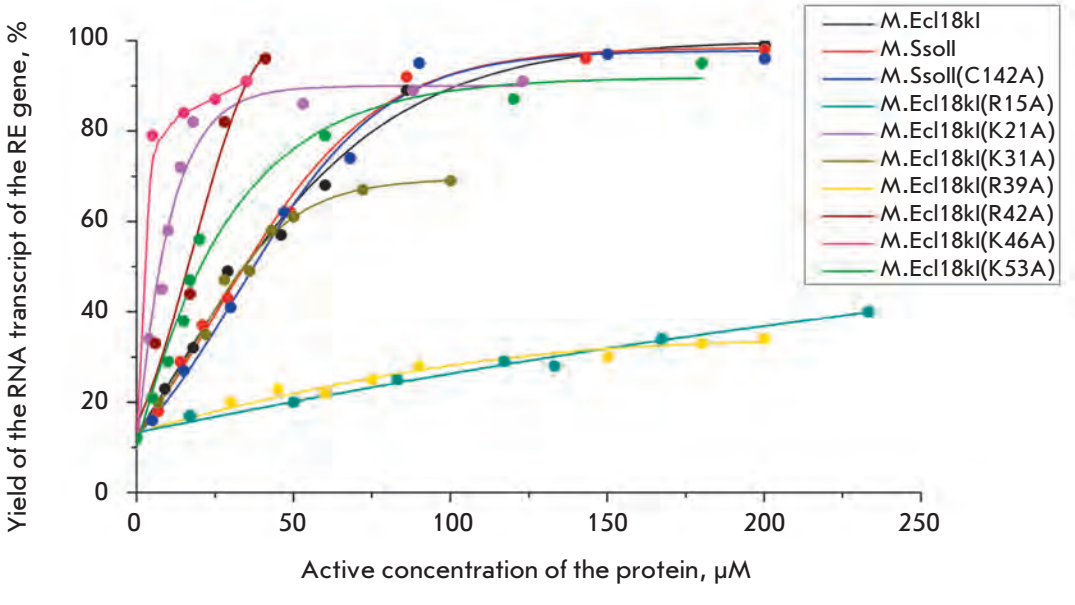


Fig. 7. Dependence of the transcript yield from the *ecl18kIR* gene promoter on the active concentration of M.Ecl18kI, M.SsoII, or the following mutants: M.SsoII(C142A), M.Ecl18kI(R15A), M.Ecl18kI(K21A), M.Ecl18kI(K31A), M.Ecl18kI(R39A), M.Ecl18kI(R42A), M.Ecl18kI(K46A), and M.Ecl18kI(K53A)

of transcription by the MTases exhibiting lower affinity had been expected to weaken. Indeed, in the presence of M.Ecl18kI(R35A) and M.Ecl18kI(R38A), which poorly interact with the regulatory site (*Table*, [9]), the result is identical to that observed in the absence of the protein: the RNA transcript from the MTase gene promoter is predominant in the reaction mixture. Evidently, these mutants cannot regulate gene transcription in the Ecl18kI R–M system.

In the presence of the rest of the M.Ecl18kI mutants, which could bind efficiently to the regulatory site, changes in the amount of the RNA transcript from the RE gene promoter were observed (*Fig. 7, Table*). However, a certain correlation between the MTase affinity to the regulatory site and the transcript yield per unit of the MTase active concentration was detected only for three mutants: M.Ecl18kI(K21A), M.Ecl18kI(K31A), and M.Ecl18kI(R42A) (in addition to the aforementioned M.Ecl18kI(R35A) and M.Ecl18kI(R38A)). In the case of M.Ecl18kI(K21A) and M.Ecl18kI(R42A), the affinity to the regulatory site and the efficiency of transcriptional regulation per unit of the MTase active concentration is higher than that in the case of wild-type M.Ecl18kI. The affinity of M.Ecl18kI(K31A) to the regulatory site and the yield of the RNA transcript per unit of the MTase active concentration is comparable to the same values for the wild-type M.Ecl18kI.

Meanwhile, the mutants M.Ecl18kI(R39A) and M.Ecl18kI(R15A), which are characterized by a higher affinity to the regulatory site as compared with that of wild-type M.Ecl18kI (2.5- and 4-fold, respectively), regulate transcription in the Ecl18kI R–M system less efficiently. The RNA transcript yield per unit of active concentration for M.Ecl18kI(K46A) and M.Ecl18kI(K53A) is 13- and 1.8-fold higher than that for the wild-type M.Ecl18kI, respectively, despite the fact that the dissociation constants of their complexes with the 31-bp duplex **IV** containing the regulatory site are comparable.

The absence of a correlation between the affinity to the regulatory site and the RNA transcripts yield can be attributed to the fact that the K_d value shows the thermodynamic stability of the MTase–DNA complex, while the relative yield of the transcription product per unit of the MTase active concentration characterizes the rate of the MTase–DNA complex formation indirectly.

Methylation of the substrate DNA

We have studied the impact onto the M.Ecl18kI methylation function caused by the replacement of residues Lys21, Lys31, Lys46, or Lys53 in the M.Ecl18kI N-terminal region by Ala. For this purpose, the K_d values of the complexes between the MTase mutants and the 30-bp duplex **V** containing the methylated site, together with the methylation rate of this substrate, were determined (*Table*). The mutants M.Ecl18kI(K21A) and M.Ecl18kI(K31A) efficiently bind to the substrate duplex **V**. The rate of DNA methylation by these proteins is 30- to 40-fold higher as compared to that of the wild-type enzyme. Contrariwise, the mutants M.Ecl18kI(K46A) and M.Ecl18kI(K53A) are characterized by low affinity to the duplex **V** and cannot methylate it.

Similar studies were conducted for the mutant proteins where one of the Arg residues (R15, R35, R38, R39, or R42) located in the M.Ecl18kI N-terminal region was replaced by Ala (*Table*) [9]. The following conclusions can be drawn when comparing the results. The affinity of M.Ecl18kI mutants to the substrate DNA **V** generally decreases when the amino acid substitution approaches the M.Ecl18kI region responsible for methylation (*Table*). Thus, the K_d values of the M.Ecl18kI(R15A), M.Ecl18kI(K21A), and M.Ecl18kI(R38A) complexes with the duplex **V** coincide within the experimental error with those for the wild-type M.Ecl18kI. The M.Ecl18kI(K31A) affinity to the DNA duplex **V** is even 3-fold higher. In M.Ecl18kI(R35A), the affinity to the methylation region is 1.6-fold lower than in the wild-type MTase; in M.Ecl18kI(R39A) and M.Ecl18kI(R42A), it is 3-fold lower; whereas M.Ecl18kI(K46A) and M.Ecl18kI(K53A) virtually do not bind to the duplex **V**.

M.Ecl18kI(K21A), M.Ecl18kI(K31A), and M.Ecl18kI(R39A) methylate the substrate duplex **V** very efficiently; however, no correlation in the affinity to the methylation site or in the transcription regulation capacity is observed among these mutant forms (*Table*). A higher degree of the duplex **V** methylation (as compared with that of M.Ecl18kI) was also observed for M.Ecl18kI(R38A) with the regulatory function turned off. The efficiency of M.Ecl18kI(R35A) methylation remains virtually unchanged, although its affinity to the methylation site decreases 1.6-fold when the regulatory function is turned off. The replacement of Arg15 and Arg42 by Ala results in a 2.5- to 3-fold decrease in the enzyme's methylation ability. The mutants M.Ecl18kI(K46A) and M.Ecl18kI(K53A) cannot methylate the duplex **V** because of their extremely low affinity.

Thus, our results demonstrate that amino acid substitutions in the M.Ecl18kI region responsible for regulation affect the ability of this protein to bind to the substrate DNA and methylate it, although no clear pattern is observed among the mutant forms.

CONCLUSIONS

The modification enzyme (cytosine-5)-DNA MTase Ecl18kI *in vitro* regulates transcription of the genes in

the Ecl18kI restriction–modification system. The inhibition of the MTase gene transcription is caused by competition between RNAP and the modification enzyme for the binding site near the MTase gene promoter. Transcription of the restriction endonuclease Ecl18kI gene is activated due to the attenuation of transcriptional interference resulting from the modification enzyme binding to the regulatory site. It is demonstrated for the first time that the presence of the MTase region responsible for methylation is required for this enzyme to function as a transcription factor. The point mutation turning off the MTase catalytic function increases the mutant affinity to the regulatory sequence and does not affect its ability to act as a transcription factor. On the other hand, the mutants M.Ecl18kI(K46A) and M.Ecl18kI(K53A), which efficiently regulate transcription in the Ecl18kI R–M system, do not modify the substrate DNA because of the extremely low affinity to the methylation site. The replacement of Arg35 or Arg38 in MTase Ecl18kI by Ala not only impairs protein binding to the regulatory site, but also impedes its performing of the regulatory function; however, the efficiency of DNA methylation is considerably enhanced in this case. Evidently, there is a relationship between the functioning of the two DNA recognition centers in the SsoII-like MTases. ●

This work was supported by the Russian Foundation for Basic Research (grants № 12-04-01399 and 12-04-32103) and the Target-Oriented Program "Scientific and Scientific-Pedagogical Personnel of Innovative Russia" for 2009–2013 (state contract № P1045).

REFERENCES

1. Kobayashi I. // Nucl. Acids Res. 2001. V. 29. P. 3742–3756.
2. Nagornykh M.O., Bogdanova E.S., Protsenko A.S., Zakharova M.V., Severinov K.V. // Russian Journal of Genetics. 2008. V. 44. P. 523–532.
3. Karyagina A., Shilov I., Tashlitskii V., Khodoun M., Vasil'ev S., Lau P.C.K., Nikolskaya I. // Nucl. Acids Res. 1997. V. 25. P. 2114–2120.
4. Denjmukhametov M.M., Brevnov M.G., Zakharova M.V., Repyk A.V., Solonin A.S., Petrauskene O.V., Gromova E.S. // FEBS Lett. 1998. V. 433. P. 233–236.
5. Zakharova M.V., Beletskaya I.V., Denjmukhametov M.M., Yurkova T.V., Semenova L.M., Shlyapnikov M.G., Solonin A.S. // Mol. Genet. Genomics. 2002. V. 267. P. 171–178.
6. Miyahara M., Ishiwata N., Yoshida Y. // Biol. Pharm. Bull. 1997. V. 20. P. 201–203.
7. Ibáñez M., Alvarez I., Rodríguez-Peña J.M., Rotger R. // Gene. 1997. V. 196. 145–158.
8. Protsenko A., Zakharova M., Nagornykh M., Solonin A., Severinov K. // Nucl. Acids Res. 2009. V. 37. P. 5322–5330.
9. Fedotova E.A., Protsenko A.S., Zakharova M.V., Lavrova N.V., Alekseevsky A.V., Oretskaya T.S., Karyagina A.S., Solonin A.S., Kubareva E.A. // Biochemistry (Moscow). 2009. V. 74. P. 85–91.
10. Vorob'eva O.V., Vasil'ev S.A., Karyagina A.S., Oretskaya T.S., Kubareva E.A. // Molekulyarnaya Biologiya. 2000. V. 34. P. 921–926.
11. Shilov I., Tashlitsky V., Khodoun M., Vasil'ev S., Alekseev Y., Kuzubov A., Kubareva E., Karyagina A. // Nucl. Acids Res. 1998. V. 26. P. 2659–2664.
12. Ryazanova A.Yu., Molochkov N.V., Abrosimova L.A., Alexeevsky A.V., Karyagina A.S., Protsenko A.S., Friedhoff P., Oretskaya T.S., Kubareva E.A. // Mol. Biol. (Moscow). 2010. V. 44. P. 807–816.
13. Karyagina A.S., Alexeevski A.V., Golovin A.V., Spirin S.A., Vorob'eva O.V., Kubareva E.A. // Biophysics. 2003. V. 48. Suppl. 1. P. S45–S55.
14. Severinov K., Darst S.A. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. P. 13481–13486.
15. Scatchard G. // Annals New York Acad. Sci. 1949. V. 51. P. 660–672.
16. Callen B.P., Shearwin K.E., Egan J.B. // Trends Genet. 2005. V. 21. P. 339–345.
17. Vijesurier R.M., Carlock L., Blumenthal R.M., Dunbar J.C. // J. Bacteriol. 2000. V. 182. P. 477–487.
18. Ryazanova A.Yu., Winkler I., Friedhoff P., Viryasov M.B., Oretskaya T.S., Kubareva E.A. // Nucleosides, Nucleotides and Nucleic Acids. 2011. V. 30. P. 632–650.
19. Cheng X., Kumar S., Posfai J., Pflugrath J.W., Roberts R.J. // Cell. 1993. V. 74. P. 299–307.

Factors Affecting Aggregate Formation in Cell Models of Huntington's Disease and Amyotrophic Lateral Sclerosis

V. F. Lazarev*, D. V. Sverchinskyi, M. V. Ippolitova, A. V. Stepanova, I. V. Guzhova, B. A. Margulis
*E-mail: vl.lazarev@gmail.com
Institute of Cytology, Russian Academy of Sciences, Tikhoretsky ave., 4, St. Petersburg, 194064
Received 29.09.2012

ABSTRACT **Most neurodegenerative pathologies stem from the formation of aggregates of mutant proteins, causing dysfunction and ultimately neuronal death. This study was aimed at elucidating the role of the protein factors that promote aggregate formation or prevent the process, respectively, glyceraldehyde-3-dehydrogenase (GAPDH) and tissue transglutaminase (tTG) and Hsp70 molecular chaperone. The siRNA technology was used to show that the inhibition of GAPDH expression leads to a 45–50% reduction in the aggregation of mutant huntingtin, with a repeat of 103 glutamine residues in a model of Huntington's disease (HD). Similarly, the blockage of GAPDH synthesis was found for the first time to reduce the degree of aggregation of mutant superoxide dismutase 1 (G93A) in a model of amyotrophic lateral sclerosis (ALS). The treatment of cells that imitate HD and ALS with a pharmacological GAPDH inhibitor, hydroxynonenal, was also shown to reduce the amount of the aggregating material in both disease models. Tissue transglutaminase is another factor that promotes the aggregation of mutant proteins; the inhibition of its activity with cystamine was found to prevent aggregate formation of mutant huntingtin and SOD1. In order to explore the protective function of Hsp70 in the control of the aggregation of mutant huntingtin, a cell model with inducible expression of the chaperone was used. The amount and size of polyglutamine aggregates were reduced by increasing the intracellular content of Hsp70. Thus, pharmacological regulation of the function of three proteins, GAPDH, tTG, and Hsp70, can affect the pathogenesis of two significant neurodegenerative diseases.**
KEYWORDS **neurodegenerative pathologies; glyceralgehyde-3-phosphate dehydrogenase; chaperones; mutant proteins; aggregation.**
ABBREVIATIONS: EGFP – enhanced green fluorescence protein; ALS – amyotrophic lateral sclerosis; HSP – heat shock protein; HD – Huntington's disease; GAPDH – glyceraldehyde-3-phosphate dehydrogenase; HNE – hydroxynonenal; SDS – sodium dodecyl sulfate; PAAG – polyacrylamide gel; SOD – superoxide dismutase; tTG – tissue transglutaminase; PBS – phosphate buffer saline.

INTRODUCTION

Progressive neuronal death in certain parts of the brain is the culprit in most neurodegenerative disorders. The development of these pathologies starts from intra- (Parkinson's and Huntington's diseases) or extracellular accumulation (Alzheimer's disease) of the aggregates of mutant proteins or their oligomers [1]. These structures are toxic for brain cells; they cause immediate neuronal death, although there is some evidence that they can exist in neurons for dozens of years and turn into an active toxic factor only at some moment in time [2].

There are two hypotheses for the aggregate formation of mutant proteins. According to one, the aggregates can form due to the formation of hydrogen bonds between the β-sheets of a damaged or a mutant protein molecule [3]. These structures are inaccessible to strong dissociating solvents, in particular to sodium dodecylsulfate (SDS). The high density of the aggregating material presumably prevents the cell from using proteolytic systems, proteasomes, and phagosomes to fight the aggregates [4]. According to the second hypothesis, amyloid aggregates can form due to covalent cross-links between mutant protein molecules and other cell proteins. The formation of such cross-links is typical of the so-called polyglutamine pathologies, which are based on mu-

tations resulting in the formation of anomalously long glutamine repeats in huntingtin molecules, various ataxins, and the androgen receptor [5–7]. Polyglutamine repeats can also be deleted in the reaction catalyzed by tissue transglutaminase (tTG) and covalently bind to donors of lysine ε-amino groups. The hypothesis about the key role of tTG in the formation of insoluble aggregates of mutant huntingtin is supported by data that indicate that the polyglutamine domains of huntingtin act as the active substrate for this enzyme; aggregation almost stops in the absence of tTG [8]. The glycolytic enzyme glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) can be used as a lysine donor in the reaction catalyzed by tTG [9]. We recently showed that GAPDH is indeed capable of forming aggregates with mutant huntingtin [10]. The data regarding the detection of GAPDH in aggregates or deposits of other mutant proteins (e.g., amyloid precursor and α-synuclein) [11, 12] provide grounds to regard GAPDH as a certain universal substrate for the aggregation processes [13]. One of the aims of our study was to determine whether GAPDH and tTG can not only participate in the aggregation of mutant huntingtin, but also contribute to the pathogenesis of a completely different disease, amyotrophic lateral sclerosis (ALS).

ALS is one of the most common neurodegenerative disorders, which manifests itself in the degeneration of neurons in the spinal cord, brain stem, and cortex [14]. ABS is inherited in approximately 14% of all cases; among those, 20% are caused by mutations in the superoxide dismutase 1 (SOD1) gene. Mutations occur in all exons of the *SOD1* gene; some of them result in folding disturbance and protein product aggregation. These mutations include the replacement of glycine 93 by alanine, G93A [15]. Thus, the first section of our study is devoted to the analysis of the functions of GAPDH and tTG in cells simulating HD and ALS.

In addition to the proteins involved in the formation of cytotoxic oligomers and aggregates, the cell contains factors that impede this process. First of all, these factors include molecular chaperons (in particular, proteins that belong to the Hsp70 family and the co-chaperons Hdj1/2) [16, 17]. An enhanced expression of the genes of these factors in the cell or in a transgenic organism inhibits the aggregation and reduces pathogenic symptoms [18]. It is believed that during the initial stages of aggregate formation chaperon Hsp70 binds their components, thus inhibiting the aggregation process [10, 19, 20]. The features of Hsp70 functioning during the aggregation of mutant huntingtin in a cell culture model of HD are considered in the second section of our study.

EXPERIMENTAL

Cells

Human neuroblastoma cell lines (SK-N-SH and SH-SY-5Y) were provided by D. Rubinztein (Cambridge University, UK); neuronal embryonic cells HNSC3148 were provided by L.I. Korochkin (Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia). The cloned cell line SK-N-SH/hsp70 had been obtained at the Laboratory of Cell Protection Mechanisms (Institute of Cytology, Russian Academy of Sciences) via cell transfection with a plasmid with an inserted *hsp70* gene under the control of the metallothionein promoter [10, 21]. The cells were grown in a DMEM medium (Biolot, St. Petersburg, Russia) with addition of *L*-glutamine, a 10% fetal bovine serum, and 50 mg/ml gentamicin (PanEco, Moscow, Russia) at 37°C in a 5% CO_2 atmosphere. SK-N-SH/hsp70 cells were grown in the presence of 100 μM genecitin.

Cell transfection

Plasmids containing exon 1 of the gene encoding huntigtin with a normal (Q25) and pathogenic (Q103) number of glutamine residues (hereinafter referred to as the *Q25* and *Q103* genes, respectively) linked to the *EGFP* gene encoding an enhanced green fluorescent protein (plasmids were provided by D. Rubinztein) and plasmids containing the wild-type superoxide dismutase 1 ($SOD1_{wt}$) gene and the mutant variant $SOD1_{G93A}$ linked to the EGFP gene (provided by M. Cheetham, University College London, UK) were used in this study. Small interfering RNA (siRNA) specific to GAPDH was purchased from Ambion (Ambion/Life Technology/Invitrogen, USA).

The cells were seeded into the wells of a 24- or 6-well plate 24 h prior to the transfection at a concentration of 3×10^5 cells/ml. Transfection was performed using the Lipofectamine–PLUS reagent (Invitrogen, USA) according to the manufacturer's recommendations.

The cells were transfected with *GAPDH* siRNA 24 h prior to transfection with the *Q103* and *Q25* genes or simultaneously with transfection with the $SOD1_{G93A}$ or $SOD1_{WT}$ gene.

Confocal microscopy

The cells were seeded onto glass coverslips placed into the wells of a 24-well plate at a concentration of 3×10^5 cells/ml. In order to determine the colocalization of GAPDH and tTG with Q103, 48 h after the transfection, the cells were fixed in 4% formaldehyde (Sigma, USA) in a phosphate buffer saline (PBS) for 30 min, washed with pure PBS, and permeabilized with cold 96% ethanol for 5 min at −20°C. The cells were incubated with specific anti-GAPDH (Abcam, UK) or anti-tTG

antibodies (Sigma, USA) overnight. After the cells had been washed in PBS, they were incubated with secondary antibodies conjugated to a CY3 fluorescent label (JacksonLab., USA). The specimens were studied using a Leica TCS SL confocal microscope (Germany). The sequential scan mode was used to avoid nonspecific interference from fluorochromes. The aggregate size was determined using a LSM510 Zeiss confocal microscope and the Zeiss LSM Image Examiner software, version 2.80.1123 (Carl Zeiss, Germany).

Determination of cell viability

In order to test the viability of the cells synthesizing pathogenic peptides against repressed GAPDH expression, we used the Mossman assay [22]. Neuroblastoma SK-N-SH cells were placed into the wells of a 96-well plate, transfected with siRNA, and subsequently with the *Q103* gene (as described above). Seventy-two hours after the onset of the experiment, the medium was removed from the wells. 90 μl of a fresh medium and 10 μl of a MTT solution (3-4,5-dimethylthiazol-2-yl-2,5-tetrazolium bromide, Sigma, USA), 5 mg/ml, in sterile PBS, were added to each well. The cells were incubated with MTT for 4 h at 37°C; then, the medium containing MTT was removed, and 100 μl of acidulated isopropanol (0.04 N HCl) was added into each well to dissolve blue formazane crystals in living cells. The signal was measured on a Fluorofot immunochemistry analyzer system (OOO "PROBANAUCHPRIBOR", Russia) at 570 and 630 nm.

Analysis of protein aggregation

Protein aggregation was analyzed using two systems (*ex vivo* and *in vitro*). In the *ex vivo* system, neuroblastoma SK-N-SH cells were transfected with plasmids containing exon 1 of the huntington gene with a normal (Q25) and pathogenic (Q103) number of glutamine residues. After 8 h, when the amount of the mutant protein accumulated in the cells was sufficient for the analysis but no aggregates had yet been formed, the cells were lysed in a buffer with the following composition: 25 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM EDTA. After the triple freeze-thaw cycle, the lysates were centrifuged at 10,000 *g*; the total protein concentration in the supernatant was measured using the Bradford protein assay [23]. The lysates were incubated at 37°C for 48 h and analyzed using the filter trap assay.

In the *in vitro* system, SK-N-SH neuroblastoma cells were transfected with plasmids encoding exon 1 of the *Q25* and *Q103* genes or with the $SOD1_{wt}$ and $SOD1_{G93A}$ genes. Twenty-four h after cell transfection with either *Q25* or *Q103* or 48 h after the transfection with *SOD1*, the cells were collected, washed thrice in cold PBS, and precipitated by centrifugation at 800 *g* for 5 min. A lysing buffer of composition 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 2% SDS was added to the dry cellular precipitate. Following ultrasonic treatment for 1 min and incubation at 98°C for 2 min, the lysates were used to study the aggregate formation using the filter trap assay or the gel retardation assay.

Filter trap assay

The filter trap assay already described by Novoselova *et al.* [24] was employed in this study. The lysates of the transfected SK-N-SH neuroblastoma cells obtained in accordance with the above-described procedure were applied onto an acetate nitrocellulose membrane placed into a Dot-Blot apparatus (Hemel Hempstead, UK) connected to a vacuum pump. The membrane was washed with a buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% SDS) under pressure before and after application of the lysates. The presence of Q103 or SOD1 in the aggregates was determined using anti-EGFP antibodies.

SDS PAGE electrophoresis and immunoblot assay

In order to prepare specimens, the cells were collected, washed with cold PBS thrice, and centrifuged at 800 *g* for 5 min. A lysing buffer of composition 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% Triton X-100, 2 mM EDTA was added to the dry cellular precipitate. The lysates were centrifuged at 10,000 *g*; total protein concentration in the supernatants was measured using the Bradford protein assay. The amount of protein per specimen was 50 μg. After the electrophoresis, the proteins were transferred from the gel to a nitrocellulose membrane (Immobilon-P (PVDF), pore size 0.45 μm, Millipore Corporation, USA) using a TransBlot system (Bio-Rad, USA).

The zones of proteins of interest were detected using primary mono- or polyclonal antibodies and secondary antibodies against mouse or rabbit immunoglobulin conjugated to horseradish peroxidase. The peroxidase reaction was identified via enhanced chemiluminescence using the Chemidoc XRC visualization system (Bio-Rad, USA).

The immunoblot assay was conducted using anti-EGFP (Abcam, UK) and anti-GAPDH (Abcam, UK) monoclonal mouse antibodies; polyclonal rabbit antibodies against Hsp70 (R22) and tissue transglutaminase (Sigma, USA). Antibodies against mouse or rabbit immunoglobulins conjugated to horseradish peroxidase (Sigma, USA) were used as secondary antibodies.

Modified SDS PAGE electrophoresis to analyze the SDS-insoluble cellular fraction (gel retardation assay)

The modified SDS PAGE electrophoresis procedure was used to analyze the level of the proteins under study in the SDS-insoluble cellular fraction. This pro-

cedure involved the retardation of insoluble complexes in the stacking gel. The cellular precipitates were dissolved in a buffer with the following composition: 62.5 mM Tris-HCl pH 8.0, 2.5% SDS, 10% glycerol, 0.1 mM EDTA, 0.02 bromophenol blue. The specimens were subjected to ultrasonic treatment for 30 s and incubated at 98°C for 5 min.

The stacking gel had the following composition: 2% acrylamide, 0.15 bisacrylamide, 0.125 mM Tris-HCl pH 6.8, 0.1 SDS, 0.06% ammonium persulfate, and 0.06% N,N,N',N'-tetramethylene diamine. The immunoblots were obtained for both the running and stacking gels.

RESULTS AND DISCUSSION

GAPH affects aggregation formation in cell models of Huntington's disease

The functions of three proteins---GAPDH, tTG, and Hsp70---during the aggregation of mutant huntingtin (model of HD) and mutant SOD1 (model of ALS) were analyzed in this study. Three cell lines, SK-N-SH and SH-SY-5Y human neuroblastoma cells and HNSC3148 human embryonic brain cells, were used as the model of HD [25]. The cells were transfected with a plasmid containing exon 1 of the *Q103* gene linked to the *EGFP* gene. Twelve h following the transfection, small bright spots emerged in the cells, which merged to form large fluorescent complexes over 100 nm in size during the subsequent 36 h (*Fig. 1A*). It should be mentioned that despite some time divergences, the patterns of aggregate formation were identical in all three types of cells. By using specific antibodies recognizing GAPDH, we demonstrated that the enzyme colocalized with aggregated polyglutamine chains (*Fig. 1A*).

It has been ascertained previously that GAPDH in SK-N-SH neuroblastoma cells is concentrated at the sites where oligomers, and subsequently large aggregates of mutant huntingtin, emerge [10]. Furthermore, this enzyme has been detected in deposits of mutant huntingtin in brain slices from patients with HD [26]. These data and the fact that this localization is observed in human embryonic cells indicate that this phenomenon is rather common and that GAPDH (whose fraction in the cellular protein can be as high as 2–4%) can affect aggregate formation during the development of HD.

The effect of GAPDH on the size of the growing aggregates of mutant huntingtin was assessed by reducing the level of this protein using siRNA. To this end, 24 h prior to launching the aggregation of mutant protein Q103, the SK-N-SH cells were transfected with the corresponding siRNA. After the reduction in the GAPDH level had been confirmed, the amount of aggregated material was determined. According to the immunoblot data, the use of this technology made it possible to reduce the amount of GAPDH in SK-N-SH human neuroblastoma cells by 50–60% (*Fig. 1B*). The filter trap assay was employed to study the aggregation process. This method allows one to determine the amount of aggregated material (with the polyglutamine repeat Q103 its major component (model of HD)) that remains on the acetate cellulose membrane after the cellular extracts obtained using SDS have passed through it. It can be clearly seen that a significant content of aggregates was detected in the control cell extract (the control cells had not been treated with siRNA) (*Fig. 1B*). The action of nonspecific siRNA (Mock siRNA) induced no changes in the aggregation process, whereas specific siRNA reduced the amount of aggregated SDS-insoluble mutant proteins (*Fig. 1B*). Moreover, it was demonstrated by counting the number of transfected cells with the aggregates and the diffusely distributed protein that the number of cells containing aggregates of the mutant protein decreased by 10% when enzyme synthesis was supressed (*Fig. 1C*).

It has been assumed that there is a direct relationship between the aggregation of mutant proteins and a reduction in viability for neuronal cells. Thus, we assessed how the suppression of GAPDH synthesis and, hence, a decrease in the aggregation level of the mutant protein affect the number of surviving cells. These experiments were conducted according to the procedure described above; however, cell viability was assessed using the Mossman assay 48 h after the transfection of siRNA and, subsequently, of the construct encoding the polyglutamine sequence (*Fig. 1D*). Indeed, the expression of mutant huntingtin was shown to reduce the number of viable cells by 40%. The preliminary transfection with a plasmid carrying the control siRNA does not affect the viability of both the control cells and those carrying Q103. The supression of GAPDH synthesis using specific siRNA resulted in a 18% increase in the number of survived cells expressing *Q103* (as compared with the cells with a normal level of GAPDH) (*Fig. 1D*).

In order to demonstrate the significance of GAPDH as a pharmacological target, we searched for low-molecular-weight compounds that exhibit affinity with GAPDH among the published data and found several compounds, including hydroxynonenal (HNE). HNE is known to be capable of reacting with cysteine and histidine residues in the enzyme molecule, thus inducing its inactivation [27]. HNE was introduced into the SK-N-SH human neuroblastoma cell culture expressing *Q103* linked to the marker gene *EGFP*. The cell lysate was subsequently obtained and analyzed via a filter trap assay. The results of these experiments give grounds to consider that GAPDH is a target for small molecules: the aggregation degree of the complex of

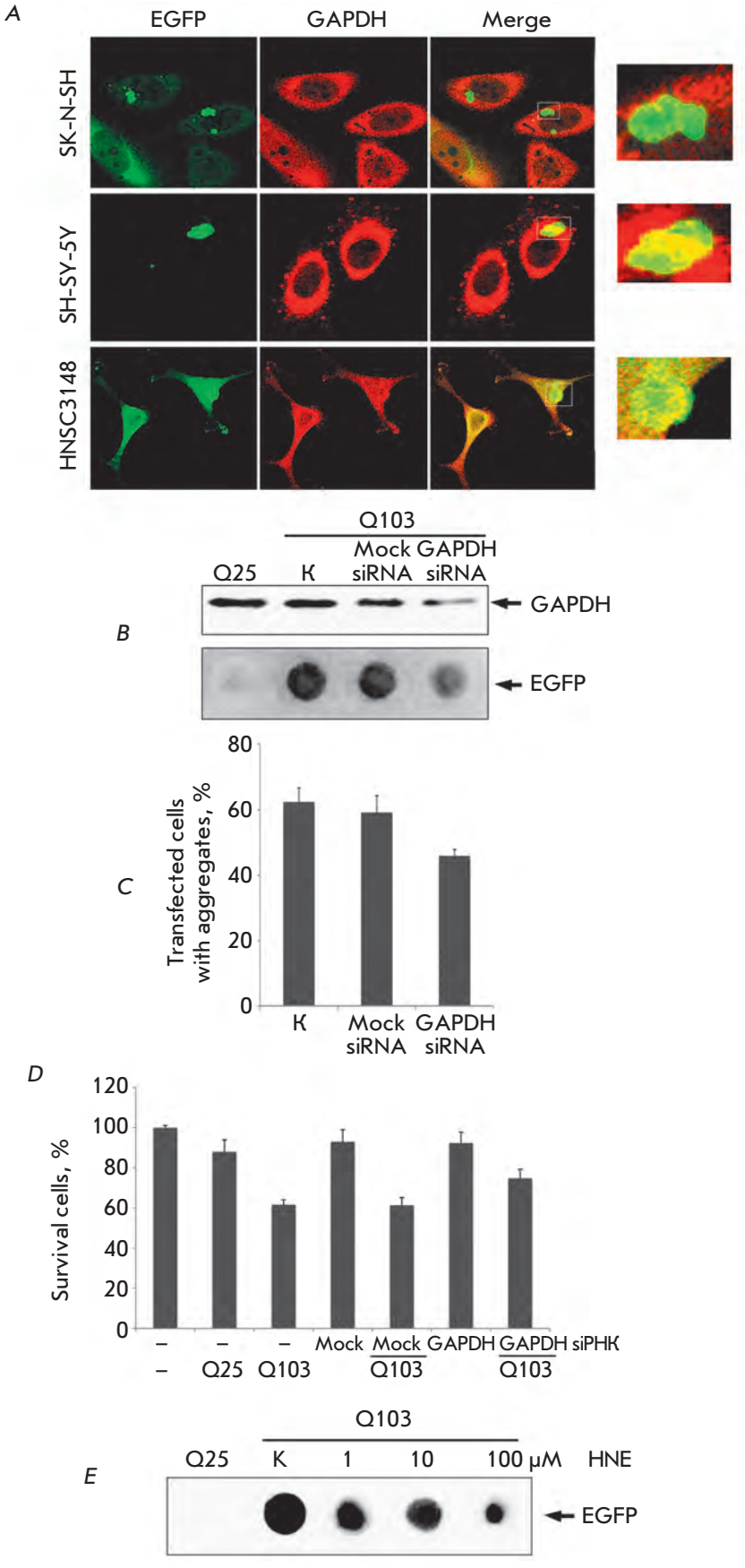


Fig. 1. GAPDH affects the aggregation process in the cell model of Huntington's disease. (a) GAPDH localizes to aggregates of mutant huntingtin. Human neuroblastoma SK-N-SH, SH-SY-5Y cells and HNSC3134 human embryo brain cells were transfected with the plasmid containing the sequence of 103 glutamines and was linked to the gene of the green fluorescent protein (EGFP). 48 h after the transfection, the cells were stained with anti-GAPDH antibody (2nd antibody is linked to a red fluorescent dye). Insert: yellow stain corresponds to the area of GAPDH and polyglutamine co-localization (488 nm and 543 nm channels). (b) Reduction in the GAPDH level attained with a specific siRNA proved by Western blotting (top panel) leads to downshifting of polyglutamine aggregation as revealed by a filter trap assay (bottom panel), (c) a decrease in the amount of aggregate-containing cells and (d) upregulation of the amount of surviving cells (data of the MTT assay) as compared with the control cells (Mock siRNA); (e) hydroxynonenal, HNE, specifically inactivating GAPDH reduces aggregation of polyglutamines in a filter trap assay

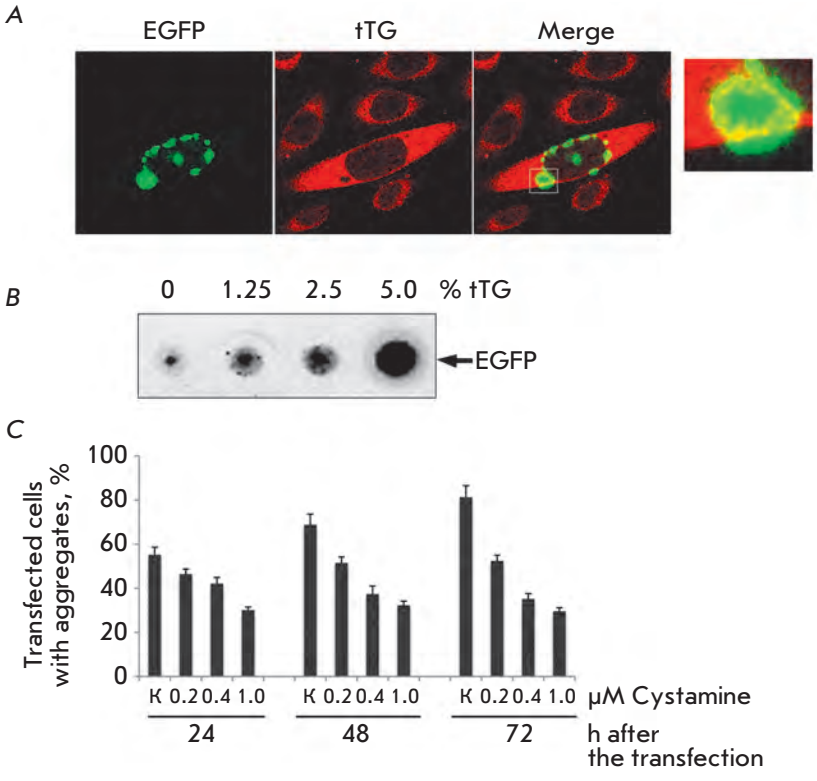


Fig. 2. Tissue transglutaminase promotes the aggregation of mutant huntingtin. (a) tTG localizes to the aggregates of mutant huntingtin. Q103, conjugated with EGFP (green), tTG – red stain. Insert: co-localization of Q103 and tTG results in yellow staining; (b) addition of tTG to the extract of cells with polyglutamine causes an increase in the amount of aggregating material in a filter trap assay (the amount of tTG is given in percents of total protein); (c) inhibitor of tTG, cystamine, reduces the aggregation of polyglutamine in a dose-dependent fashion (the ordinates represent the number of cells with aggregates)

mutant huntingtin with GAPDH decreased by 45–50% under the effect of HNE at a pharmacologically relevant concentration (1 μM). This value is a fairly good therapeutic index; the aggregation degree decreased even more dramatically with increasing concentration of the compound (*Fig. 1E*). In our opinion, the effect of HNE is based on its ability to isolate some GAPDH molecules from its complex with mutant huntingtin; the aggregation process is supposed to be inhibited.

Thus, GAPDH localizes in pathogenic aggregates, along with mutant huntingtin, and seems to participate actively in the aggregation process at its early stages. This interpretation of the results is supported by the fact that a specific decrease in the amount of enzyme molecules accessible for aggregation, which can be attained using siRNA or a high-affinity compound, results in inhibition of the aggregation of the Q103–GAPDH complex.

Investigation of the effect of tissue transglutaminase on the aggregation of mutant huntingtin

Aggregates of mutant huntingtin, ataxin, and some other pathogenic proteins can be formed via cross linking anomalously long polyglutamine chains with proteins that donate reactive lysines (in particular, GAPDH) in the tTG-catalyzed process [9, 28]. Immunofluorescence microscopy was used to determine the localization of tTG in cells that express *Q103* in order to elucidate the role of this enzyme in our cell model. It turned out that tTG molecules are uniformly distributed over the cell cytoplasm, while enzyme clusters are observed around the Q103–GAPH aggregates. One can see that tTG co-localizes with mutant huntingtin (*Fig. 2A*). The participation of tTG in aggregate formation was proved by introducing the purified enzyme into an extract of cells transfected with the Q103–EGFP construct prior to the onset of aggregate formation. The *ex vivo* analysis of aggregation (see the EXPERIMENTAL section) conducted using the filter trap assay shows that the introduction of tTG dose-dependently increases the amount of Q103 that is SDS-insoluble and remains on the membrane (*Fig. 2B*).

To what extent does the inhibition of tTG activity affect the process of aggregate formation? In order to answer this question, we incubated SK-N-SH cells with the well-known enzyme inhibitor cystamine, right after the onset of aggregation (i.e., 5 h following the transfection with the *Q103–EGFP* gene). The counting of aggregate-containing cells demonstrated that the cystamine effect manifested itself 24 h after the start of the incubation; after a day, the effect of the inhibitor became more pronounced. Finally, cell-counting 3 days after the introduction of the inhibitor into the medium showed that cystamine at a concentration of 0.4 μM re-

duces by twofold the number of cells containing Q103–EGFP aggregates, while further increase in inhibitor concentration suppresses the aggregation to a greater extent (*Fig. 2C*). Interestingly, when using cystamine at a concentration of 1 μM, the fraction of aggregate-containing cells remains constant with time and, in our case, was equal to 25–28% of the total number of transfected cells. In the untreated population, the fraction of aggregate-containing cells increased and was equal to 82% 72 h after the transfection. This fact can be an indication that the formation of polyglutamine aggregates is caused not only by the action of tTG, but also by another mechanism (e.g., that in the "polar zipper" model) [3].

Participation of GAPDH and tTG in aggregate formation in mutant SOD1

Assuming that GAPDH and tTG promote the aggregation not only of huntingtin but of other mutant proteins as well, we analyzed the functions of these enzymes in aggregate formation by the example of the cellular model of ALS. To this end, plasmids carrying the $SOD1_{G93A}$ and $SOD1_{wt}$ genes linked to the green fluorescent protein gene were used. A microscopic analysis of SK-N-SH neuroblastoma cells transfected with these plasmids demonstrated that the mutant $SOD1_{G93A}$, unlike $SOD1_{wt}$, can form aggregates in 36–48 h (*Fig. 3A*).

Does GAPDH play a role in the formation of aggregates of mutant SOD1 as important as that in the model of HD? In order to answer this question, we used the technology of specific small interfering RNAs. Lysates of SK-N-SH cells simultaneously transfected with specific or control siRNA and $SOD1_{wt}$ or $SOD1_{G93A}$ were analyzed by the gel retardation assay and immunoblot assay and using the filter trap assay (*Fig. 3B*).

As follows from the electrophoresis data, the content of GAPDH that can penetrate the running gel in the lysate of the cells treated with specific siRNA is considerably lower than that in the lysates of the control (intact) cells and cells carrying $SOD1_{wt}$ (*Fig. 3B*, middle panel). Both in the control cells and in the cells transfected with Mock siRNA (both cell types carrying mutant SOD1), the level of GAPDH that can penetrate the running gel is reduced. However, it is the lysates obtained from these cells that contain a significant amount of the aggregates remaining in the stacking gel (*Fig. 3B*, top panel). These results have also been supported in an experiment employing the filter trap assay. Aggregates of mutant SOD1 (presumably containing GAPDH) were detected in the lysates of these cells (*Fig. 3B*, bottom panel).

In addition to GAPDH-specific siRNA, HNE was also used to repress GAPDH. We demonstrated using the filter trap assay that HNE at a concentration of 1 μM

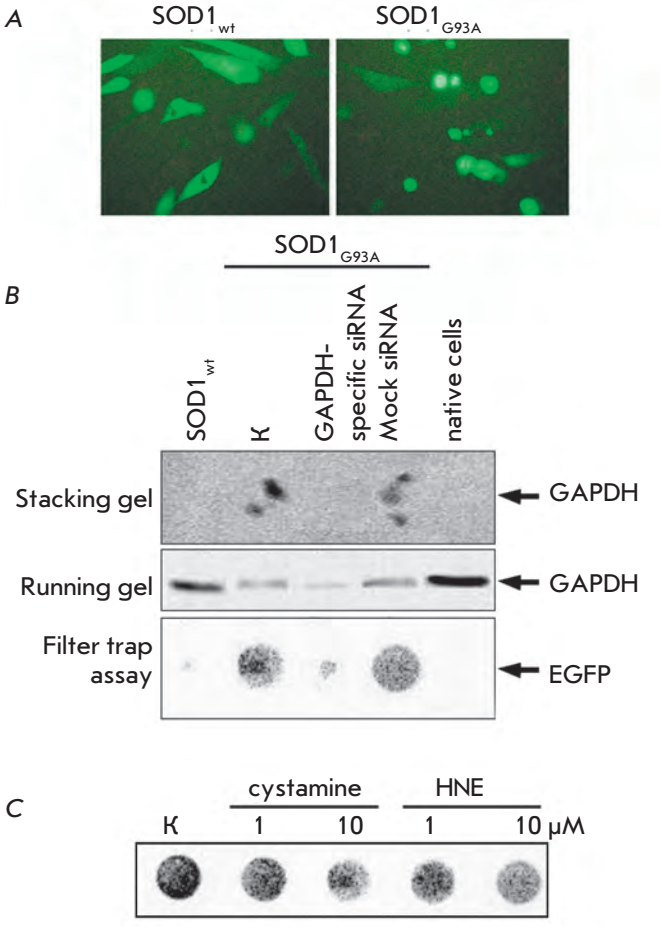


Fig. 3. GAPDH participates in the formation of aggregates in a cell model of amyotrophic lateral sclerosis (ALS). (a) SK-N-SH human neuroblastoma cells 48 h after the transfection with the gene of mutant SOD1 (G93A) or wild-type SOD1, conjugated with the gene of a green fluorescent protein (EGFP). Right panel: SOD1-G93A forms insoluble aggregate structures; (b) application of GAPDH-specific siRNA prevents aggregation of mutant SOD1 (data of gel shift assay, two top panels and data of filter trap assay, bottom panel); (c) incubation of ALS-imitating cells with cystamine and HNE down-regulates SOD1 aggregation (filter trap assay)

represses the aggregation of mutant SOD1; an increase in HNE concentration to 10 μM strengthened this effect (*Fig. 3C*). The effect of HNE can be attributed to the fact that the formation of free radicals is increased in patients with ALS, as well as in those with numerous other pathological processes, while the oxidative stress disrupts the GAPDH structure. The regions of the enzyme molecule are exposed and bind to mutant proteins, giving rise to large complexes [29]. We hy-

pothesize that HNE impedes the formation of these complexes; i.e., it inhibits SOD1 aggregation.

The participation of tTG in the formation of SOD1–GAPDH aggregates has also been demonstrated using inhibitory analysis. We used cystamine to ascertain that the suppression of tTG activity reduces the weight of the aggregating material on a filter. However, this effect can be achieved at high cystamine concentrations (at least 10 μM) exceeding pharmacological values (*Fig. 3C*). It is harder to interpret the fact of suppression of aggregation of mutant SOD1 when using cystamine. It is possible that he inhibition with tTG prevents the formation of covalent bonds both between GAPDH molecules and between GAPDH molecules and other proteins.

Hsp70 chaperon represses the aggregation of mutant huntingtin in the cellular model of HD

Hsp70 chaperon plays a significant role in preventing complex formation between damaged or mutant polypeptides [30]. The effect of Hsp70 on the aggregation of mutant huntingtin was studied using SK-N-SH neuroblastoma cells transfected with the *Hsp70* gene under the control of an inducible metallothionein promoter as a model. *Hsp70* expression was induced using a zinc salt ($ZnSO_4$), which can be used to increase the protein level in a dose-dependent manner (*Fig. 4A*), 6 h prior to the transfection of SK-N-SH cells with a plasmid carrying the *Q103* gene. The diameter of the aggregates of mutant huntingtin was determined 48 h after the transfection. The mean diameter of the aggregates in the transfected cells treated with 50 μM $ZnSO_4$ was 3.15 ± 0.69 μm, while in the control cells it was equal to 6.82 ± 0.98 μm. A further decrease both in the number of aggregate-carrying cells and aggregate size (the mean diameter being equal to 1.52 ± 0.19 μm, *Fig. 4B*) was observed as the $ZnSO_4$ concentration was increased to 100 μM. The effect of Hsp70 on the amount of aggregating material was analyzed using the filter trap assay. It turned out that an increase in $ZnSO_4$ concentration and, hence, in the Hsp70 level results in a decrease in the amount of material containing the EGFP marker retained on the filter surface (*Fig. 4C*).

It has been known that molecular chaperons, in particular Hsp70, participate in the prevention of aggregate formation in pathogenic or damaged proteins; however, the mechanisms underlying this effect remain unclear. In this context, data indicating that the chaperon forms a complex with the aggregating mono- and oligomers of mutant huntingtin are of significant interest [31]. This complex is of a dynamic nature. It is assumed that Hsp70 impedes the incorporation of polyglutamine molecules into the aggregates of mutant chromatin being formed. However, the results of

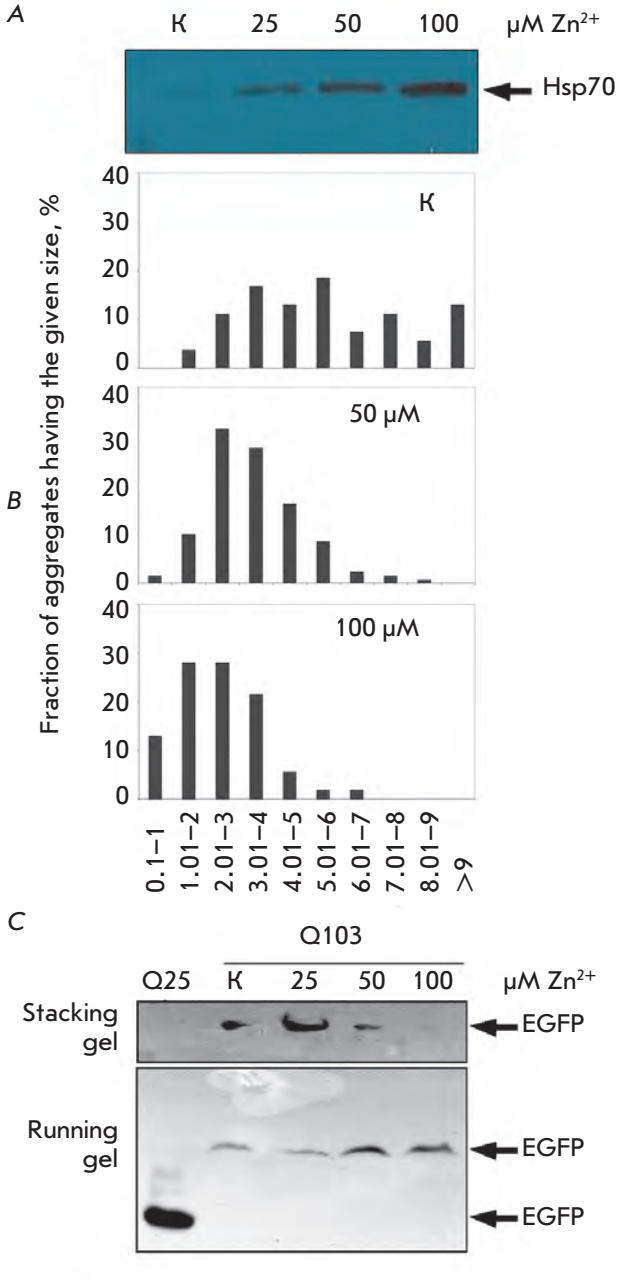


Fig. 4. Hsp70 prevents aggregation of mutant huntingtin in a cell model of HD. (a) Expression of Hsp70 gene controlled by the metallothionein promoter in SK-N-SH cells is dose-dependently induced by zinc; (b) histograms demonstrating the distribution of the Q103 aggregate size dependent on the zinc concentration and, accordingly, the level of Hsp70 expression; (c) the stepwise increase in Hsp70 content in SK-N-SH neuroblastoma cells transfected with Q103 leads to a reduction in polyglutamine aggregation in a filter trap assay

a recently published study [10] indicate that Hsp70 affects not only mutant huntingtin, but also GAPDH molecules, which significantly enhance the aggregation. Based on the results of this study an assumption can be made that at least in the cell models of HD, Hsp70 is capable of impeding the formation of the polyglutamine–GAPDH complex, thus protecting the enzyme against the cross-linking effect of tTG. This hypothesis does not contradict the theories of the function of chaperons in protecting cells against neuropathogenic stimuli. However, it undoubtedly requires thorough verification.

CONCLUSIONS

It has been ascertained in this study using cell models of Huntington's disease and amyotrophic lateral sclerosis that two enzymes–glyceraldehyde-3-phosphate-dehydrogenase and tissue transglutaminase–play a significant role, along with the pathogenic proteins specific to these disorders. The former enzyme participates in complex formation with pathogenic proteins in both models of the diseases; its blockage reduces the aggregation rate. Transglutaminase presumably catalyzes the formation of the GAPDH complex with pathogenic cellular proteins. Hsp70 chaperon is the factor that prevents the aggregation; an increase in its expression reduces the pathogenic symptoms in a dose-dependent fashion. The data obtained provide ground to regard all three proteins as promising pharmacological targets. ●

This work was supported by the Program of the Presidium of the Russian Academy of Sciences "Basic Research for Medical Applications" and the Russian Foundation for Basic Research (grants № 11-04-12047-ofi_m and 12-08-31523).

REFERENCES

1. Margulis B., Kinev A., Guzhova I. Heat Shock Proteins in Biology and Medicine. Kerala, India: Research Signpost, 2006. P. 305–329.
2. Evert B.O., Wüllner U., Klockgether T. // Cell Tissue Res. 2000. V. 301. № 1. P. 189–204.
3. Perutz M.F., Johnson T., Suzuki M., Finch J.T. // Proc. Natl. Acad. Sci. USA. 1994. V. 91. № 12. P. 5355–5358.
4. Bence N.F., Sampat R.M., Kopito R.R. // Science. 2001. V. 292. № 5521. P. 1552–1555.
5. Kahlem P., Terre C., Green H., Djian P. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. № 25. P. 14580–14585.
6. Fischbeck K.H. // Brain Res. Bull. 2001. V. 56. P. 161–163.
7. Cooper A.J., Sheu K.R., Burke J.R., Strittmatter W.J., Gentile V., Peluso G., Blass J.P. // J. Neurochem. 1999. V. 72. P. 889–899.
8. Lai T.S., Tucker T., Burke J.R., Strittmatter W.J., Greenberg C.S. // J. Neurochem. 2004. V. 88. № 5. P. 1253–1260.
9. Orru S., Ruoppolo M., Francese S., Vitagliano L., Marino G., Esposito C. // Protein Sci. 2002. V. 11. P. 137–146.
10. Guzhova I.V., Lazarev V.F., Kaznacheeva A.V., Ippolitova M.V., Muronetz V.I., Kinev A.V., Margulis B.A. // Hum. Mol. Genet. 2011. V. 20. № 20. P. 3953–3963.
11. Wang Q., Woltjer R.L., Cimino P.J., Pan C., Montine K.S., Zhang J., Montine T.J. // FASEB J. 2005. V. 19. P. 869–871.
12. Tsuchiya K., Tajima H., Kuwae T., Takeshima T., Nakano T., Tanaka M., Sunaga K., Fukuhara Y., Nakashima K., Ohama E., et al. // Eur. J. Neurosci. 2005. V. 21. № 2. P. 317–326.
13. Butterfield D.A., Hardas S.S., Lange M.L. // J. Alzheimers Dis. 2010. V. 20. № 2. P. 369–393.
14. Oster C., Pagnini F. // Front. Psychol. 2012. V. 3. A. 530.
15. Al-Chalabi A., Jones A., Troakes C., King A., Al-Sarraj S., van den Berg L.H. // Acta Neuropathol. 2012. V. 124. № 3. P. 339–352.
16. Kobayashi Y., Kume A., Li M., Doyu M., Hata M., Ohtsuka K., Sobue G. // J. Biol. Chem. 2000. V. 275. № 12. P. 8772–8778.
17. Muchowski P.J., Wacker J.L. // Nat. Rev. 2005. V. 6. P. 11–22.
18. Jana N.R., Tanaka M., Wang G. // Hum. Mol. Genet. 2000. V. 9. № 13. P. 2009–2018.
19. Hartl F.U., Hayer-Hartl M. // Science. 2002. V. 295. № 5561. P. 1852–1858.
20. Ben-Zvi A., De Los Rios P., Dietler G., Goloubinoff P. // J. Biol. Chem. 2004. V. 279. № 36. P. 37298–37303.
21. Tikhonov N.S., Moskalev O.S., Margulis B.A. Guzhova I.V. // Cytology. 2008 . T. 50 . № 5. P. 467–472.
22. Mosmann T. // J. Immunol. Meth. 1983. V. 65. № 1–2. P. 55–63.
23. Bradford M.A. // Anal. Biochem. 1976. V. 72. P. 248–254.
24. Novoselova T.V., Margulis B.A., Novoselov S.S., Sapozhnikov A.M., van der Spuy J., Cheetham M.E., Guzhova I.V. // J. Neurochem. 2005. V. 94. № 3. P. 597–606.
25. Poltavtseva R.A., Marey M.V., Aleksandrova M.A., Revishchin A.V., Korochkin L.I., Sukhikh G.T. // Brain Res. Dev. Brain Res. 2002. V. 134. № 1–2. P. 149–154.
26. Kish S.J., Lopes-Cendes I., Guttman M., Furukawa Y., Pandolfo M., Rouleau G.A., Ross B.M., Nance M., Schut L., Ang L., et al. // Arch. Neurol. 1998. V. 55. № 10. P. 1299–1304.
27. Uchida K., Stadtman E.R. // J. Biol. Chem. 1993. V. 268. № 9. P. 6388–6393.
28. Cooper A.J., Sheu K.R., Burke J.R., Onodera O., Strittmatter W.J., Roses A.D., Blass J.P. // Proc. Natl. Acad. Sci. USA. 1997. V. 94. № 23. P. 12604–12609.
29. Naletova I., Schmalhausen E., Kharitonov A., Katrukha A., Saso L., Caprioli A., Muronetz V. // Biochim. Biophys. Acta. 2008. V. 1784. № 12. P. 2052–2058.
30. Meriin A.B., Sherman M.Y. // Int. J. Hyperthermia. 2005. V. 21. № 5. P. 403–419.
31. Muchowski P.J., Schaffar G., Sittler A., Wanker E.E., Hayer-Hartl M.K., Hartl F.U. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. № 14. P. 7841–7846.

Identification of Novel IGF1R Kinase Inhibitors by Molecular Modeling and High-Throughput Screening

R. Moriev[1], O. Vasylchenko[1], M. Platonov[1], O. Grygorenko[1,2], K. Volkova[1] and S. Zozulya[1,*]
[1]Enamine Ltd, Chervonotkatska Str., 78, Kyiv, Ukraine, 02094
[2]Kyiv National Taras Shevchenko University, Volodymyrska Str., 64, Kyiv, Ukraine, 01601
*E-mail: s.zozulya@enamine.net
Received 04.12.2012

ABSTRACT **The aim of this study was to identify small molecule compounds that inhibit the kinase activity of the IGF1 receptor and represent novel chemical scaffolds, which can be potentially exploited to develop drug candidates that are superior to the existing experimental anti-IGF1R therapeutics. To this end, targeted compound libraries were produced by virtual screening using molecular modeling and docking strategies, as well as the ligand-based pharmacophore model. High-throughput screening of the resulting compound sets in a biochemical kinase inhibition assay allowed us to identify several novel chemotypes that represent attractive starting points for the development of advanced IGF1R inhibitory compounds.**
KEYWORDS **IGF1 receptor; tyrosine kinase inhibitor; anti-cancer drug candidate; high-throughput screening; virtual screening.**
ABBREVIATIONS **IGF1R – insulin-like growth factor type 1 receptor; InsR – insulin receptor; RTK – receptor tyrosine kinase; TKI – tyrosine kinase inhibitor; HTS – high-throughput screening; ATP – adenosine triphosphate; ADP – adenosine diphosphate.**

INTRODUCTION

The receptor of insulin-like growth factor type 1 (IGF1R) is a transmembrane receptor tyrosine kinase (RTK) widely expressed in various cell types and the tissues of all vertebrates. IGF1R is the key biological regulator of cell growth and survival, both in the developmental and adult states. The receptor is a very close phylogenetic relative of the insulin receptor (InsR), the major regulator of carbohydrate homeostasis, as well as lipid and protein metabolism. IGF1R shares almost 60% overall homology with InsR; the similarity is much higher (~ 90%) in the catalytic domain area of the receptors. The IGF pathway is commonly dysregulated in many human cancers, including breast, prostate, liver, lung, bladder, thyroid, renal cancers, Ewing's sarcomas, rhabdomyosarcoma, lymphomas, leukemias, multiple myeloma, etc., primarily via increased expression of IGF1R or its ligands, IGF-1 and IGF-2, and autocrine loops [1, 2, 3]. The IGF-1 receptor is needed for the transformation of cells by oncogenes; enhanced IGF-1 receptor expression can cause ligand-dependent, malignant transformation and tumorigenesis [4]. Mutated, constitutively upregulated forms of IGF1R kinase as cancer drivers have not been documented in the literature, contrary to the paradigm for oncogenic tyrosine kinases. The general anti-proliferative and pro-apoptotic effects associated with IGF1R inhibition, as well as the broad expression of IGF1R in tumors, are suggestive of a high clinical potential for IGF1R inhibitors in combination therapies. Because of the broad malignant neoplasia linkage to the ubiquitous IGF signaling pathway, therapeutic strategies that inhibit the IGF1R receptor using either small-molecule kinase inhibitors (TKIs) or monoclonal antibodies (mABs) have been actively explored in various types of cancers by a large number of pharmaceutical and biotechnology companies over the past 10–15 years. At least a dozen IGF1R inhibitors, both small-molecules and antibodies, are currently in late preclinical or clinical development.

Due to the very high degree of homology among the catalytic domains of IGF1R and InsR RTKs, all of the known advanced IGF1R-targeting TKIs inhibit InsR to a significant degree, as well. As a result, these TKIs obviously could impair glucose homeostasis and lead to hyperglycemia and the concomitant diabetic complications in TKI-treated patients. Indeed, such hyperglycemic effects have been observed in pre-clinical models and, more recently, in the clinical trials of small-molecule IGF1R inhibitors, casting some doubt on the perspectives for their long-term clinical development and therapeutic use. Over the past decade, the obvious

selectivity problem associated with small molecules has led to a shift in interest towards the development of an intrinsically, highly selective monoclonal antibody or protein-based IGF1R blockers which target either the receptor itself or its ligands. However, due to the peculiarities of IGF signaling axis biology and the resulting substantial cross-talk between IGF1R and InsR-driven signaling, some degree of InsR co-inhibition is believed to be beneficial in oncology settings by most experts [5, 6]. On the other hand, the concept of precisely blocking IGF1R signaling by pharmacological agents that are highly selective at the molecular level turns out to be a gross oversimplification when applied to the actual systemic action of these agents *in vivo*. Both a lack of efficacy and hyperglycemic effects were found in the late pre-clinical and clinical studies of several advanced anti-IGF1R protein/antibody-based therapeutic candidates despite their ultimate molecular selectivity for the target [4, 7]. Some of the apparent underlying mechanisms of this non-selectivity *in vivo* are as follows: significant cross-activation of IGF1R receptors by insulin and vice versa; activity of anti-IGF1R antibodies on InsR-IGF1R heterodimers, compensatory mechanisms of the living organism, such as upregulation of InsR or induction of IGF1 and insulin biosynthesis upon depletion of IGF-1/IGF1R pools in the body [7, 8]. In addition to the selectivity aspects mentioned above, there are multiple mechanisms of resistance to IGF1R-targeted therapy, which might necessitate co-inhibition to achieve efficacy [9].

Thus, despite the solid academic validation of IGF1R as a highly attractive drug target in oncology, as well as the sustained effort to develop therapeutically useful IGF1R pathway blockers of diverse molecular nature and mechanisms of action, so far the results of late-stage clinical trials remain less than exciting [7, 10]. This controversial and complicated landscape, in addition to the certain inherent advantages of small molecule drugs over antibody/protein-based therapeutics, creates a persistent drive for a continued search for clinically superior chemical inhibitors of IGF1R. Such target compounds might differ from their predecessors by virtue of their different mechanisms of inhibition, more selective tissue distribution, co-inhibition of other targets, and altered pharmacodynamics or a better balanced selectivity for IGF1R versus InsR.

In this study, we report on the identification of several small molecule IGF1R inhibitors as a result of the screening of a focused library of 2,935 compounds generated by the combined use of pharmacophore- and target structure-based models. The compound series found represent novel chemotypes and are potentially developable into clinical IGF1R inhibitors with favorably altered properties as compared to existing ones.

EXPERIMENTAL

Reagents and materials

All the reagents for screening, including ADP-Glo™ Kinase Assay (Cat. V9401), Kinase System kits for IGF1R (Cat. V3581), InsR (Cat. V9411), Met (Cat.V3361), Syk (Cat. V3801), and BTK (Cat. V2941) kinases, were obtained from Promega Corporation (Madison, WI, USA) and used according to the manufacturer's recommendations. The reference kinase inhibitors PQ401 (Cat. P0113), AG538 (Cat. T7697), staurosporine (Cat. S5921), as well as poly(Glu4,Tyr1), sodium salt (Cat. P0275), and dimethyl sulfoxide (Cat. 41640), were all purchased from Sigma-Aldrich (St. Louis, MO, USA). The low volume, U-bottom, white NBS 384-well microplates (Cat. 3673) used for all luminescent assays were from Corning (Lowell, MA,USA), and robotic liquid handler 384-channel tips (Cat. 5316) were from Thermo Scientific/Matrix (Hudson, NH, USA). The polypropylene 384-well and V-bottom plates (Cat. 784201) were purchased from Greiner Bio-One (Monroe, NC, USA), and the 96-well plates, from Matrix (Cat. 4271), or similar ones, were used for compound storage and dilutions. The reagents and buffers for robotic multichannel pipetting were kept in disposable modular reservoirs (Cat. N372790) from Beckman Coulter (Indianapolis, IN, USA).

All the compounds iteratively tested in this study were selected from the ~1,900,000 compound collection of Enamine, Ltd. (www.enamine.net, Kyiv, Ukraine) and supplied by Enamine's library formatting facility as frozen 10 mM solutions in dimethyl sulfoxide (DMSO) in heat-sealed 96- or 384-well polypropylene plates.

Molecular modeling and chemoinformatics

All computations were done using the QXP/Flo+ software package developed by McMartin *et al.* [11]. We used the computer cluster configuration HPC Linux cluster (164 CPU cores in 5 nodes). All the manipulations with chemical structures and databases were conducted in the Instant JChem software (ChemAxon, software version 5.10.1).

Screening equipment and data analysis

Multi-well liquid dispensing for setting up assays was performed using either the robotic liquid handler PlateMate Plus or the manual electronic multichannel micropipettes Matrix Impact (Thermo Scientific, Hudson, NH, USA). High-throughput Screening (HTS), kinase selectivity, and dose-response (IC_{50}) assays were read in the luminescence mode using the PolarStar Omega reader (BMG Labtech, Ortenberg, Germany) or M5 reader (Molecular Devices Corp., Sunnyvale, CA). IGF1R ADP-Glo data in relative luminescence units

(RLU) were collected from the plate readers, and the percentage of activity (% Activity) was determined for each point as follows: % Activity = 100*(RLU sample − RLU no kinase control avg)/(RLU kinase control avg − RLU no kinase control avg). The screening data were processed and visualized using Microsoft Excel templates designed to calculate the inhibition values, the Z'-factor, and GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, CA) for IC_{50} analysis. The dose-response curves of Percent Activity were fit in Prism, using a sigmoidal variable slope fit with the maximum % activity and the minimum % activity fixed at 100% and 0%, respectively. The Z' factor, a statistical measure of variability and reproducibility for HTS assays, was determined using the following formula: $Z' = 1-[3\times(SD_{sample}+SD_{contol})/|M_{sample}-M_{control}|]$ [12], where SD denotes the standard deviation and M denotes the mean for the samples and controls, respectively. Prior to starting HTS, assay conditions were optimized and validated with regard to the maximum ATP turnover (never exceeding 20%), an acceptable signal-to-background ratio ("assay window") of at least 6, an acceptable Z'-factor of at least 0.6, as well as day-to-day and plate-to-plate reproducibility of the screening data. All primary screening was performed at 20 μM compound concentrations, and some of the weak hits were subsequently re-tested at higher concentrations (40 or 80 μM) for confirmation. Statistically significant HTS hits in the primary screening were defined as those that produced a kinase activity signal at least three standard deviations lower than the mean of the assay plate run (not including the plate controls). Lineweaver-Burk plots were created in Excel or Prism using the standard algorithms [13]. All assay development and validation and high-throughput screening procedures were carried out according to the general guidelines as published on the U.S. National Chemical Genomics Center web site (NCGC Assay Guidance Manual and High-throughput Assay Guidance Criteria, http://www.ncbi.nlm.nih.gov/books/NBK53196/).

High-throughput screening (HTS)

IGF1R high throughput screening (HTS) assays using the IGF1R Kinase System and ADP-Glo readout system (Promega Corp.) were performed in a final volume of 7 μL per test compound using a 384-well small volume microplate format. All liquid dispensing was done using the PlateMate Plus robotic liquid handler. 3 μL aliquots of the enzyme/substrate mixture containing 1 μg of the IGF1Rtide peptide substrate and 4 ng of recombinant IGF1R kinase in a 0.66 × assay buffer were transferred to the plate wells. The assay buffer (1 ×) consisted of 40 mM Tris chloride, 20 mM magnesium chloride, 0,1 mg/ml bovine serum albumin, 2 mM manganese chloride, and 250 μM dithiothreitol (DTT). To add the tested compounds to the reaction, starting stocks of 10 mM compounds in dimethyl sulfoxide (DMSO) were diluted with DMSO to 2 mM. Next, 3 μL aliquots of 2 mM compound stock solutions in DMSO were transferred to 83 μL volumes of a 1 × reaction buffer and thoroughly mixed. Aqueous compound solution aliquots of 2 μL were then transferred into the assay plate to get a final concentration of each compound of 20 μM in 1% DMSO. The plates were pre-incubated at 27˚C for 10 min with gentle shaking (300 rpm). The assay was started by adding 2 μL of the ATP stock solution to the reaction mixture to achieve a final ATP concentration of 50 μM. After 1.5 h of gentle shaking (300 rpm) at 27˚C, the ADP-Glo reagent (7 μL) and, after an additional 40 min, the detection reagent (14 μL) were added. After the final incubation for 20 min at 25˚C, luminescence was read using a PolarStar Omega multimode plate reader at a gain setting of 3500 and integration time of 0.2 s.

The compounds were tested in quadruplicates or duplicates during the screening. One quadruplicate sample per plate of 2 μM final staurosporine was used as a control inhibitor sample. For the columns 1 and 2 of each 384-well plate, 3 μL of the 0.66 × assay buffer, instead of the enzyme/substrate mixture, and DMSO-spiked 1 × assay buffer without test-compounds (final 1% DMSO in the reaction mixture) were used to produce a positive (no kinase reaction) control. For the columns 23 and 24 of each plate, a DMSO-containing buffer without test-compounds (final 1% DMSO) was used during the compound addition step to produce a negative (no kinase inhibition) control. Prior to the dose-response and selectivity studies, all primary screening hits were confirmed by re-testing the single concentration point inhibition of IGF1R by compound solutions freshly prepared from solid compound stocks under the same conditions as described above ("confirmation from powders").

Dose-response curves (IC_{50}) and kinase selectivity measurements

Kinase selectivity assays with InsR (dose-response measurements) and Met, Syk and BTK kinases (single point compound concentrations) using the ADP-Glo readout system were run under optimized experimental conditions similar to the IGF1R assay with regard to the kinetic parameters of kinase reactions.

Dose-response and IC_{50} measurements for the confirmed screening hits were conducted for IGF1R and InsR kinases. The conditions were similar to the HTS conditions described above, except that the compounds were plated in 8-point curves serially diluted 1:2 from 100 μM top concentrations and with a 1% final DMSO

throughout, in quadruplicates for each compound dilution point. The time of incubation of kinase reaction mixtures was 2.5 or 4.5 hours at 27°C in different experiments both for the insulin kinase receptor and IGF1R, the dithiotreitol (DTT) concentration was 500 μM, and the amount of recombinant insulin receptor kinase was 2 ng per well. DTT concentration was elevated to enhance the stability of the enzymes during 4.5 h incubation experiments. Compounds were typically serially 2-fold diluted in pure DMSO starting from 10 mM down to 19.5 μM to produce the final concentrations ranging from 100 μM to 195 nM in a 1% DMSO-aqueous reaction buffer.

The amounts of the Met, Syk, and BTK enzymes and the incubation time were optimized to ensure an ATP conversion not higher than 20% in all cases. The typical assay window (signal/background) was 3–5 for all kinases. The final volumes of the Met, Syk, and BTK kinase assay reaction mixtures were 5 μL. Two μL of the enzyme solution in a 1 × reaction buffer (6 ng Met, 8 ng BTK, 4 ng Syk), 1 μL of the compound solution in a 2 × reaction buffer, the ATP and substrate mixture in a 0.5 × reaction mixture were added sequentially. The composition of the 1 × reaction buffer was 40 mM Tris, 20 mM magnesium chloride, 0.1 mg/ml bovine serum albumin, 500 μM DTT at pH 7.5 with 1% final concentration of the DMSO, 50 μM ATP, and 0.2 mg/ml Poly(Glu4, Tyr1) substrates. The buffer was supplemented with 2 mM manganese (II) chloride in the case of BTK kinase. Compounds were tested at 40 μM concentrations. In each experiment, 6 wells with the enzyme, but no added compounds, were used as negative inhibition controls; 6 wells without tested compounds and the enzyme and 6 wells with the enzyme and 0.5–1 μM staurosporine were used as positive inhibition controls. Pre-incubation of the reaction mixture with compounds prior to ATP addition was performed for 20 min at 25°C in all experiments. The incubation of the reaction mixture lasted for 25 min at 37°C. All compounds were tested in 4 to 6 repeats. Conditions of ADP detection were as follows: 40 min incubation with an ADP-Glo reagent, followed by 30 min with a detection reagent at 25°C. Luminescence was read using a BMG Polarstar Omega reader at a gain setting of 4095 and measurement time of 0.5 s.

Measurements of the Michaelis–Menten kinetics

Serial dilutions of ATP and substrate polypeptide poly(Glu4,Tyr1) were tested in IGF1R kinase assays with the inhibitors L1 and T4 to produce the kinetic data for Lineweaver-Burk plots. In the ATP competition measurements, solutions of ATP and the compounds being tested were subjected to twofold serial dilution. Compound L1 was tested at concentrations of 100, 50, 25, 12, and 0 μM in combination with eight ATP concentrations ranging from 519 to 4 μM. Compound T4 was tested at the same concentrations in combination with eight ATP concentrations ranging from 1 mM to 8 μM. IGF1Rtide at a concentration of 143 μg/ml was used as a peptide substrate in both cases. All concentration points were quadruplicated. The amount of IGF1R kinase was 1 ng per well, the DTT concentration was 500 μM, and the the kinase reaction was incubated for 4 h at 27°C. The range of ATP concentrations used for the plot was narrowed to 6 points for L1 to get the best fit.

In the substrate competition measurements, the poly(Glu4,Tyr1) substrate was titrated by twofold dilutions to obtain 8 concentrations from 0.9 to 114.3 μM, assuming the average molecular mass of the substrate to be 12.5 kDa. Three concentrations of two hit compounds – L1 (50, 25, 0 μM) and T5 (50, 12.5, 0 μM) – combined with 8 peptide concentrations were assayed; the ATP concentration was 250 μM for the peptide-competitive assay. The DTT concentration was 250 μM; the amount of the IGF1R enzyme was 2 ng per well. In order to build the plot, the range of the used peptide concentrations was narrowed to 5 points to get a linear fit. Prior to ATP addition, the reaction mixture with compounds was incubated for 20 min at 27°C in all the experiments.

RESULTS AND DISCUSSION

Virtual screening – Target-based selection

The general concept of this study was to implement the "smart screening" strategy relying on the iterative physical screening of small, focused compound libraries selected from the vast off-the-shelf collection of ~1.9 million compounds at Enamine (www.enamine.net). The selection was based on ligand- and target-based virtual screening supplemented with knowledge of the published data on existing IGF1R inhibitors and the crystal structure of its kinase domain. Compounds containing potential toxicophoric and reactive structural fragments were removed using the medicinal chemistry filtering criteria described elsewhere [14]. Such an approach assists in the elaboration of new pharmacologically active compounds without resorting to random, large-scale screening of chemical diversity. The rationale was to deviate from the known IGF1R inhibitor chemotypes and from the paradigm of catalytic site binding and direct ATP competition. Several diverse *in silico* modeling approaches were used to generate mini-sets consisting of several hundred compounds each, which were subjected to experimental screening in a biochemical kinase assay. A total of approximately 4,000 compounds were screened as a result of this effort, including the "hit expansion" screens of active compound ("hit") analogs. Two of the approaches used, which were based on screening of 2,935 molecules, led

to the discovery of chemotypes with attractive properties and structural novelty (described below).

A series of allosteric inhibitors of the IGF1R kinase domain have recently been reported [15]. The mechanism of action of these compounds is based on their binding to the allosteric protein surface pocket, which does not spatially overlap with the catalytic site and is located in the vicinity of the kinase domain activation loop that is triple-phosphorylated upon enzyme activation [16, 17]. The identified compounds were characterized by moderate potency; however, some of them exhibited up to a tenfold selectivity for IGF1R versus InsR. Based on these results, we concluded that the binding site mentioned above has relevance for designing selective inhibitors of IGF1R. In order to design a IGF1R inhibitor screening set, we created the pharmacophore model (Fig.1) of interaction between the compound series mentioned above and the allosteric site using the available X-ray structure of the IGF1R kinase domain (PDB code 3LWO). The model included a H-donor, a H-acceptor, an aromatic/pseudo-aromatic ring, as well as any group distanced from the main molecular cluster.

Hydrogen bonding with the carboxyl group of Val1063 is one of the key determinants of binding at the allosteric site. The candidate binder molecule must contain a fragment identifiable as a strong hydrogen bond donor. Ambiguities in the definitions of such donors in various commercially available chemistry search programs led us to establish internal definition criteria for it.

All chemical compounds containing strong hydrogen bond donors were selected from the available compound database of approximately 1,900,000 entries (www.enamine.net) for further filtering. Those included all aliphatic amines, including tertiary amines (which are capable of becoming hydrogen bond donors upon protonation), as well as all other compounds with non-amide and non-sulfonamide NH groups (which were selected using the following SMART string: ([#1][#7;H1]([!$([#6,#16;X3,X4]=[O])])[!$([#6,#16;X3,X4]=[O])])). All compounds lacking an aromatic ring or an H-acceptor were subsequently removed from the selection. The resulting reduced database (approximately 400,000 compounds) has been further filtered to comply with the created pharmacophore model. All degrees of freedom were allowed for the rotatable bonds, and the additional "forbidden volume" rule was imposed on the protein atoms. Upon processing of the starting database according to the rules described above, 42,031 compounds strictly corresponding to the model criteria were identified. This final filtered set was subjected to a molecular docking study.

Docking was done with the flexible ligand and fixed receptor model, using a systematic docking algorithm (SDOCK+), which demonstrates sufficient ability to

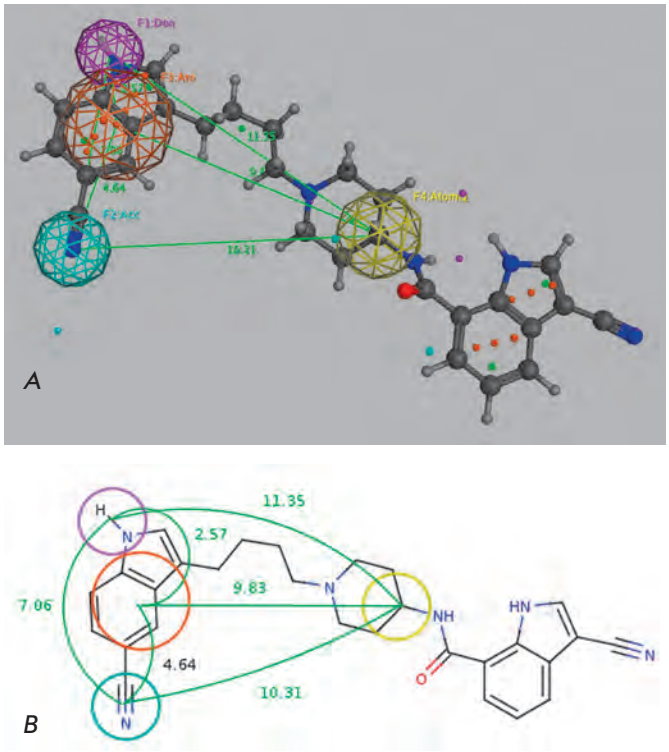


Fig. 1. The pharmacophore model used for virtual library filtering. **A** — Pharmacophore model mapping of IGF1R inhibitor 3-cyano-N-{1-[4-(5-cyano-1H-indol-3-yl)butyl] piperidin-4-yl}-1H-indole-7-carboxamide derived from the ligand orientation in the crystal structure. The IGF1R inhibitor is shown in ball-stick representation. **B** — The generated pharmacophore model is shown with its inter-feature distance constraints. Magenta — hydrogen bond donor; blue — hydrogen bond acceptor; orange — aromatic ring; yellow — any heavy atom; green — distances between the centers of the pharmacophore groups

reproduce ligand conformations with a minimal root-mean-square deviation (RMSD) with regard to the crystallographic data [18]. The maximum number of computational steps was set at 300, and the 20 best complexes (based on internal QXP scoring functions) were retained for analysis. The binding site model was formed based on the available X-ray data for the complex (3LWO). Amino acid residues with at least some atoms within a 1.0-nm radius around the initial inhibitor were taken into account when designing the binding-site model.

Post-docking processing and analysis of the results were performed according to the general logic of the pharmacophore model, which incorporates the key determinants of ligand-site binding strength. The following main geometrical filters were used: hydrogen bonding with Val1063, stacking with Met1054 and Met1079, as well as the secondary filters – electrostatic inter-

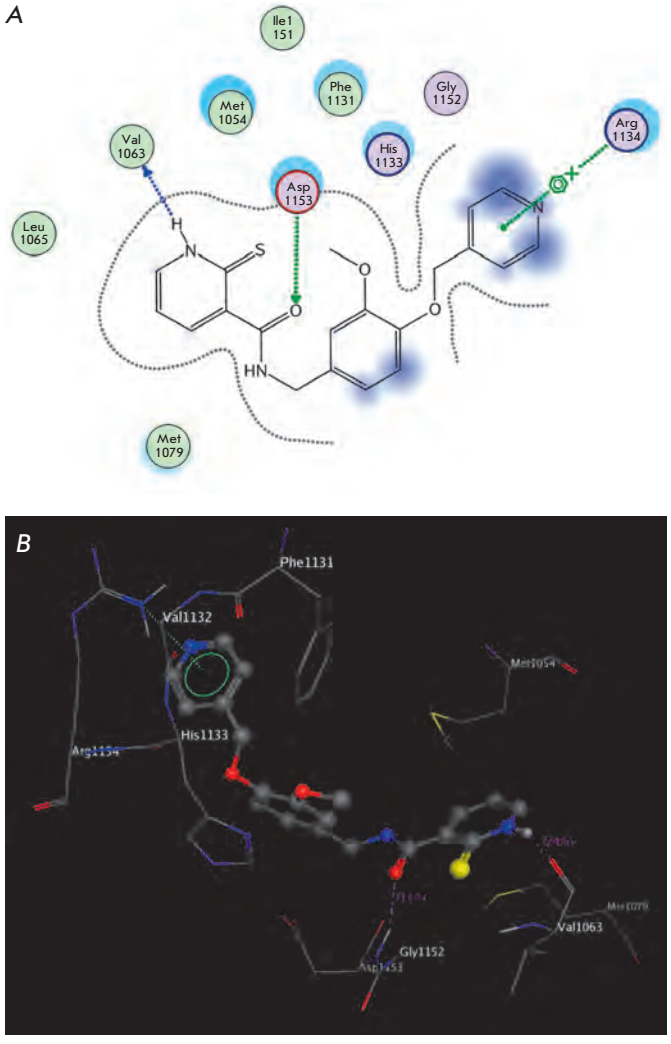


Fig. 2. Key interactions of the T2 compound with the IGFR1 binding site model (**A** — two-dimensional diagram showing the key interactions; **B** — ligand-target complex obtained by docking)

actions with Lys1033 and/or formation of hydrogen bonds with Asp1153 and/or Glu1050, Arg1134 (PDB code 3LW0). The main filters, as well as one or several secondary ones, were always used for selecting compounds. Visual inspection of the automated filtering output was conducted to ensure overall correspondence of the filtering rules to the model. The resulting 1,746 compounds designated as the T(target)-type selection set were submitted to high throughput screening as described below. The interaction between the 1,2-dihydropyridine-2-thione derivative (compound T2) and the allosteric binding site model is illustrated in Fig. 2 as an example. The compound meets all the basic requirements, and three additional interactions, namely, hydrogen bonding/stacking with Arg1134, His1133, and Asp1154, are possible as well. We consider these factors to be sufficient for inhibitory activity according to the mechanism postulated above.

Virtual screening – Ligand-based selection

Another productive approach for discovering IGF1R inhibitors was the "ligand-based," rather than the "target-based," one; it relied on the published data on IGF1R inhibitors discovered by Levitzky's group [19]. Since these compounds have been reported to be ATP-noncompetitive and some of them have exhibited substantial selectivity for IGF1R versus InsR kinase, we considered them an attractive starting point for exploring Enamine's collection in the search for structurally distant novel analogs. A SAR analysis of these active compounds, some of which are shown in Fig. 3, allowed us to identify a number of potentially preferable structural features. In particular, the actives contained 2- or 3-substituted benzene rings linked by saturated NH-CH_2 or CH_2-N-CH_2 linkers; further elongation of the linker by an additional atom decreased potency. At least one group with one hydrogen bond acceptor atom (N or O) must be located at the *para*- and/or *meta*-positions of the linked benzene rings to ensure activity. It was evident that average potency declined for the series: catechols > salicylic acid derivatives > benzodioxols. We hypothesized that the acceptor atoms directly linked to the rings at the *para*- and/or *meta*-positions were the most efficient pharmacophore groups that could be freely rotated to effectively bind to the target site. Additional hydrogen acceptor atoms seemed to provide higher potencies, and a similar level of IGF1R potency was achieved with the H-bond acceptor located either in the condensed aromatic rings or in aliphatic substituents (see Fig. 3A,B,C). Acylation of *para*-/*meta*-hydroxy groups (see Fig. 3A,D) did not significantly change the activity but seemed to have increased selectivity against the insulin kinase receptor and SRC kinase. Fully substituted benzodioxol compounds without hydrogen donors were also active.

These observations were combined in the Markush formula (Fig. 3G). The proposed structure contained at least two 5- or 6- atom aromatic systems, with at least one R-group from the ones listed below at positions 3 and/or 4 of the aromatic system. The R-groups present in the inhibitors described in the literature (i.e. *O*- and *N*-containing substituents), as well as fluoro- and α-fluoroalkyl, were selected as potential H-acceptors. The R-groups were allowed to incorporate the rings. One- to three-atom linkers formed by any nonring bonds (single, double, triple or aromatic) were used to link the aromatic rings. All the atoms in the rings and the linker were set to "any element except hydrogen" type during the database search. A search of Enamine's ~1.9 million

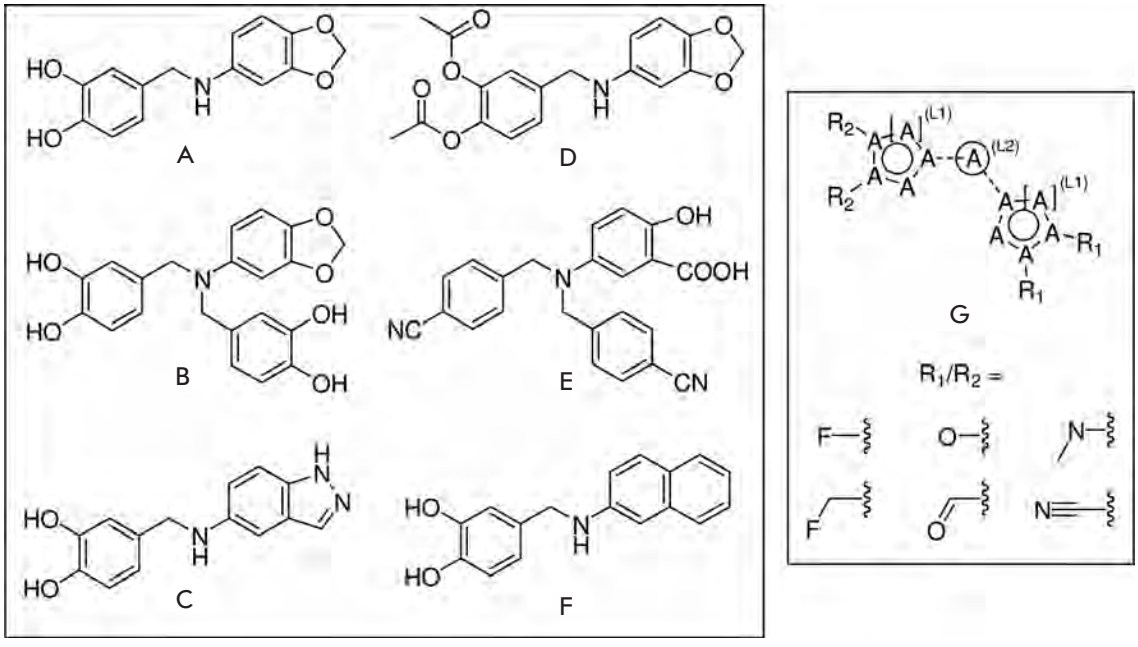


Fig. 3. Exemplary IGF1R inhibitors used as a basis of the L-type selection (*A*–*F*) and the corresponding Markush formula (*G*) for analog search

compounds collection *in silico* using the Instant JChem software resulted in the identification of 607 final compounds for HTS selected from 1,327 Markush-compliant compounds after application of the set of medicinal chemistry filters discussed above [14] and setting cut-offs for logP and logS to < 5. This screening selection was designated as the L(ligand)-type compound set.

High throughput screening and dose-response measurements

Screening of the T-type (1647 compounds) and L-type (607 compounds) sets, which were selected as described above, was performed using the commercially available biochemical IGF1R ADP-Glo kinase assay system (Promega Corp.). This assay utilizes a recombinant intracellular kinase domain fragment of IGF1R and is based on quantitation of adenosine diphosphate (ADP), a universal product of any kinase reaction, via enzymatic conversion of ADP to ATP, followed by detection of the luciferase-based luminescent signal [20]. Prior to performing HTS, the assay was validated for enzyme inhibition using the known IGF1R inhibitors — diarylurea derivative PQ401 [21] and tyrphostin AG538 [22], as well as the pan-kinase inhibitor staurosporine. The dose-response curves for all the reference inhibitors were in agreement with the data in the literature. In addition, performance of the HTS assay was tested for day-to-day and plate-to-plate reliability and reproducibility according to the standard HTS guidelines outlined in the Experimental section.

All the compounds exhibiting statistically significant inhibitory activity in the primary high-throughput screening runs ("screening hits") were re-tested separately at least once. Structural analogs of the confirmed hits were identified in Enamine's compound repository by chemoinformatics searches, and the resulting sets were additionally screened in the same assay ("hit expansion" screening) as shown in Fig. 4. Hit expansion was done by selecting all the nearest structural relatives of the actives from the collection, whereby all constituent groups in the molecules were subjected to structural variability where possible (sub-structure search). For the T-type compounds, conservation of the key pharmacophores depicted in Fig. 1 was imposed as an additional condition.

Screening of the T-type compound sets in the IGF1R ADP-Glo assay resulted in the identification of 3 confirmed active hits; the fourth hit (L4) was identified after the "hit expansion" screening. Screening of L-type compound sets led to the identification of 3 confirmed hits — L1, L2, and L3; the fourth hit (L4) was also identified after the "hit expansion" screening. These inhibitors were selected from all detected primary screening hits for additional experimental characterization. The selection was based on their estimated potency, reproducibility of the inhibition, and attractive chemical features. In particular, chemical tractability of the inhibitors is facilitated by their structural novelty, potential for synthetic improvements, as well as the absence of undesirable functionalities that might hinder further development of the compound.

Selectivity of the inhibitors

Selectivity for the IGF1R of eight hits identified as the result of the HTS campaign and follow-up hit expansion screenings were tested against the insulin receptor

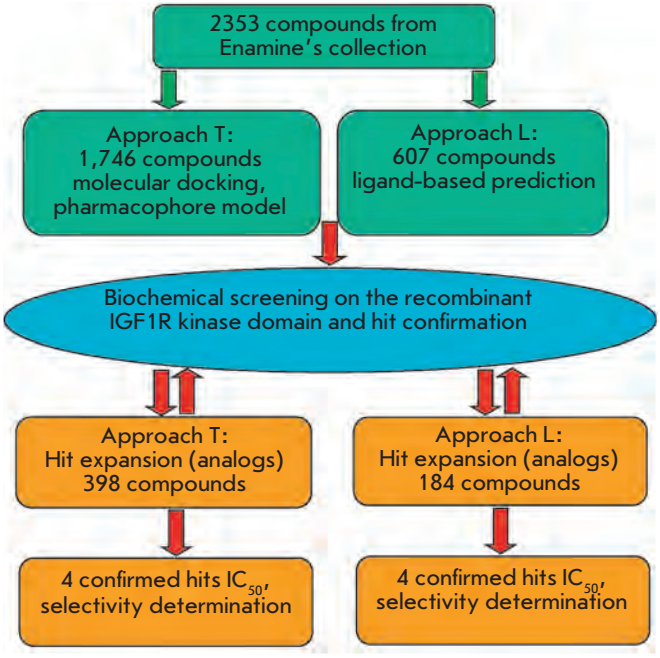


Fig. 4. Flowchart of the high-throughput screening procedure

(InsR) kinase, as well as the tyrosine kinases Met, Syk, and BTK. While InsR is the closest evolutionary relative of IGF1R, the other 3 tyrosine kinases are more distantly related receptor-type (Met) or cytoplasmic (Syk, BTK) tyrosine kinases. When comparing IGF1R and InsR, the IC_{50} values were measured for the inhibitors (Table 1), while the inhibition of the remaining kinases was evaluated at a single concentration point. All the kinase assays were run under experimental conditions similar to those of the IGF1R assay, with regard to the kinetic parameters of kinase reactions, and were read using the same commercial ADP detection system ADP-Glo (Promega Corporation) in order to ensure maximum uniformity for comparing the inhibition degrees. The small kinase panel used in this study cannot provide a comprehensive profile of the kinase selectivity of the tested inhibitors; however, it offers a general indication of the selectivity for the IGF1R target within the most closely evolutionarily related tyrosine kinase subfamily of over 500 human protein kinases. The data (Table 1) indicates limited selectivity of the compounds between IGF1R and InsR kinases, with some of the compounds being essentially nonselective (L1, L3, L4), while the others reproducibly exhibited 1.5-4-fold selectivity for IGF1R versus InsR (L2, T2, T4). Interestingly, compounds T1 and T3 exhibited a substantially stronger (5-10-fold) inhibition of InsR versus IGF1R in our experiments. This selectivity was similar or higher than that of almost all the known small molecule inhibitors of IGF1R demonstrated in a

Table 1. Inhibition of IGF1R and InsR (IC_{50}) by L- and T-series hit compounds.

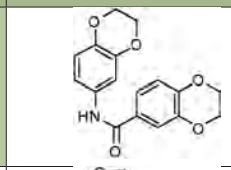
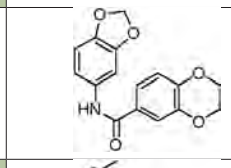
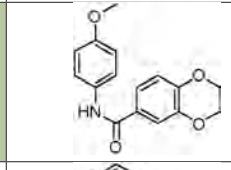
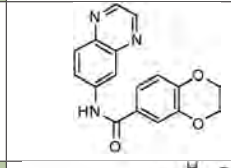
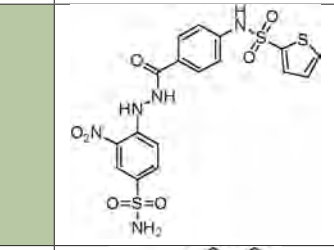
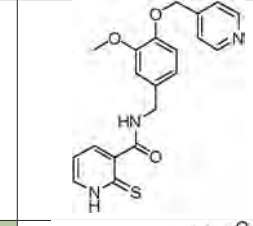
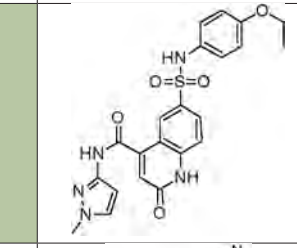
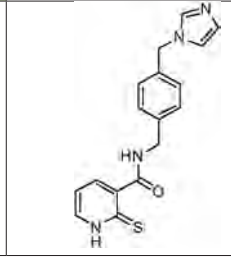
Compound	Structure	IC_{50}, μM	
		IGF1R	InsR
L1		18	22
L2		25	100
L3		26	29
L4		25	30
T1		~100	20
T2		18	30
T3		~100	10
T4		7	10

Table 2. Inhibition of Met, Syk and BTK kinases by L- and T-series hit compounds.

Compound*	Activity Met, %	±SD	Activity Syk, %	±SD	Activity Btk, %	±SD
L1	102	3	109	16	98	10
L2	63	2	48	5	106	6
L3	77	4	88	4	106	8
L4	60	2	53	9	113	11
T1	78	2	84	10	92	11
T2	62	4	72	13	83	8
T3	98	3	115	12	81	8
T4	75	3	115	19	82	5

*Concentration of compounds is 40 μM.

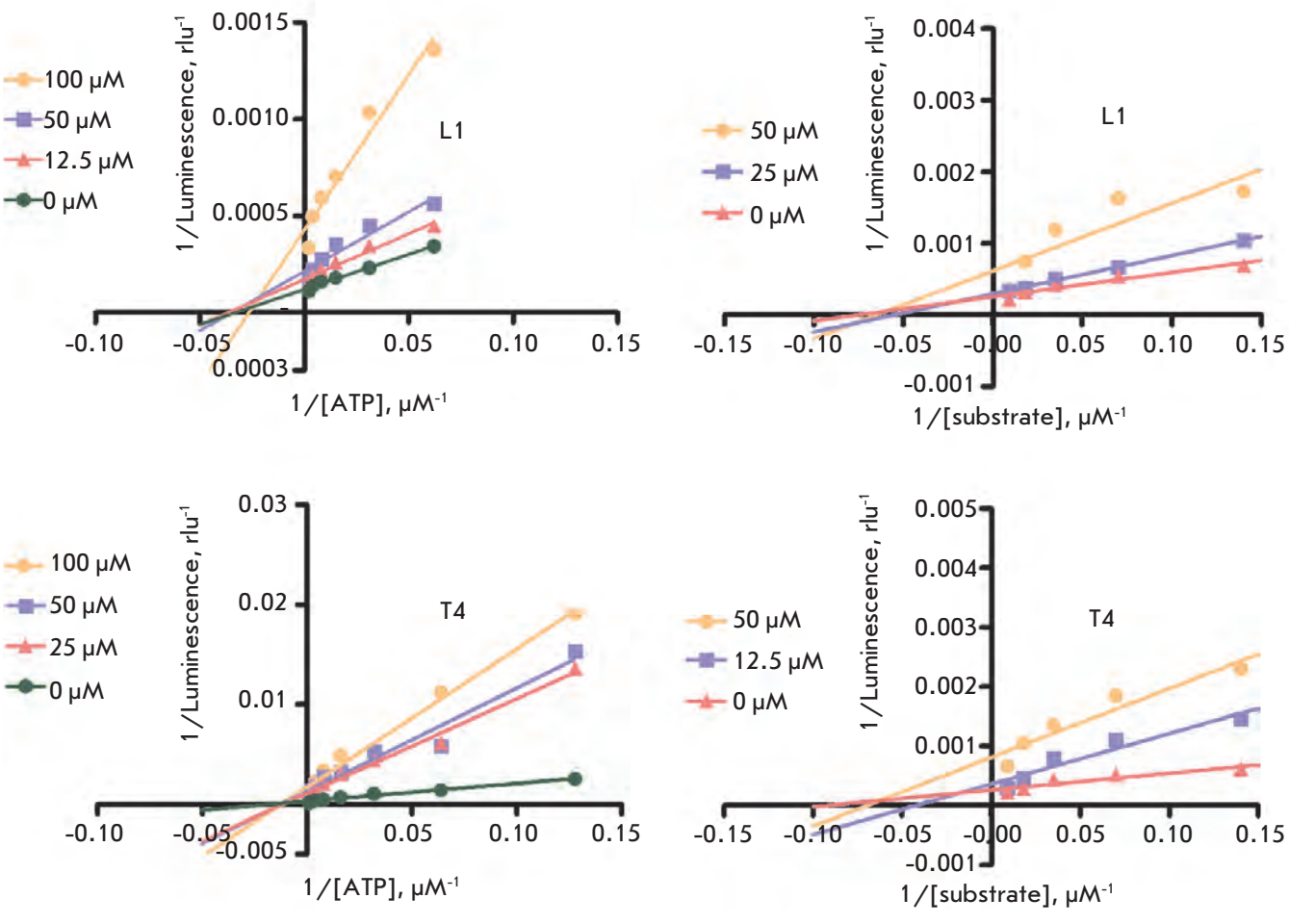


Fig. 5. Lineweaver-Burk plots for compounds L1 and T4

valid biochemical assay [23], apparently reflecting a very high degree of similarity between the receptors at their catalytic sites and in their close vicinity. The single concentration data on the inhibition of the more distantly related tyrosine kinases Met, BTK, and Syk (Table 2) suggests no inhibition (compounds L1, T3) or weak inhibition for some of these kinases, with estimated IC_{50} values in excess of 50 μM.

Mechanistic kinetic studies

Two arbitrarily selected compounds representing both series (L1 and T4) were used in the experiments of competitive inhibition kinetics with IGF1R kinase to investigate the inhibition mechanism (Fig. 5). The Lineweaver-Burk plot analysis indicates that both compounds exhibit non-competitive inhibition with regard to both ATP and the substrate. This experimental conclusion is consistent with the rationales used for the virtual selection of compounds for HTS and suggests an allosteric inhibition mode of IGF1R kinase by both the L- and T-type compounds. The binding site for T-type compounds is likely to align with the allosteric site defined for the prototypic indolealkylamines that were used to establish our pharmacophore model [15] and is spatially separated from the enzyme catalytic site. In the case of L-type compounds, localization of their putative binding site(s) on the kinase domain fragment is unclear. Some prototypic compounds used for our modeling have also been reported to be noncompetitive with ATP [19]; however, no substrate competition or extensive molecular modeling data are available. Due to the generally higher variability of the kinase domain regions distant from the conserved active sites, the allosteric mode of binding has more potential for fine-tuning the selectivity profiles of the inhibitors during their synthetic optimization stage.

Therefore, the set of IGF1R inhibitors described above meets the conventional requirements imposed on high-throughput screening hits: reproducibility and dose-dependence of the pharmacological response, acceptable potency of the molecular target inhibition (IC_{50}=10–25 μM) and selectivity versus related targets, absence of structural elements undesirable from a medicinal chemical perspective, novelty of the compounds, and availability of synthetic routes for their modification. In addition, the compounds exhibit allosteric inhibitor properties, which was one of the objectives of the project. In conclusion, the preliminary characterization of the two inhibitor series identified in the course of the screening campaign suggests that these compounds can serve as attractive starting points for a medicinal chemistry optimization towards novel, small molecule therapeutics targeting IGF1R. ●

The authors are grateful to Promega Corporation (Madison, WI, USA) for providing all screening kits and reagents for the high-throughput screening for IGF1R inhibitors as well as the mechanistic and selectivity studies of the discovered hit compounds. The authors would like to particularly express their gratitude and appreciation to John Watson and Tetyana Ruda (Promega Corp.) for their help and support during this work.

REFERENCES

1. Pollak M.N., Schernhammer E.S. and Hankinson S.E. // Nat. Rev. Cancer. 2004. V. 4. P. 505–518.
2. Pollak M. // Nat. Rev. Cancer. 2008. V. 8. P. 915–28.
3. Khandwala H.M., McCutcheon I.E., Flybjerg A. and Friend K.E. // Endocrine Reviews. 2000. V. 21. № 3. P. 215–244.
4. López-Calderero I., Sánchez Chávez E., García-Carbonero R. // Clin. Transl. Oncol. 2010. V. 12. P. 326–338.
5. Buck E. and Mulvihill M. // Expert Opin. Investig. Drugs. 2011. V. 20, № 4, P. 605-621.
6. Belfiore A., Frasca F., Pandini G., Sciacca L., and Vigneri R. // Endocrine Reviews, 2009, V. 30. № 6. P. 586–623.
7. Pollak M. // Nat. Rev. Cancer. 2012. V. 12, P. 159–169.
8. Buck E., Gokhale P.C., Koujak S., Brown E., Eyzaguirre A., Tao N., Rosenfeld-Franklin M., Lerner L., Chiu M.I., Wild R. et al. // Mol Cancer Ther. 2010. V. 9. № 10. P. 2652–2654.
9. Ludwig J.A., Lamhamedi-Cherradi S-E., Lee H-Y, Naing A. and Benjamin R. // Cancers. 2011. V. 3. № 3. P. 3029–3054.
10. Yee D. // JNCI J. Natl. Cancer. Inst. 2012. V. 104. № 13. 975–981.
11. McMartin C., Bohacek R. J. // Comput.-Aided Mol. Des. 1997. V. 11. № 4. P. 333–344.
12. Zhang J. Chung T.D., Oldenburg K.R. // J. Biomol. Screen. 1999. V.4. P. 67–73.
13. Lineweaver H., Burk B. // J. Am. Chem. Soc., 1934. V. 56. № 3. P. 658–666.
14. Chuprina A., Lukin O., Demoiseaux R., Buzko A., Shivanyuk A. J. // Chem. Inf. Model. 2010. V. 50. P. 470–479.
15. Heinrich T., Grodler U., Bottcher H., Blaukat A. and Shutes A. // ACS Med. Chem. Lett. 2010. V. 1. P. V. 199–203.
16. Munshi S., Kornienko M., Hall D.L., Reid J.C., Waxman L., Stirdivant S.M., Darke P.L., Kuo L.C. // J. Biol. Chem. 2002. V. 277. P. 38797–38802.
17. Li W., Favelyukis S., Yang J., Zeng Y., Yu J., Gangjee A., Miller W.T. // Biochem. Pharmacol. 2004. V. 68. P. 145–154.
18. Warren G.L, Andrews C.W, Capelli A.M., Clarke B., LaLonde J., Lambert M.H., Lindvall M., Nevins N., Semus S.F., Senger S., et al. // J. Med. Chem. 2006. V. 49, № 20, P. 5912–5931.
19. Steiner L., Blum G., Friedmann Y., Levitzki A. // Eur. J. Pharmacol. 2007. V. 562. № 1–2. P. 1–11.
20. Li H., Totoritis R.D., Lor L.A., Schwartz B., Caprioli, P., Jurewicz A.J. and Zhang, G. // Assay Drug Dev. Technol. 2009. V. 7. № 6. P. 598–605.
21. Gable K.L., Maddux B.A., Penaranda C., Zavodovskaya M., Campbell M.J., Lobo M., Robinson L., Schow S., Kerner J.A., Goldfine I.D., et al. // Mol. Cancer Ther. 2006. V. 5, № 4. P. 1079–1086.
22. Blum G., Gazit A., Levitzki A. // Biochemistry. 2000. V. 39. № 51. P. 15705–15712.
23. Chene P., Hau J.-C., Blechschmidt A., Fontana P., Bohn J., Zimmermann C., De Pover A., Erdmann D. // Open Enz. Inhib. J. 2010. V. 3 № 1. P. 27–37.

GENERAL RULES

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

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

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

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

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

MANUSCRIPT FORMATTING

The manuscript should be formatted in the following manner:

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

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

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

RECOMMENDATIONS ON THE TYPING AND FORMATTING OF THE TEXT

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

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

GUIDENESS FOR ILLUSTRATIONS

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

REVIEW AND PREPARATION OF THE MANUSCRIPT FOR PRINT AND PUBLICATION

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

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

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

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

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

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

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

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

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

FORMAT OF REFERENCES

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

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

For periodicals: the last name and initials of the author, title of the journal, year in which the work was published, volume, issue, first and last page of the article. Must specify the name of the first 10 authors. Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D., Platzer M., Howell G.R., Burrows C., Bird C.P., et al. // Nature. 2005. V. 434. № 7031. P. 325–337.

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) 930-87-07

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

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

NANOTECHNOLOGIES *in Russia*

Peer-review scientific journal

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

→ Russian edition: http://nanoru.ru

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

Issued with support from:

The Ministry of Education and Science of the Russian Federation

IN THE FRAME OF ST. PETERSBURG
INTERNATIONAL HEALTH FORUM

BIOINDUSTRY

OCTOBER 16–18, 2013

III INTERNATIONAL EXHIBITION AND CONFERENCE

A COMPLEX OF THE SPECIAL EXHIBITION ZONES,
SCIENCE, EDUCATION AND BUSINESS EVENTS AIMED
AT THE DEVELOPMENT OF BIOSPHERE AS A COMFORTABLE
HUMAN BEING HABITAT

Saint Petersburg
Exhibition complex LENEXPO
+7 812 240 40 40 add. 229, 279, 127

Organiser EXPOFORUM

www.bioindustry.ru

www.bio.expoforum.ru