JANUARY-MARCH 2020 VOL. 13 № 1 (48)

ISSN 2075-8251

ActaNaturae

INSTITUTE OF GENE BIOLOGY

30 years

ActaNaturae
JANUARY-MARCH 2021 VOL. 13 № 1 (48)
since april 2009, 4 times a year

Founders
Acta Naturae, Ltd,
National Research University
Higher School of Economics

Adress: 101000, Moscow, Myasnitskaya Ulitsa, 13, str. 4
Phone/Fax: +7 (495) 727 38 60
E-mail: actanaturae@gmail.com

Номер подписан в печать 30 марта 2021 г.
Тираж 100 экз. Цена свободная.
Отпечатано в типографии: НИУ ВШЭ,
г. Москва, Измайловское шоссе, 44, стр. 2

Indexed in Pubmed, Web of Science, Scopus, and RISC

Impact Factor: 1.360

CONTENTS

REVIEWS

RESEARCH ARTICLES

Super-Enhancers in the Regulation of Gene Transcription: General Aspects and Antitumor Targets

A. V. Bruter[1,2*], M. D. Rodionova[3], E. A. Varlamova[1,2], A. A. Shtil[1,2]
[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia
[2]Blokhin National Medical Research Center of Oncology, Moscow, 115478 Russia
[3]ITMO University, Saint-Petersburg, 197101 Russia
*E-mail: aleabruter@gmail.com
Received June 30, 2020; in final form, September 07, 2020
DOI: 10.32607/actanaturae.11067

ABSTRACT **Super-enhancers (genome elements that activate gene transcription) are DNA regions with an elevated concentration of transcriptional complexes. These multiprotein structures contain, among other components, the cyclin-dependent kinases 8 and 19. These and other transcriptional protein kinases are regarded as novel targets for pharmacological inhibition by antitumor drug candidates.**
KEYWORDS **transcription, super-enhancers, transcriptional protein kinases, targeted therapy, tumors.**

INTRODUCTION

The template synthesis of molecules (and gene transcription in particular) is one of the most essential processes in nature. This evolutionary conservative mechanism is found in all organisms, without exception: from viruses to higher mammals. Its biological role consists of transmitting and consolidating genetic information from the template macromolecule in the offspring. The cornerstone role of transcription is not limited to the "normal" processes, such as ontogenesis and phylogenesis, speciation, biodiversity, control of heredity, etc. Deepening our understanding of the molecular mechanisms of transcription allows us to grasp its fundamental significance in pathological processes. Today, it is impossible to interpret the etiology and pathogenesis of diseases without an analysis of the regulation of gene expression in a pathological site. It appears reasonable to assert that differential gene expression (changes in the set of functioning genes, activity (intensity) and temporal regulation of expression compared to the physiological pattern) defines the essence of a disease as a "transcriptional imbalance."

Our modern approach to therapy (targeted manipulation with specific transcription mechanisms) is currently rooted in this understanding. These mechanisms in mammalian cells are unusually diverse, interchangeable, and they remain insufficiently understood; in addition, a number of them remain unsusceptible to pharmacological regulation (the so-called non-druggable targets). Therefore, targeted "transcriptional therapy" is only in its first steps.

Which structural and functional elements of the transcription apparatus can be influenced to regulate gene expression? Which mechanisms should one regulate and how is this problem solved in terms of the spatial organization of transcription? Research focused on the regulatory regions of genes (promoters and enhancers) is necessary as, among the functions of these regions, are ensuring proper localization of multiprotein transcriptional complexes, transcription initiation, and regulation of the transcription rate. Since protein kinases are perhaps the most common targets in modern drug design, it is no coincidence that, among the various mechanisms of gene expression regulation, transcriptional kinases (a separate class of serine/threonine phosphotransferases) are emerging as the study object, the potential therapeutic target.

This review analyzes genomic elements where the presence of the transcriptional machinery is especially potent: super-enhancers and the proteins associated with them (transcription factors, cofactors, and protein kinases). We consider these elements as the structural

and functional units of the transcription apparatus, and as therapeutic targets in tumor cells.

SUPER-ENHANCERS: SPECIAL ENHANCERS!

Definition of the concept

The concept of super-enhancers was first formulated in a study focused on the regulation of gene expression in embryonic stem cells. Whyte *et al.* [1] disclosed a number of the traits of the regulatory regions of the genes whose active expression is associated with the maintenance of the undifferentiated pluripotent state (*Oct4*, *Sox2*, *Nanog*, *Klf4*, *Esrrb*, miR-290-295, etc.) These genomic regions differ from the conventional enhancers in terms of length and distance from the regulated gene, as well as in terms of the number and set of transcription factors associated with them. Oct4, Sox2, Nanog, Klf4, and Esrrb proved to be the prevailing transcription factors (the occupancy of the latter two factors in conventional enhancers is particularly different from that in super-enhancers). They are the key transcription factors that support, and can even induce, the pluripotent state of embryonic cells, as well as Med1, a component of the Mediator complex. The identified areas were named super-enhancers. An important feature of super-enhancers was discovered already in that first study: when the level of transcription factors in the cell changes (e.g., when the amount of Oct4 or the Mediator complex partially decreases), transcription of the corresponding genes stops, while transcription of the genes regulated by conventional enhancers changes insignificantly [1].

An attempt to provide a generalized definition of a "super-enhancer" makes it necessary to draw a distinction between these regions of the genome and conventional enhancers. This boundary turns out to be conditional (see below). The definition of super-enhancers is empirical and is based on two criteria. Super-enhancers include genomic regions with the following features: (1) regions containing extended (up to 12.5 kb) groups of enhancers; (2) regions with abnormally high binding of a certain set of transcription factors (these typically are transcription factors that are essential for the physiology of cells of a given type: Oct4, Sox2, and Nanog in embryonic stem cells; MyoD, a crucial tissue-specific transcription factor in muscle cells [2], etc.) and cofactors. In practice, these two structural criteria correlate with two functional criteria: a high expression level of the genes regulated by super-enhancers and an abrupt change in the expression level in response to small changes in the concentration of transcription factors [1, 3].

Although several thousand enhancers regulate the expression of thousands of genes, only a few hundred super-enhancers regulate the expression of the genes whose products are particularly important for cells of a given type [1, 4–6]. In addition, some super-enhancers function according to the positive feedback mechanism: super-enhancers regulate the expression of the genes encoding transcription factors that enhance the transcription of the genes regulated by super-enhancers. These genes include *Oct4*, *Sox2*, *Nanog*, *Klf4*, *Esrbb*, and *Prdm14* [6].

It is noteworthy that during evolution, super-enhancers were acquired by many of the genes playing a key role in cell biology, but not by the so-called housekeeping genes, which are characterized only by a consistently high expression level. Super-enhancers act as the end target of the main signal cascades more often than conventional enhancers do. In addition to an increased level of binding to the transcription factors (Oct4, Sox2, Nanog, etc.) that regulate the maintenance of the undifferentiated state of embryonic stem cells, increased binding of super-enhancers to the transcription factors closing the main signaling pathways (TCF3 (the WNT signaling pathway), SMAD3 (the TGF-β signaling pathway), STAT1 (the JAK-STAT signaling pathway), and STAT3 (LIF)) was also detected [7, 8].

Identification of super-enhancers in the genome

The most common method used to identify active super-enhancers is based on the characteristic epigenetic state inherent to active enhancers: monomethylation, instead of trimethylation, of lysine at position 4 of histone 3 (H3K4me1; it allows one to distinguish between enhancers' active promoters) and acetylation of lysine at position 27 of histone 3 (it "marks" active enhancers as opposed to inactive regulatory elements) [1, 3, 9]. At the first (experimental) stage, H3K27ac chromatin regions are immunoprecipitated, with subsequent sequencing of the DNA fragments associated with them. The obtained data are processed using bioinformatics (see [1, 4]). The DNA sequences found during sequencing are compared with the corresponding genome, and the regions that appear repeatedly (the so-called peaks) are identified. Peaks separated from each other by less than 12.5 kb are combined into single extended enhancers. The density of H3K27ac in the enriched sites is then normalized to the average density of H3K27ac for a given genomic region, and the enhancers are arranged in increasing order of enrichment. The resulting curve is characterized by an abrupt increase within the region of high enrichment in histone 3 with acetylated lysine-27. The enhancers contained in this area are referred to as super-enhancers; the criterion is that the enhancer is located on the plot (*Fig. 1*) to the right from the point

at which the derivative of the enrichment function equals 1 [10].

Along with enrichment in histone "marks," other molecular criteria for active transcription can be used to identify super-enhancers: sensitivity to DNase I, increased binding of transcription factors (Oct4, Sox2, Nanog, etc.), the presence of the activators Med1 and p300 [1, 3, 6]. The SEdb database [11] contains more than 300,000 super-enhancers, from 542 samples obtained from human cell lines. The differences in the number of super-enhancers between individual lines of non-tumor and tumor cells are specified. This database makes it possible to analyze, in detail, the nucleotide sequences of super-enhancers and identify binding sites for the transcription factors, polymorphisms, etc.

Super-enhancers identified by any of these methods consist of only a few single enhancers, and about 15% of the super-enhancers consist of just one enhancer [12]. Such an unclear empirical definition, which is also based on a conditional choice of distinction between conventional enhancers and super-enhancers, allows one to raise the following question: are super-enhancers actually a separate class of regulatory elements or are they a particularly effective type of enhancers?

FUNCTIONING OF SUPER-ENHANCERS

The difference between conventional enhancers and super-enhancers is clearly manifested in the nature of the dependence of the transcription activity ensured by the regulatory element and the number of transcription factors and cofactors associated with it. This dependence is linear for conventional enhancers, while, for super-enhancers, it acquires an "all or nothing" form [1, 13] resembling the dependences describing phase transitions in the framework of statistical thermodynamics. In practice, this manifests itself as a high sensitivity of super-enhancers to changes in conditions. Deletion of a small area or a reduced concentration of one of the cofactors (BRD4, CDK7) can completely inactivate the super-enhancer [4, 6, 14, 15]. A detailed analysis of these mechanisms is provided below.

A phase separation model was proposed to explain the regularities of super-enhancer function [13]. The high concentration of intensely interacting molecules gives rise to a membraneless organelle that is phase-separated from the rest of the nucleus. The model takes into account two numerical indicators: the number of molecules in a given volume (DNA, histones, transcription factors, and cofactors) (it is assumed that on average it is equal to 10 for a conventional enhancer and 50 for a super-enhancer) and the "valence" of these molecules (a number describing how many interactions are available to the molecule). In this model, transcriptional activity depends on

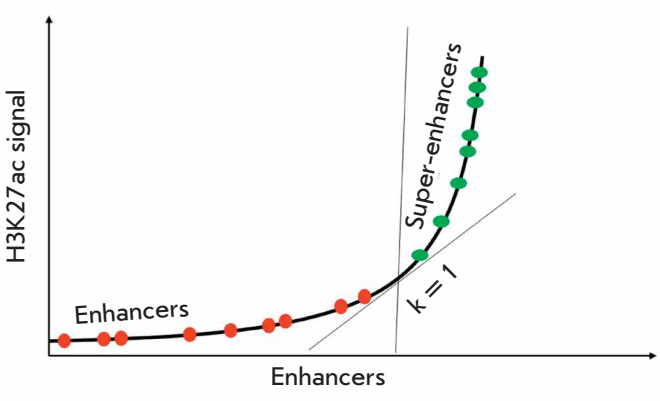


Fig. 1. Conventional enhancers and super-enhancers. The distribution of enhancers depending on the number of bound H3K27Ac molecules is shown. Adapted from [3, 10]

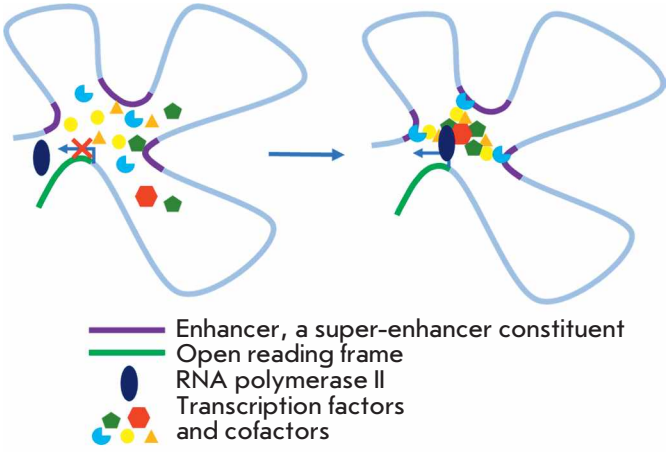


Fig. 2. The phase separation model of the structure and function of super-enhancers. Transcription factors and coactivators interact with different regions of the super-enhancer and with each other. The high intensity of these interactions leads to phase separation of DNA-protein transcription complexes and an abrupt transcription activation. *Left*: the situation before the start of interaction and separation. mRNA synthesis by RNA polymerase II does not occur. *Right*: after separation and transcriptional activation. Adapted from [14]

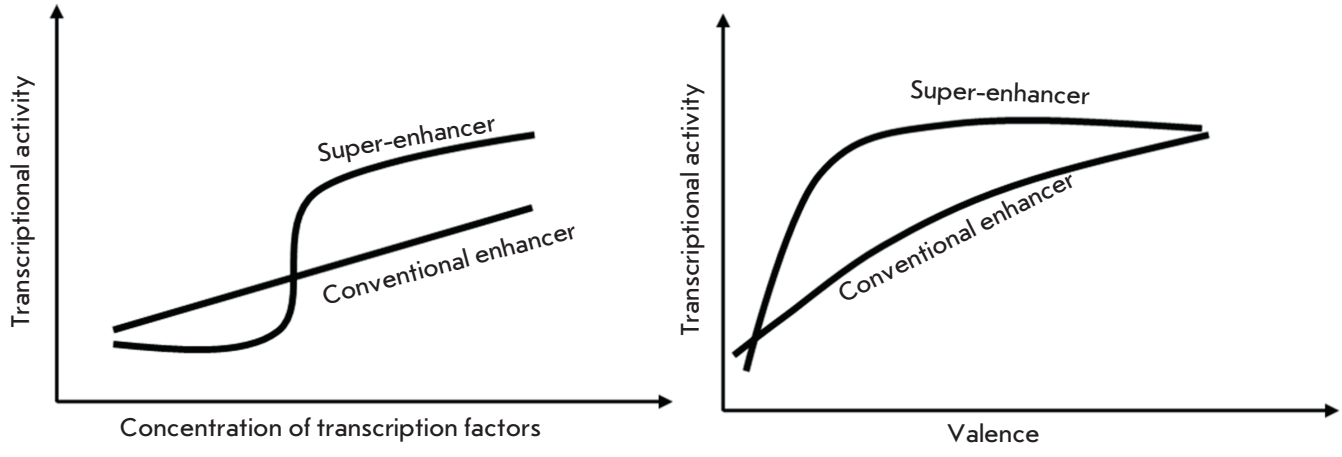


Fig. 3. *Left*: General dependence of transcriptional activity on the concentration of transcription factors for enhancers and super-enhancers. *Right*: Dependence of enhancer and super-enhancer activity on valence (i.e., the number of available intermolecular interactions according to the phase separation model). The regular enhancer is modeled by a system consisting of 10 molecules, whereas the super-enhancer is modeled by 50 molecules. Adapted from [14]

the percentage of molecules interacting with each other at a given time (*Figs. 2, 3*). The state of phase separation occurs when almost all molecules interact (i.e., the fraction of interacting molecules approaches unity). In this state, the transcriptional activity is at its maximum. The valence of the molecules in the system can grow (e.g., during chromatin remodeling and activation in the enhancer or super-enhancer region). Mathematical modeling has shown that the transcriptional activity of a conventional enhancer depends linearly on the valence of the system, and for a super-enhancer, at relatively low valence values, phase separation occurs and transcriptional activity increases abruptly almost to a maximum.

According to this model, valence decreases upon inhibition of a cofactor or deletion of the binding site. In the case of a super-enhancer, it causes a significant drop in transcriptional activity from the maximum to the minimal value.

In addition to DNA and protein molecules, the complex also contains enhancer RNAs (eRNAs): non-coding RNAs transcribed from the enhancers. Among them, there are short-lived short RNAs without poly(A) regions that can be transcribed in both directions, and longer ones with poly(A) regions (transcribed only in the 5' → 3' direction). eRNAs are involved in the organization of promoter–enhancer interactions: they increase the strength of the binding of transcription factors to DNA, recruit and activate cofactors, and shorten the transcriptional pause. Super-enhancers express eRNA at a higher level than conventional enhancers do; in addition, eRNAs are more often expressed from the super-enhancers rather than from conventional enhancers [10]. eRNAs can be involved in the activation of the expression of the corresponding gene but can also activate other genes, including those located on other chromosomes, thus spreading the impact of the enhancer (distant regulation of the genome) [16].

SUPER-ENHANCERS IN THE REGULATION OF NORMAL AND PATHOLOGICAL PROCESSES

Super-enhancers are much more likely to act as regulators of the key processes in normal cells and pathological processes compared to conventional enhancers [6]. The IgH 3'RR super-enhancer located in the 3'-regulatory region of the *IgH* locus on chromosome 14 of the human genome regulates recombination in B cells (in particular, V(D)J recombination in B1 cells [17] and isotype switching, depending on the external signal in B2 cells [18]). Another element that is important in this process is the super-enhancer of the *Aicda* locus. Enzymes belonging to the TET family and ensuring demethylation of this super-enhancer are required for isotype switching [19]. Conversion of adipocytes from brown fat to white fat is accompanied by the activation of a super-enhancer associated with the gene encoding the nuclear receptor PPARγ [20]. Furthermore, because of the activation of the super-enhancer, renin synthesis is induced by renal cells that do not synthesize renin under normal conditions.

A super-enhancer is activated only in the offspring of embryonic renin-producing cells [21]. CircRNA (circNfix), regulated by a super-enhancer that is specific to mature cardiomyocytes [22], precludes the division of mature cardiomyocytes, and its suppression improves tissue regeneration after an experimental myocardial infarction in mice.

Point mutations in the noncoding regions of the genome account for about 90% of all disease-related mutations (according to genome-wide association studies (GWAS)). Such mutations are more common in super-enhancers than they are in conventional enhancers. This conclusion was bolstered by comparing the super-enhancers in different types of cells from the same patient: in the abnormal focus and outside of it. Mutations (single-nucleotide polymorphisms, SNPs) in super-enhancers are associated with Alzheimer's, systemic lupus erythematosus, type 1 diabetes mellitus, etc. Thus, several SNPs have been found in the super-enhancer of the *BIN1* gene, whose increased expression is associated with the risk of developing Alzheimer's. In the case of type 1 diabetes mellitus, an increased amount of mutations was found in super-enhancers in T-helper cells. Polymorphisms associated with systemic lupus erythematosus were found to be concentrated in the super-enhancers of the key genes for B-cells [6].

Super-enhancers can also be epigenetically activated in response to external stimuli. Thus, during inflammation, activation of the transcription factor NF-κB in endothelial cells can lead to the formation of active super-enhancers and maintenance of a high expression of genes whose products promote the adhesion of leukocytes (*SELE*, *VCAM1*), as well as chemokine CCL2. Super-enhancers have been activated due to binding of the acetylated form of NF-κB to BRD4. No super-enhancers were activated upon BRD4 inhibition in [23].

Super enhancers in tumor cells

The high level of expression of oncogenes in malignant cells may be due to the emergence of a new super-enhancer. Accelerated proliferation of malignant cells is regulated by signaling cascades. The proliferation intensity of colon adenocarcinoma cells (HCT116 line) depends on the activation of the Wnt signaling pathway; in this lineage, a super-enhancer is activated at the *c-MYC* locus. Along with this, increased binding of the transcription factor TCF4, an effector of the Wnt cascade, to the super-enhancer was discovered. A similar regulation mechanism was encountered in the cells of estrogen-dependent breast cancer [7].

The emergence of super-enhancers in tumor cells occurs according to the same mechanisms as any changes in gene expression. Both genetic (chromosomal translocations [6], amplification [24–27], deletions, insertions [28], and point mutations) and epigenetic (activation of oncogen expression [4, 6, 11, 29–31] or reduced expression of anti-oncogenes [8, 31–33]) mechanisms can participate in malignant cell transformation. In the first variant, the emergence of new super-enhancers and the disappearance of previously existing ones is possible, as well as the transfer of potential oncogenes under the control of active super-enhancers that are unusual for them. Epigenetic regulation is represented by activation or deactivation of the corresponding super-enhancers.

The *c-MYC* locus, especially, frequently acquires super-enhancers during carcinogenesis. In multiple myeloma cells, the super-enhancer regulating the gene expression of the *igH* locus appears in the *c-Myc* locus via translocation [6]. In some patients with acute T-cell leukemia, reduplications were found in the non-coding region, consistent with the super-enhancer regulating the *c-Myc* gene. The amplified fragment ensures the Notch-dependent functioning of the super-enhancer [25]. The amplified region was associated with the proteins of the SWI/SNF complex, which remodels chromatin and is key in the proliferation of tumor cells [34]. Focal amplifications (copying of a small region of the genome, followed by the transferring of copies to arbitrary genomic regions) of super-enhancers in the 3'-regulatory region of the *c-Myc* gene were found in lung and endometrial tumors [26]. In a similar manner, focal amplification of a super-enhancer in the 3'-regulatory region of the *KLF4* gene and its increased expression were found in squamous cell carcinoma of the head and neck [27]. In T-cell leukemias, a small (2–12 bp) insertion forming a binding site for the Myb transcription factor mediates the formation of an active 8-kb super-enhancer, thus recruiting additional transcription factors [28] (*Fig. 4*).

An important genetic mechanism of malignant cell transformation is represented by a violation of the boundaries of topologically associating domains (TADs), which are chromosomal segments approximately 1 Mbp long that are transcriptionally isolated from each other. The fragments of one TAD interact with each other much more often than the fragments of different TADs. Division into these domains is an evolutionarily conserved process that probably arose to prevent the long-range interaction of enhancers and super-enhancers with "foreign" promoters. The boundaries between the TADs are the binding sites of the CTCF transcription repressor (CCGCGNGGNGGCAG) and the CTCF–cohesin protein complexes associated with them. Mutations in the genes encoding cohesin and CTCF, as well as in their DNA-binding sites, are

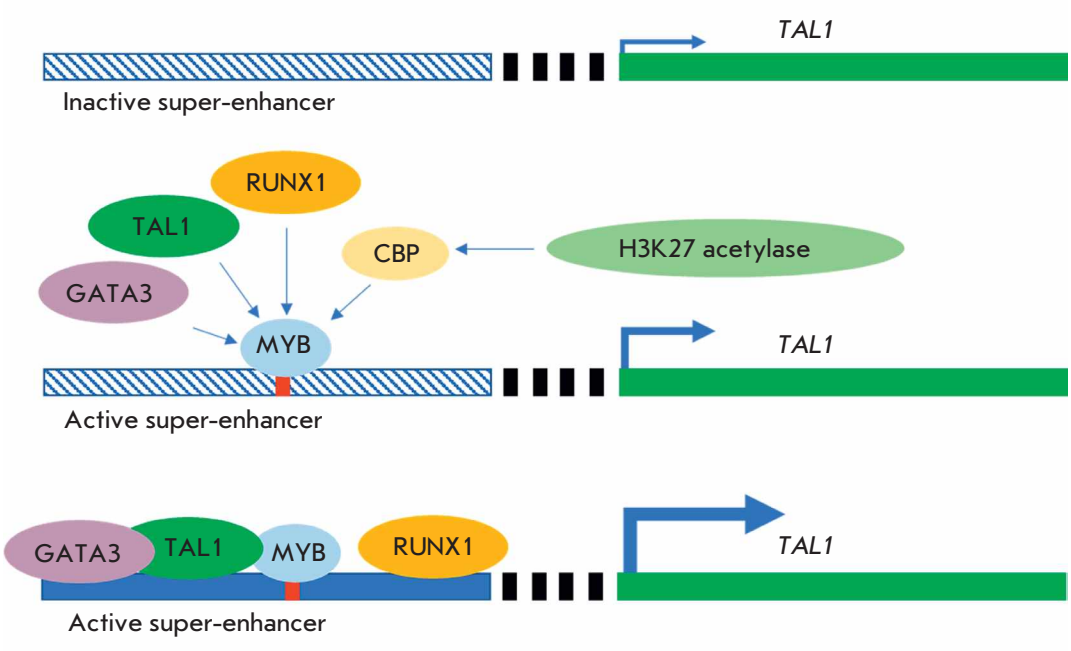


Fig. 4. Insertion of the Myb transcription factor binding site activates the super-enhancer-driven expression of the *TAL1* oncogene. The DNA-bound Myb recruits the cyclic AMP response element binding protein (CBP) and its partner, histone acetylase H3K2, to activate chromatin in the super-enhancer region. Additional transcription factors are also recruited

often present in transformed cells [9]. Small deletions were found at the TAD boundaries in the Jurkat cell line ($CD4^+8^+$ thymocytes). When reproducing these deletions in epithelial cells (HEK293 line) using the CRISPR/Cas9 system, a super-enhancer-dependent activation of the oncogenes *TAL1* and *LMO1* was detected [35].

Changes in the nucleotide sequences of super-enhancers may both affect their binding to proteins and also cause changes in the sequences and the number of eRNA molecules. Certain eRNAs associated with "oncogenic" super-enhancers have oncogenic properties themselves. These eRNAs are involved in the regulation of the key processes: proliferation, apoptosis, autophagy, epithelial–mesenchymal transition, and angiogenesis [16].

The bromodomain protein BRD4, which binds to acetylated lysine residues in histones (i.e., to active chromatin), plays an important role in the epigenetic regulation of super-enhancer activity. A study focused on diffuse large B-cell lymphomas showed that one-third of the BRD4 molecules in a cell are concentrated in super-enhancers. Expression of the corresponding genes was found to be very sensitive to the pharmacological inhibition of BRD4 [14]. A low-molecular-weight inhibitor of BRD4, compound JQ1, reduced the expression of super-enhancer-dependent genes, in particular, the *c-Myc* oncogene, in myeloma cells [4]. The high level of expression of *c-Myc* in colon cancer cells (HCT116 and DLD1 lines) is also supported by a super-enhancer. Knockdown of the *BRD4*, *MED12*, and *MED13/13L* genes decreased enhancer-dependent gene expression and inhibited the proliferation of colorectal cancer cells [29].

An unexpected variant of epigenetic activation of super-enhancers was discovered in the study of B-cell infection with the Epstein–Barr virus. In infected cells, this virus synthesizes its own transcription factors, EBNA2, 3A, 3C, and EBNA-LP, and activates some cellular ones (RelA, RelB). These transcription factors form the active super-enhancers that regulate the *c-Myc*, *MIR155*, *IKZF3*, and *Bcl-2* genes that are crucial to cell survival. The activation of super-enhancers is sensitive to BRD4 inhibition: it was blocked by the compound JQ1 [36].

Super-enhancers also regulate the differentiation status of tumor stem cells; this fact can be used in elaborating therapeutic strategies. For maintaining the pluripotent state of glioma stem cells, the ELOVL2 (elongation of the very long chain fatty acids protein 2) protein, whose expression in these cells is particularly high and is triggered by the epigenetic activation of the corresponding super-enhancer, is important. ELOVL2 plays a key role in the synthesis of polyunsaturated fatty acids (components of the plasma membrane) and is also involved in the signaling from the epidermal growth factor receptor (EGFR). Selective inactivation of the ELOVL2 super-enhancer by dCas9-KRAB

leads to a post-transcriptional decrease in the EGFR level [30]. Loss of functional activity by the B-cell transcriptional regulator Ikaros is associated with a poor prognosis in patients with B-cell acute lymphoblastic leukemia. This is because Ikaros is required for the terminal differentiation of rapidly proliferating B-cell progenitors. The "two-facedness" of Ikaros is rather interesting: it maintains the inactive state of chromatin in the regions of super-enhancers of the genes whose expression determines the undifferentiated state of B-cells, but it also maintains active chromatin in the super-enhancers of the genes whose products are important for differentiation [31].

This example illustrates a situation where the malignant potential of a cell is ensured by the inactivation of the expression of anti-oncogenes dependent on super-enhancers rather than by the high expression of oncogenes caused by super-enhancers. In this regard, therapeutic strategies aimed at reactivating anti-oncogene super-enhancers appear reasonable. The possibility of implementing this strategy is analyzed below.

SUPER-ENHANCERS AS A FOCUS OF THERAPEUTIC TARGETING

As mentioned above, super-enhancers make it possible to fundamentally alter the transcriptional program, even in response to a relatively weak stimulus. Malignant transformation is often associated with the emergence (formation) or activation of an existing, but non-functioning, super-enhancer. Therefore, super-enhancers and the associated proteins are gaining interest as targets for the development of anticancer drugs. It is hoped that a therapeutic effect will be achieved at relatively low concentrations of such drugs (see below).

Two classes of proteins associated with super-enhancers are considered as therapeutic targets: proteins with a bromodomain (primarily BRD4) and the cyclin-dependent protein kinases CDK4/6, CDK7, CDK8, and CDK12/13.

Proteins containing a bromodomain

Proteins carrying a conserved lysine-binding amino acid sequence, the bromodomain, ensure the functioning of super-enhancers by maintaining the active state of chromatin through interaction with the acetylated lysine residues in chromatin proteins. As a result of this interaction, transcription factors and RNA polymerase II are recruited to super-enhancers. Inhibition of proteins carrying the bromodomain can lead to chromatin inactivation.

BRD4 inhibitors (small-molecule compounds JQ1 and I-BET151) have shown encouraging results in preclinical models of acute myeloid leukemia and multiple myeloma. Tumor growth retardation, as well as suppression of *Myc* expression and downstream transcription programs, was observed [9]. A number of compounds that inhibit BRD2/3/4/T by competitive binding are currently undergoing clinical trials [37]. The ABBV-075 inhibitor (Mivebresib) was tested on 10 patients with acute myeloid leukemia resistant to standard therapy and/or recurrent forms of the disease; one of the patients achieved complete remission, the number of blast cells in bone marrow in four patients was reduced at least twofold, and good treatment tolerance was observed. Combination therapy is also promising [38].

Cyclin-dependent protein kinases

The cyclin-dependent kinases CDK4 and CDK6 play a key role in the phase change of the G1-S cell cycle. The CDK4/6 inhibitors palbociclib, ribociclib, and abemaciclib have been included in hormone-sensitive HER2-negative breast cancer protocols as monotherapy. As part of combination therapy, CDK4/6 inhibitors are undergoing clinical trials for other types of breast cancer [39]. Selective inhibitors of CDK4/6 have a cytostatic effect and cause death of Ewing sarcoma cells in culture and *in vivo*, and they also reduce the expression of a number of genes dependent on super-enhancers (in particular, cyclin D1) [40].

A special group of cyclin-dependent protein kinases does not participate in the regulation of cell cycle phases but functions as a structural and functional component of the transcription apparatus. In particular, such "transcriptional" protein kinases include CDK7, CDK8 and its paralog CDK19 (CDK8/19), as well as CDK9 and CDK12/13 [41]. CDK7 is a component of the TFIIH transcription-initiating complex; it mediates the phosphorylation of the C-terminal domain of RNA polymerase II and transcription initiation. CDK9 within the p-TEFb complex also regulates the transition to elongation by phosphorylation of the C-terminal domain of RNA polymerase II [41]. CDK12 and CDK13 directly activate mRNA elongation and processing [42].

THZ1, an inhibitor of CDK7 (and, to some extent, CDK9 and CDK12) [43, 44], reduces transcription in cells of various tissue origins; the transformed cells were found to be sensitive to low THZ1 concentrations. The compound suppressed the oncogenes associated with super-enhancers, in particular, at the *c-Myc* locus [45]. Clinical trials of a more selective inhibitor of CDK7, the SY5609 compound, were launched in 2020 [46].

Super-enhancer-dependent expression of the *RUNX1*, *MYB*, *TAL1*, and *GATA3* oncogenes decreases in Jurkat cells under the influence of THZ531, an inhibitor of CDK12 [47]. Since the first specific inhibitors of CDK12 have been synthesized recently, their

clinical trials are yet to be started. However, it has been shown in cell cultures and tumor models in mice that inhibition of CDK12 has a pronounced effect on osteosarcoma, liver, breast, and ovarian tumors, as well as neuroblastoma [48].

CDK8 plays a special role in transcription regulation. This serine/threonine protein kinase, in cooperation with cyclin C (CCNC), the MED12, and MED13 proteins, forms the regulatory CDK module of a crucial transcriptional complex: Mediator. The components of this complex are conserved in all eukaryotes. It is important to understand that CDK8/19, unlike other CDKs, does not regulate phase transitions in the cell cycle [49]. The main function (but not the only one) of CDK8/19 is to regulate the phosphorylation of the C-terminal domain of RNA polymerase II at the serine-2 and serine -5 residues of the heptapeptide repeat constituting this domain. This phosphorylation was shown in a cell-free system. In cells, this event is necessary at different stages of transcription (initiation, pause release, and elongation of the primary transcript); however, the role of CDK8/19 in this phosphorylation needs to be proved experimentally. In contrast to CDK7 and CDK9, which function on all promoters, CDK8/19 is involved only in the regulation of the activity of RNA polymerase II on actively transcribed genes (inducible genes and genes functioning in the development of the organism) [50–53]. The selectivity of expression activation indicates that CDK8/19 is one of the key mechanisms of transcriptional reprogramming. This unique feature has been the subject of extensive research in recent years.

Transcriptional reprogramming is not vital for an adult organism under homeostatic conditions; long-term inhibition of CDK8/19 has no phenotypic manifestations. Genetic (mediated by the Cre/Lox system) knockout of the *cdk8* gene also has no significant manifestations in adult mice [54]. However, reprogramming of transcription is necessary for the development of the organism: knockout of *cdk8* in mice is lethal at the preimplantation embryo stage [55], and null mutations in the genes encoding the cdk8 or ccnc proteins in *Drosophila melanogaster* leads to death at the late third instar larva and prepupal stages [56, 57].

Importantly, the *CDK8* gene knockout and pharmacological inhibition of kinase activity have different effects on the general patterns of gene expression, which indicates two fundamentally different mechanisms of CDK8/19 action: those dependent on and independent of kinase function.

Is there a connection between CDK8/19 and super-enhancers? In immunoprecipitation experiments, an increased presence of CDK8 in the regions of individual super-enhancers was detected. According to the RNA sequencing data, two-thirds of the genes whose expression is affected by CDK8 inhibition are the genes regulated by super-enhancers. Among super-enhancers whose relationship with CDK8 was established in both of the aforementioned experimental systems, there were super-enhancers of the genes encoding the transcription factors Nanog, Oct3/4, and SOX2, as well as a significant number of super-enhancers of the genes regulated by the Wnt signaling pathways [32].

SUPER-ENHANCERS AND CDK8/19: THE TARGETS OF ANTITUMOR ACTION

Transcriptional reprogramming is fundamental in the development of many pathological processes, especially tumor ones. Deregulation of CDK8/19 is often encountered in tumors in which CDK8 is involved in the activation of important signaling pathways mediated by Wnt/β-catenin [58], NF-ϰB [51], TGF-β [59], HIF1α [51], or the estrogen receptor [41] regulating the response to changes in the serum concentration [50]. CDK8 was found to be an oncoprotein associated with the development of colorectal cancer [47], tumors of the pancreas [60] and mammary glands [52, 61–63], and melanoma [64]. CDK8 is responsible for the phenotype of cancer stem cells [65].

Since CDK8/19 inhibition is practically safe in an adult organism, it is promising to use CDK8/19 as therapeutic targets [41, 66, 67]. Compounds belonging to various chemical classes acting as a platform for the proposed pharmacological blockers of CDK8/19 kinase activity and the so-called degraders for complete elimination of these proteins (overcoming of the kinase-independent function) are being intensively studied [68, 69]. The issues related to inhibitor selectivity have been discussed in reviews and original studies [53, 70–73].

Cortistatin A is a relatively selective inhibitor of CDK8/19 kinase activity. Inhibition of these protein kinases by cortistatin A in acute myeloid leukemia cells (MOLM-14 line) increased the expression of the "antitumor" genes controlled by super-enhancers (*CEBPA*, *IRF8*, *IRF1,* and *ETV6*), which slowed down cell proliferation. Meanwhile, the expression of 20% of the genes associated with super-enhancers increased, while only a 3% increase in expression was observed for the genes with the conventional enhancers. CDK8 was found to be associated with the super-enhancers of all activated genes (versus 67% of activated genes with the conventional enhancers). There were only three inhibited genes regulated by the super-enhancers (1% of all the genes regulated by super-enhancers). These ratios allow [33] one to infer that the proteins associated with super-enhancers are the direct targets of cortistatin A in MOLM-14 cells.

Treating these cells with the I-BET151 compound, an inhibitor of BRD4, reduced the expression of the genes regulated by super-enhancers, although the result (an antitumor effect) was the same as that observed after exposure to cortistatin A. Proliferation of tumor cells is likely to depend on the "dosage" of gene expression determined by super-enhancers. It is noteworthy that the effects were not summarized for I-BET151 and cortistatin A used together, and that the changes in the gene expression profile were in full alignment with those caused by I-BET151 alone. BRD4 is probably required for transcription activation in response to cortistatin A [33].

Acute myeloid leukemia cells are sometimes characterized by constitutive activation of the JAK-STAT signaling pathway. The STAT1 transcription factor is one of the main targets of CDK8 kinase activity. The content of the phosphorylated (transcriptionally competent) form pSTAT1S727 is increased in the super-enhancer regions [8]. It turns out that CDK8-dependent phosphorylation of STAT1 is required for rapid proliferation of leukemia cells. Exposure to cortistatin A led to a slowdown in the proliferation and activation of the super-enhancer-dependent expression of the GATA1, GATA2, and ID2 transcription factors mediating proliferation slowdown or cessation, as well as the activation of the megakaryocyte-specific PLEK, CFLAR, and UBASH3B factors. As a result, transition of cells from the stem state to the differentiated state and inhibition of proliferation were observed [8].

CDK8/19 inhibitors based on modified pyridines slowed the proliferation of colorectal cancer cells [32]. The gene expression patterns changed in the same fashion as when a number of super-enhancers were activated. Under the influence of CDK8/19 inhibitors, tumor cells went from the stem phenotype to a differentiated state. As mentioned above, this process is associated with the activation of super-enhancers. Activation of the *c-Myc* oncogene, which is also regulated by a super-enhancer, is consistent with the concept of increased super-enhancer-dependent expression upon CDK8 inhibition. However, despite the activation of *c-Myc*, the inhibitors were found to exhibit an overall moderate antitumor effect [32].

Kuuluvainen *et al.* [29] attempted to devise a way to selectively inactivate the super-enhancers ensuring high oncogene expression in colorectal cancer cells. The reduction in the CDK8 level by RNA interference led to an integral decrease in the expression of the super-enhancer-regulated genes, but no selective decline in the expression of the super-enhancer-regulated oncogenes was observed. This decrease was caused by knockdown of the *MED12* and *MED13/13L* genes [29].

In the examples discussed above, CDK8/19 inhibition did not affect the activated enhancers of oncogenes but led to the activation of the super-enhancers of the anti-oncogenes [8, 32, 33]. The antitumor effect is attributed to the restoration of a differentiated phenotype and slowdown in cell proliferation. Hence, the use of CDK8/19 inhibitors in the treatment of certain tumor types should be considered a super-enhancer-mediated restoration of normal gene expression in malignant cells. Meanwhile, the pathological process is not limited to transcription disorders: post-transcriptional and post-translational events also play a role.

Early stages of clinical trials of CDK8/19 inhibitors are currently underway. For instance, SEL120 compounds are being tested as candidate drugs against acute myeloid leukemia, and BCD-115 are studied for possible treatment of HER2-negative estrogen-dependent breast cancer. Trial identifiers, as well as the analysis of the causes of the toxicity of CDK8/19 inhibitors, were presented in [71].

CONCLUSION. SUPER-ENHANCERS AS THERAPEUTICALLY SIGNIFICANT ELEMENTS OF THE GENOME

Detailed studies of the structural and functional features of genome organization have made it possible to formulate the concept of super-enhancers as regions with an increased content of transcription complexes. It is not surprising that these regions are important in pathogenesis: the molecular mechanisms of diseases are associated, in one way or another, with dysregulation of gene transcription. Super-enhancers acquire a special role in tumor biology: uncontrolled proliferation of transformed cells and their evasion of therapeutic action (chemotherapy and radiation therapy) are caused by both transcription activation and by adaptive changes in their gene expression profile. Consequently, super-enhancers (the DNA regions carrying multiprotein transcriptional "machines") become targets of antitumor action.

The question related to the prospects of the low-molecular-weight chemical modulators of transcriptional CDKs in tumor therapy is especially important. The effectiveness of the first CDK inhibitors turned out to be insufficient, and general resorptive toxicity was high. Subsequently, more selective inhibitors of individual transcriptional protein kinases were obtained: THZ1 for CDK7, THZ531 for CDK12/13, and palbociclib and ribociclib for CDK4/6. These compounds have a pronounced antitumor effect in clinical situations and are becoming parts of treatment regimens [74, 75].

CDK8/19s are of interest as a unique target: the special selectivity of transcription reprogramming offers a chance to replace the currently used toxic drugs

with well-tolerated agents inhibiting this mechanism. Although occurring in all age groups, acute myeloid leukemia is especially common in patients over 60 years of age. The currently used treatment regimens are difficult to tolerate due to the cardio- and myelotoxicity; the likelihood of early relapse is high (within the first year) [76, 77]. In our experiments, the selective CDK8/19 inhibitor senexin B caused the death of acute myeloid leukemia cells (line MV-4-11) when used at significantly lower concentrations than cytosar, one of the main chemotherapy drugs used for treating this disease. Senexin B produced the indicated effect at concentrations that were non-toxic to non-tumor cells. In the culture of chronic myeloid leukemia cells, senexin B increased the antitumor effect of targeted inhibitors of chimeric tyrosine kinase Bcr-ABL [78], thus broadening the possibilities of CDK8/19 inhibition in the therapy of blood cancers. The outcomes with chemotherapy for colorectal cancer (especially metastatic disease [79, 80]) remain unsatisfactory; therefore, the results of studies focused on this tumor and demonstrating the effectiveness of an inactivation of the Mediator complex components seem rather promising [29, 32, 81].

While it may be difficult to interpret super-enhancers as special "independent" regulatory elements of the genome from a general biology point of view, their practical importance as "accumulators" of transcription complexes for studying a pathogenesis and developing personalized therapy seems undeniable. This strategy involves identifying the role of a specific transcriptional mechanism in the patient (the transcriptional "portrait") and targeting the established mechanism. ●

This study was supported by the Megagrant (Agreement №14.W03.31.0020 between the Ministry of Science and Education of the Russian Federation and Institute of Gene Biology, Russian Academy of Sciences).

REFERENCES

1. Whyte W., Orlando D., Hnisz D., Abraham B., Lin C., Kagey M., Rahl P., Lee T., Young R. // Cell. 2013. V. 153. № 2. P. 307–319.
2. Peng X., So K., He L., Zhao Y., Zhou J., Li Y., Yao M., Xu B., Zhang S., Yao H., et al. // Nucl. Acids Res. 2017. V. 45. № 15. P. 8785–8805.
3. Tang F., Yang Z., Tan Y., Li Y. // NPJ Precis. Oncol. 2020. V. 4. № 1. P. 2.
4. Lovén J., Hoke H., Lin C., Lau A., Orlando D., Vakoc C., Bradner J., Lee T., Young R. // Cell. 2013. V. 153. № 2. P. 320–334.
5. Parker S., Stitzel M., Taylor D., Orozco J., Erdos M., Akiyama J., Bueren K., Chines P., Narisu N., Black B., et al. // Proc. Natl. Acad. Sci. USA. 2013. V. 110. № 44. P. 17921–17926.
6. Hnisz D., Abraham B., Lee T., Lau A., Saint-André V., Sigova A., Hoke A., Young R. // Cell. 2013. V. 155. № 4. P. 934–947.
7. Hnisz D., Schuijers J., Lin C.Y., Weintraub A., Abraham B., Lee T., Bradner J., Young R. // Mol. Cell. 2015. V. 58. № 2. P. 362–370.
8. Nitulescu I., Meyer S., Wen Q., Crispino D., Lemieux M., Levinee R., Pelish H., Shair M. // EBioMedicine. 2017. V. 26. P. 112–125.
9. Thandapani P. // Pharmacol. Ther. 2019. V. 199. P. 129–138.
10. Peng Y., Zhang Y. // Animal Model Exp. Med. 2018. V. 1. № 3. P. 169–179.
11. The comprehensive human Super-Enhancer database (http://www.licpathway.net/sedb) // Date of the application: 05.08.2020.
12. Pott S., Lieb J. // Nat. Genet. 2014. V. 47. № 1. P. 8–12.
13. Hnisz D., Shrinivas K., Young R., Chakraborty A., Sharp P. // Cell. 2017. V. 169. № 1. P. 13–23.
14. Chapuy B., McKeown M., Lin C., Monti S., Roemer M., Qi J., Rahl P., Sun H., Yeda K., Doench J., et al. // Cancer Cell. 2013. V. 24. № 6. P. 777–790.
15. Ball B., Abdel-Wahab O. // Trends Pharmacol. Sci. 2018. V. 39. № 12. P. 1002–1004.
16. Tan Y., Li Y., Tang F. // Mol. Cancer. 2020. V. 19. № 1. P. 74.
17. Ghazzaui N., Issaoui H., Saintamand A., Oblet C., Carrion C., Denizot Y. // Blood Adv. 2018.V. 2. № 3. P. 252–262.
18. Santos J., Braikia F., Oudinet C., Joana M., Dauba A., Khamlichi A. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 29. P. 14708–14713.
19. Lio C., Shukla V., Samaniego-Castruita D., González-Avalos E., Chakraborty A., Yue X., Schatz D., Rao A. // Sci. Immunol. 2019. V. 4. № 34. P. eaau7523.
20. Loft A., Forss I., Siersbæk M., Schmidt S., Larsen A., Madsen J., Pisani D., Nielsen R., Aagaard M., Mathison A., et al. // Genes Dev. 2014. V. 29. № 1. P. 7–22.
21. Martinez M., Medrano S., Brown E., Tufan T., Shang S., Bertoncello N., Guessoum O., Adli M., Belyea B., Sequeira-Lopez M., Gomez R. // J. Clin. Invest. 2018. V. 128. № 11. P. 4787–4803.
22. Huang S., Li X., Zheng H., Si X., Li B., Wei G., Li C., Chen Y., Chen Y., Liao W., et al. // Circulation. 2019. V. 139. № 25. P. 2857–2876.
23. Brown J., Lin C., Duan Q., Griffin G., Federation A., Paranal R., Bair S., Newton G., Lichtman A., Kung A., et al. // Mol. Cell. 2014. V. 56. № 2. P. 219–231.
24. Iwakawa R., Takenaka M., Kohno T.,Shimada Y., Totoki Y., Shibata T., Tsuta T., Nishikawa R., Noguchi M., Sato-Otsubo A., et al. // Genes Chromosomes Cancer. 2013. V. 52. № 9. P. 802–816.
25. Herranz D., Ambesi-Impiombato A., Palomero T., Sch-

nell S., Belver L., Wendorff A., Xu L., Castillo-Martin M., Llobet-Navás D., Cordon-Cardo C., et al. // Nat. Med. 2014. V. 20. № 10. P. 1130–1137.
26. Zhang X., Choi P., Francis J., Imielinski M., Watanabe H., Cherniack A., Meyerson M. // Nat. Genet. 2015. V. 48. № 2. P. 176–182.
27. Zhang X., Choi P., Francis J., Gao G., Campbell J., Ramachandran A., Mitsuishi Y., Ha G., Shih J., Vazquez F., et al. // Cancer Discov. 2018. V. 8. № 1. P. 108–125.
28. Mansour M., Abraham B., Anders L., Berezovskaya A., Gutierrez A., Durbin A., Etchin J., Lawton L., Sallan S., Silverman L., et al. // Science. 2014. V. 346. № 6215. P. 1373–1377.
29. Kuuluvainen E., Domènech-Moreno E., Niemelä E., Mäkelä T. // Mol. Cell Biol. 2018. V. 38. № 11. P. e00573-17.
30. Gimple R., Kidwell R., Kim L., Sun T., Gromovsky A., Wu Q., Wolf M., Lv D., Bhargava S., Jiang L., et al. // Cancer Discov. 2019. V. 9. № 9. P. 1248–1267.
31. Hu Y., Zhang Z., Kashiwagi M., Yoshida T., Joshi I., Jena N., Somasundaram R., Emmanuel A., Sigvardsson M., Fitamant J., et al. // Genes Dev. 2016. V. 30. № 17. P. 1971–1990.
32. Clarke P., Ortiz-Ruiz M., TePoele R., Adeniji-Popoola O., Box G., Court W., Czasch S., Bawab S., Esdar C., Ewan K., et al. // eLife. 2016. V. 5. P. e20722.
33. Pelish H., Liau B., Nitulescu I., Tangpeerachaikul A., Poss Z., Da Silva D., Caruso B., Arefolov A., Fadeyi O., Christie A., et al. // Nature. 2015. V. 526. № 7572. P. 273–276.
34. Shi J., Whyte W., Zepeda-Mendoza C., Milazzo J., Shen C., Roe J., Minder J., Mercan F., Wang E., Eckersley-Maslin M., et al. // Genes Dev. 2013. V. 27. № 24. P. 2648–2662.
35. Hnisz D., Weintraub A., Day D., Valton A., Bak R., Li C., Goldmann J., Lajoie B., Fan Z., Sigova A., et al. // Science. 2016. V. 351. № 6280. P. 1454–1458.
36. Zhou H., Schmidt S., Jiang S., Willox B., Bernhardt K., Liang J., Johannsen E., Kharchenko P., Gewurz B., Kieff E., et al. // Cell Host Microbe. 2015. V. 17. № 2. P. 205–216.
37. Alqahtani A., Choucair K., Ashraf M., Hammouda D., Alloghbi A., Khan T., Senzer N., Nemunaitis J. // Future Sci OA. 2019. V. 5. № 3. FSO372.
38. Borthakur G., Wolff J., Aldoss I., Hu B., Dinh M., Torres A., Chen X., Rizzieri D., Sood A., Odenike O., et al. // J. Clin. Oncol. 2018. V. 36. № 5. P. 7019.
39. Sobhani N., D'Angelo A., Pittacolo M., Roviello Z., Miccoli A., Corona S., Bernocchi O., Generali D., Otto T. // Cells. 2019. V. 8. № 4. P. 321.
40. Kennedy A., Vallurupalli M., Chen L., Crompton B., Cowley G., Vazquez F., Weir A., Tsherniak A., Parasuraman S., Kim S., et al. // Oncotarget. 2015. V. 6. P. 30178–30193.
41. Chou J., Quigley D., Robinson T., Feng F., Ashworth A. // Cancer Discov. 2020. V. 10. № 3. P. 351–370.
42. Greenleaf A. // Transcription. 2019. V. 10. № 2. P. 91–110.
43. Kwiatkowski N., Zhang T., Rahl P., Abraham B., Reddy J., Ficarro S., Dastur A., Amzallag A., Ramaswamy S., Tesar B., et al. // Nature. 2014. V. 511. № 7511. P. 616–620.
44. Rix U., Superti-Furga G. // Nat. Chem. Biol. 2009. V. 5. № 9. P. 616–624.
45. Chipumuro E., Marco E., Christensen C., Kwiatkowski N., Zhang T., Hatheway C., Abraham B., Sharma B., Yeung C., Altabef A., et al. // Cell. 2014. V. 159. № 5. P. 1126–1139.
46. A Study of SY 5609, a Selective CDK7 Inhibitor, in Advanced Solid Tumors (https://clinicaltrials.gov/ct2/show/NCT04247126) // Date of the application: 05.08.2020.
47. Zhang T., Kwiatkowski N., Olson C., Dixon-Clarke S., Abraham B., Greifenberg A., Ficarro S., Elkins J., Liang Y., Hannett N., et al. // Nat. Chem. Biol. 2016. V. 12. P. 876–884.
48. Liang S., Hu L., Wu Z., Chen Z., Liu S., Xu X., Qian A. // Cells. 2020. V. 9. № 6. P. 1483.
49. Galbraith M., Donner A., Espinosa J. // Transcription. 2010. V. 1. P. 4–12.
50. Donner A., Ebmeier C., Taatjes D., Espinosa J. // Nat. Struct. Mol. Biol. 2010. V. 17. P. 194–201.
51. Galbraith M., Allen M., Bensard C., Wang X., Schwinn M., Qin B., Long H., Daniels D., Hahn W., Dowell R., et al. // Cell. 2013. V. 153. P. 1327–1339.
52. McDermott M., Chumanevich A., Lim C., Liang J., Chen M., Altilia S., Oliver D., Rae J., Shtutman M., Kiaris H., et al. // Oncotarget. 2017. V. 8. P. 12558–12575.
53. Chen M., Liang J., Ji H., et al. // Proc. Natl. Acad. Sci. USA. 2017. V. 114. P. 10208–10213.
54. McClelandM., SoukupT., Liu S., Esensten J., de Sousa E Melo, Yaylaoglu M., Warming S., Roose-Girma M., Firestein R. // J. Pathol. 2015. V. 237. P. 508–519.
55. Westerling T., Kuuluvainen E., Makela T. // Mol. Cell Biol. 2007. V. 27. P. 6177–6182.
56. Loncle N., Boube M., Joulia L., Boschiero C., Werner M., Cribbs D., Bourbon H. // EMBO J. 2007. V. 26. P. 1045–1054.
57. Xie X., Hsu F., GaoX., Xu W., Ni J., Xing Y., Huang L., Hsiao H., Zheng H., Wang C., et al. // PLoS. Biol. 2015. V. 13. № 7. P. e1002207.
58. Firestein R., Bass A., Kim S., Dunn I., Silver S., Guney I., Freed E., Ligon A., Vena N., Ogino S., et al. // Nature. 2008. V. 455. P. 547–551.
59. Alarcon C., Zaromytidou A., Xi Q., Gao S., Yu J., Fujisawa S., Barlas A., Miller A., Manova-Todorova K., Macias M., et al. // Cell. 2009. V. 139. P. 757–769.
60. Xu W., Wang Z., Zhang W.,Qian K., Li H., Kong D., Li Y., Tang Y. // Cancer Lett. 2015. V. 356. P. 613–627.
61. Porter D., Farmaki E., Altilia S., Schools G., West D., Chen M., Chang B., Puzyrev A., Lim C., Rokow-Kittell R., et al. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. P. 13799–13804.
62. Broude E., Gyorffy B., Chumanevich A., Chen M., McDermott M., Shtutman M., Catroppo J., Roninson I. // Curr. Cancer Drug Targets. 2015. V. 15. P. 739–749.
63. Xu D., Li C., Zhang X., Gong Z., Chan C., Lee S., Jin G., Rezaeian A., Han F., Wang J., et al. // Nat. Commun. 2015. V. 6. P. 6641.
64. Kapoor A., Goldberg M., Cumberland L., Ratnakumar K., Segura M., Emanuel P., Menendez S., Vardabasso C., Leroy G., Vidal C., et al. // Nature. 2010. V. 468. P. 1105–1109.
65. Adler A., McCleland M., Truong T., Lau S., Modrusan Z., Soukup T., Roose-Girma M., Blackwood E., Firestein R. // Cancer Res. 2012. V. 72. P. 2129–2139.
66. Menzl I., Witalisz-Siepracka A., Sexl V. // Pharmaceuticals (Basel). 2019. V. 12. № 2. P. 92.
67. Xi M., Chen T., Wu C., Gao X., Wu Y., Luo X., Du K., Yu L., Cai T., Shen R., Sun H., et al. // Eur. J. Med. Chem. 2019. V. 164. P. 77–91.
68. Solum E., Hansen T., Aesoy R., Herfindal L. // Bioorg. Med. Chem. 2020. V. 28. № 10. P. 115461.
69. Hatcher J., Wang E., Johannessen L., Kwiatkowski N., Sim T., Gray N. // ACS Med. Chem. Lett. 2018. V. 9. № 6. P. 540–545.
70. Philip S., Kumarasiri M., Teo T., Yu M., Wang S. // J. Med. Chem. 2018. V. 61. № 12. P. 5073–5092.
71. Chen M., Li J., Liang J., Thompson Z., Kathrein K., Broude E., Roninson I. // Cells. 2019. V. 8. № 11. P. 1413.
72. Chen W., Ren X., Chang C. // ChemMedChem. 2019. V. 14. № 1. P. 107–118.
73. He L., Zhu Y., Fan Q., Miao D., Zhang S., Liu X., Zhang C. // Bioorg. Med. Chem. Lett. 2019. V. 29. № 4. P. 549–555.

74. Jia Q., Chen S., Tan Y., Li Y., Tang F. // Exp. Mol. Med. 2020. V. 52. № 5. P. 713–723.
75. Sengupta S., George R. // Trends Cancer. 2017. V. 3. № 4. P. 269–281.
76. Brinda B., Khan I., Parkin B., Konig H. // J. Cell. Mol. Med. 2018. V. 22. № 3. P. 1411–1427.
77. Tan Y., Wu Q., Zhou F. // Crit. Rev. Oncol. Hematol. 2020. V. 152. P. 102993.
78. Khamidullina A., Tatarskiy V., Yastrebova M., Nuzhina J., Ivanova E., Lim C.-U., Chen M., Broude E., Roninson I., Shtil A. // Proc. EACR conference «A Matter of Life and Death: Mechanisms, Models and Therapeutic Opportunities». Italy, 2020. P. 60.
79. Feng S., Yan P., Zhang Q., Li Z., Li C., Geng Y., Wang L., Zhao X., Yang Z., Cai H., Wang X. // Int. J. Colorectal Dis. 2020. V.35. P. 1355–1369.
80. Kotani D., Kuboki Y., Horasawa S., Kaneko A., Nakamura Y., Kawazoe A., Bando H., Taniguchi H., Shitara K., Kojima T., et al. // BMC Cancer. 2019. V. 19. № 1. P. 1253.
81. Liang J., Chen M., Hughes D., Chumanevich A., Altilia S., Kaza V., Lim C., Kiaris H., Mythreye K., Pena M., et al. // Cancer Res. 2018. V. 78. № 3. P. 6594–6606.

Chromatin Modifiers in Transcriptional Regulation: New Findings and Prospects

M. Yu. Mazina, N. E. Vorobyeva*
Institute of Gene Biology RAS, Group of transcriptional complexes dynamics, Moscow, 119334 Russia
*E-mail: vorobyeva@genebiology.ru
Received July 28, 2020; in final form, December 17, 2020
DOI: 10.32607/actanaturae.11101

ABSTRACT **Histone-modifying and remodeling complexes are considered the main coregulators that affect transcription by changing the chromatin structure. Coordinated action by these complexes is essential for the transcriptional activation of any eukaryotic gene. In this review, we discuss current trends in the study of histone modifiers and chromatin remodelers, including the functional impact of transcriptional proteins/complexes i.e., "pioneers"; remodeling and modification of non-histone proteins by transcriptional complexes; the supplementary functions of the non-catalytic subunits of remodelers, and the participation of histone modifiers in the "pause" of RNA polymerase II. The review also includes a scheme illustrating the mechanisms of recruitment of the main classes of remodelers and chromatin modifiers to various sites in the genome and their functional activities.**
KEYWORDS **transcription, chromatin, enhancer, co-regulator, remodelling, transcriptional factor.**

INTRODUCTION

The general activation process of eukaryotic gene transcription begins with the binding of an activator protein (for example, a hormone receptor) to a regulatory element. The activator protein, with the help of protein complex coregulators, promotes the recruitment of general transcription factors (GTFs) to the gene. Multiprotein coregulatory complexes coordinate the transcription process; they integrate signals from various DNA-binding activators and chromatin modifications and transmit them to GTFs (*Fig. 1A*). The principal coregulatory complexes involved in the transcription of any gene are chromatin modifiers. They are divided into two large, functionally different groups: complexes that change the position of nucleosomes and those that covalently modify histones in chromatin (*Fig. 1B*).

It is known that hundreds of different proteins are involved in the activation of transcription. Apparently, they cannot bind the regulatory elements of the activated gene simultaneously throughout the entire process of transcription activation (although this possibility had been previously assumed as part of the "histone" code hypothesis). Today, it is customary to describe the transcriptional process as extremely dynamic. Moreover, different coregulatory complexes are thought to be responsible for each of its many stages. This model mechanism of transcription regulation is called the "ratchet-clock mechanism" (*Fig. 1C*) [1]. According to this model, the intermediate markers regulating the directed exchange of transcriptional complexes at the DNA regulatory elements are covalent histone modifications [2, 3]. Covalent modifications can promote not only the recruitment, but also the removal of transcriptional complexes from the regulatory element, thereby stimulating the dynamics of the transcriptional process. It has been shown that a decrease in the time of association of transcriptional regulators with DNA enhances transcriptional activation [4]. A positive feature of the "ratchet-clock mechanism" model consists in that it illustrates the possibility of a large number of proteins functioning on a single regulatory element of a gene. The preservation of information from previously recruited coregulators in the form of a modification on chromatin allows the organism to maintain the general direction of the regulated process (movement towards the active work of the regulatory element or, conversely, suppression of its activity).

This review aims to summarize the available information on the functional properties and recruitment mechanisms of chromatin modifier complexes (this information is summarized in the form of a diagram in *Fig. 2*, which includes references to scientific studies

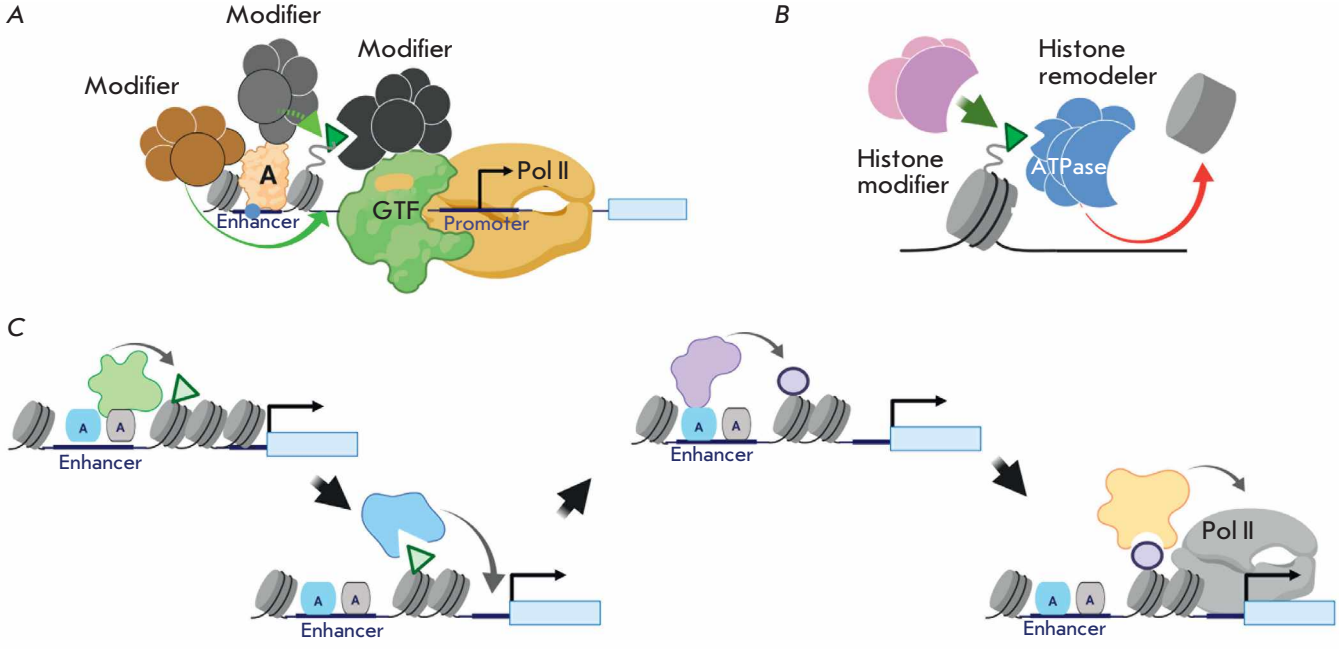


Fig. 1. (*A*) – general model of transcriptional activation of eukaryotic genes. A – transcriptional activator; GTF – General transcriptional factors; Pol II – RNA polymerase II. (*B*) – the main classes of chromatin modifiers: chromatin remodeling complexes and covalent histone-modifying complexes. (*C*) – "Ratchet-clock" model of transcriptional regulation. According to the "Ratchet-clock" mechanism, covalent histones modifications mediate the change of transcriptional complexes at the regulatory regions (play the role of connecting elements in transcription regulation). A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

describing the individual properties of the chromatin modifiers). In more detail, we describe those areas of study pertaining to chromatin modifiers that have advanced significantly in recent years. In addition, we discuss a number of issues that have not yet been resolved.

THE MOST ACTIVE FIELDS IN THE STUDY OF COACTIVATORS AFFECTING CHROMATIN

Transcriptional complexes that change the positions of nucleosomes

Since the emergence of chromatin (DNA packaged into fibrils using histone proteins) in the course of evolution, the most important way to regulate gene transcription has been to influence chromatin packaging, determining the availability of regulatory DNA elements. The protein complexes called chromatin remodelers belong to the transcriptional coregulators that affect the chromatin state [58, 59]. These transcriptional complexes are evolutionarily conserved (i.e., they are present in all eukaryotic organisms, from yeast to humans). Although the subunit composition of these complexes changes during evolution, their molecular properties (i.e., their ability to influence the position of nucleosomes in a certain direction) and the composition of their core subunits remain practically unchanged.

The molecular mechanisms of pioneer factors

DNA-binding transcription factors play the main role in the specificity of eukaryotic transcriptional regulation. It is the set of transcription factors associated with the regulatory element that affects its type of activity (which is realized by recruitment of various transcriptional complexes). It is a generally accepted fact that most transcription factors (for example, nuclear receptors) cannot bind to the regulatory DNA region occupied by nucleosomes. It is believed that a special class of DNA-binding proteins called pioneer factors is responsible for the displacement of compacted chromatin from regulatory DNA elements; the FoxA and GATA factors are prominent examples of this class [60]. These pioneer factors have a special property: the ability to bind regulatory DNA elements in a state of compacted chromatin and bring them into a state competent for binding by other transcription factors. Thus, pioneer factors are, in essence, the primary regulator-remodelers, initiating changes in the chromatin structure, which is further supported by transcriptional remodeling complexes. Despite the fact that the concept of

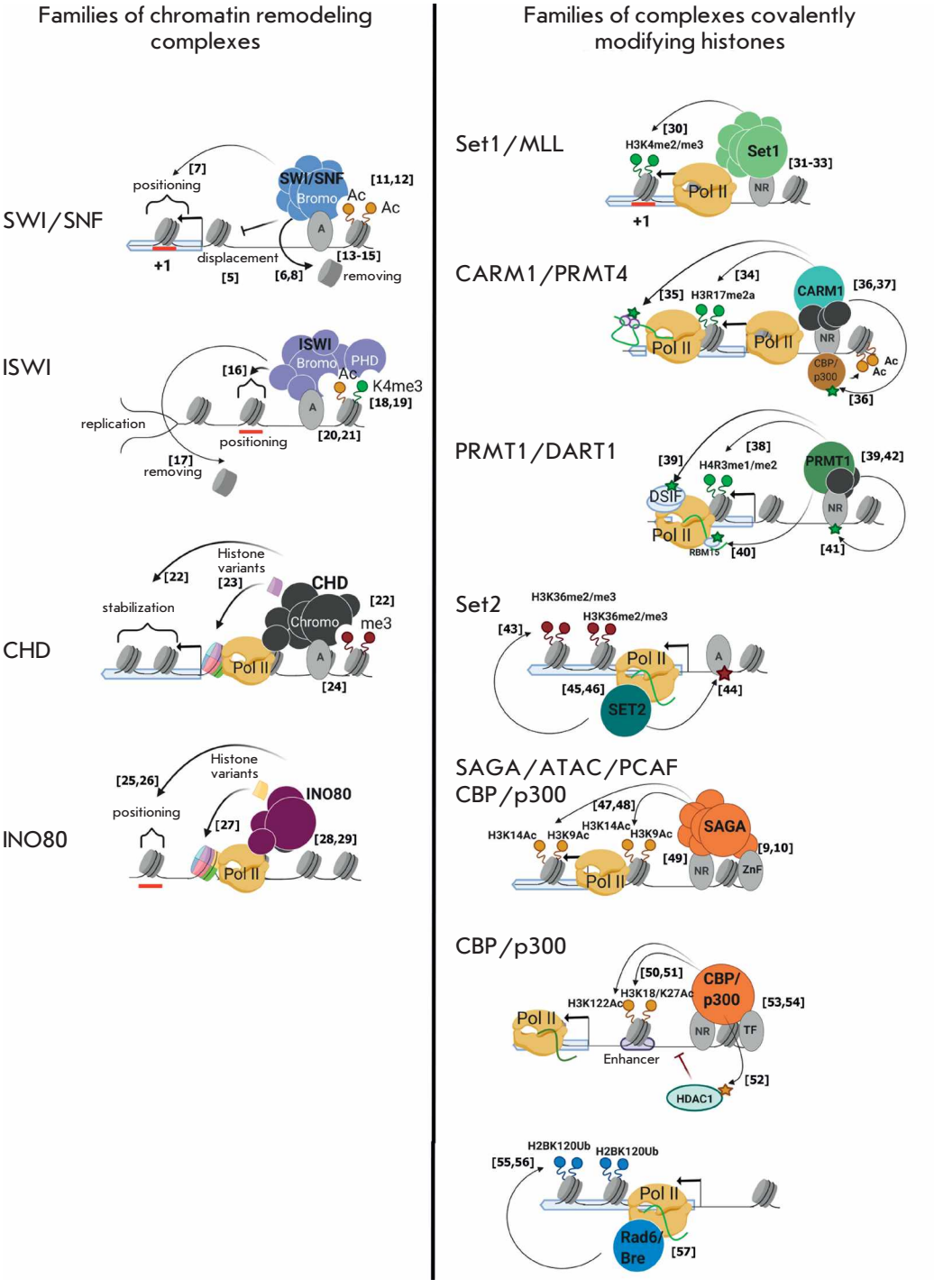


Fig. 2. Functional features and mechanisms of recruitment for histone remodelers and histone modifiers into chromatin. Abbreviations: A – activator, NR – nuclear receptor, TF – transcription factor. All models were created using the app BioRender.com

pioneer factors was formulated almost 10 years ago, the molecular mechanism of the functioning of these proteins remains not fully understood. Initially, it was thought that pioneer factors function on their own, without the participation of remodeling transcriptional complexes (this assumption was based on the ability of these proteins to bind chromatinized DNA *in vitro*) [61]. At the same time, it has long been noted that pioneering factors *in vivo* are capable of affecting chromatin in quite complex ways (for example, replacing histones H2A with H2AZ), which is hardly possible for individual monomer proteins [62].

According to current views, it is unlikely that pioneer factors function as single proteins in living cells. Most likely, their unique ability to act on compacted chromatin is a consequence of cooperative multiprotein interactions. An example of such joint functioning can be the paired work of a pioneer factor with a nuclear

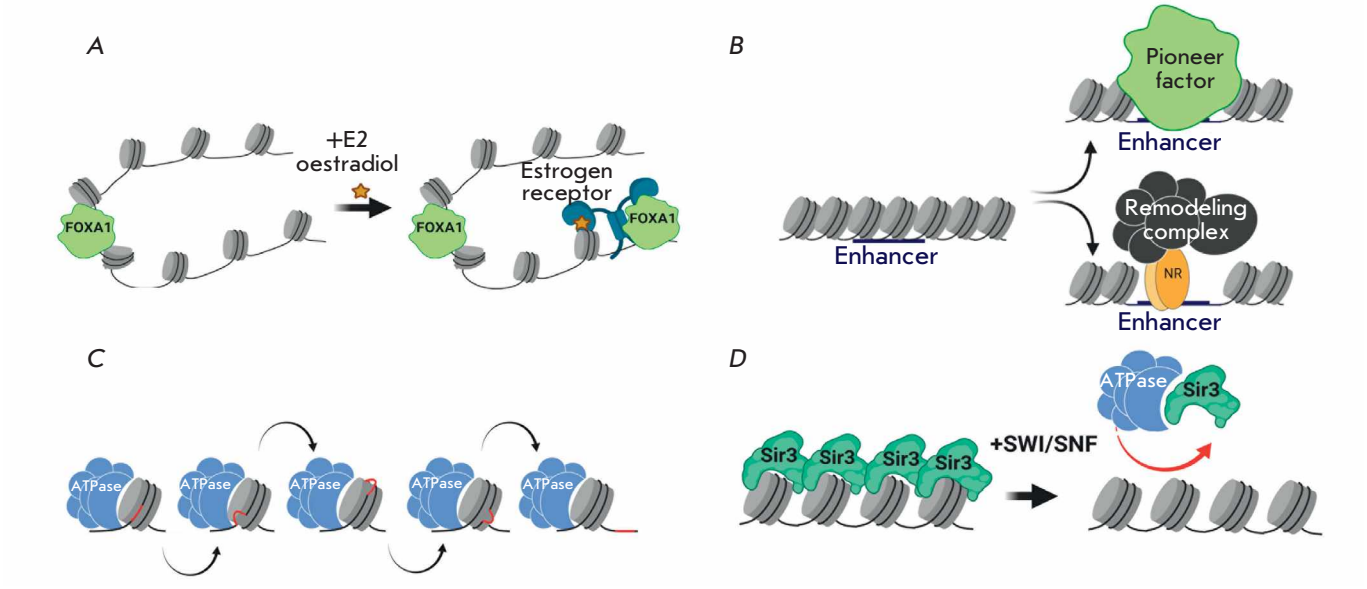


Fig. 3. (*A*) – cooperative work of the pioneer factor FoxA1 and the nuclear receptor ERα during the chromatin de-compaction at the DNA regulatory sites. (*B*) – the primary binding of transcriptional regulators to chromatinized DNA elements can be carried out both by specialized DNA-binding factors-"pioneers" or the chromatin remodeling complexes associated with the nuclear receptors. (*C*) – all families of remodeling complexes operate via the same molecular mechanism: the formation of a DNA loop on the nucleosome and the change of its position relative to the nucleosome surface. (*D*) – effect of SWI/SNF chromatin remodeler on the binding of the Sir3p repressor to chromatin. The SWI/SNF complex in yeast is able to interact with the heterochromatin repressor Sir3p and remove it from chromatin. A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

receptor (for example, the pioneer factor FoxA1 and the nuclear receptor ERα) [63]. It has long been known that the binding of FoxA1 and ERα to DNA occurs cooperatively. However, it was assumed that the pioneering factor plays a leading role in this process, as it is the suppression of FoxA1 expression that leads to the removal of 90% of ERα genomic sites with a very weak reverse effect in a reciprocal experiment [64]. Nevertheless, further studies have shown a more significant role for nuclear receptors in chromatin de-compaction at DNA regulatory sites. In fact, oestradiol treatment (of which ERα is a sensor) of MCF-7 cells leads to an increase in the number of FoxA1 binding sites by almost 30%, thereby demonstrating the ability of ERα to act as a pioneer factor, at least for some FoxA1 sites (*Fig. 3A*) [65]. Most likely, the ability of a nuclear receptor to play the role of a pioneer factor may be based on its ability to interact with transcriptional complexes and chromatin remodelers. It is known that many steroid receptors use SWI/SNF and NURF remodeling complexes to de-compact chromatin at the early stages of gene transcription activation [66, 67]. There is a hypothesis about the possibility of the formation of a common complex between the nuclear receptor and the transcriptional remodeler complex not on chromatin but in nucleosol [67]. Such a pair would be an effective pioneer factor capable of interacting with the regulatory regions within compacted chromatin (*Fig. 3B*). Further research is needed to understand how common this molecular mechanism is in nature.

The functional activity of chromatin remodeling complexes. The possibility of remodeling non-histone proteins

While the mechanism that organizes primary access to the regulatory elements of compacted chromatin remains unclear, the maintenance of nucleosome-free regions is undoubtedly the responsibility of chromatin remodeling complexes. In general, transcriptional remodeling complexes can affect nucleosomes in a variety of ways: to remove them, shift, position, or replace histones with alternative variants. However, all these mechanical functions are based on the same ability of remodelers to create a DNA loop within the nucleosome and change its position relative to the nucleosome's surface (*Fig. 3C*) [58]. The subunit composition of the remodeling complexes, as well as the structural features of the ATPase subunits (the presence of additional domains that are capable of binding a certain type of histone), determines the functional ability of the corresponding transcriptional complexes. Thus, complexes of the SWI/SNF family, which car-

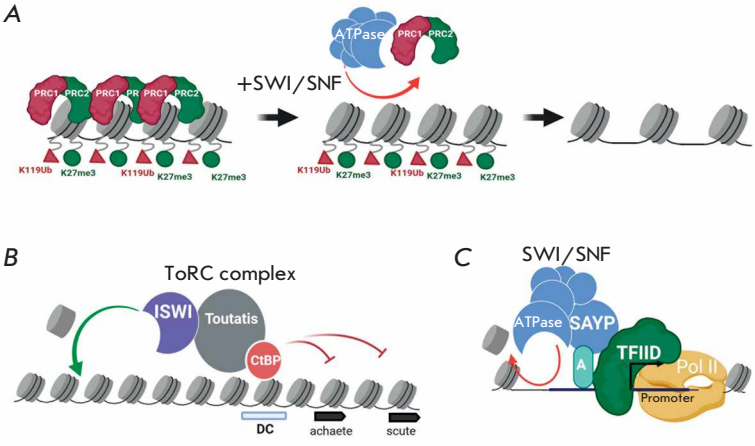


Fig. 4. (*A*) – artificial SWI/SNF recruitment leads to the decrease in PRC binding at the repressed loci. (*B*) – the ISWI remodeler, as part of the ToRC repressor complex of Drosophila, interacts with the transcriptional repressor CtBP. CtBP enhances the remodeling properties of ISWI, and ISWI is involved in the repression of the transcription of CtBP-dependent genes. (*C*) – the SAYP subunit of SWI/SNF mediates the recruitment of both SWI/SNF and TFIID to the genomic sites. A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

ry the SnAC domain that binds nucleosomes in their enzymatic subunit, are responsible for the removal of entire nucleosomes from chromatin [68]. Complexes of the INO80 family, which have a two-part translocation domain in their ATPase, are capable of replacing histones in nucleosomes with alternative variants [27]. The ISWI ATPase family, having a C-terminal HSS domain that binds unmodified histone H3 and regions of linker DNA, participates in the coreplicative assembly of chromatin, helping chaperones form high-grade nucleosomes within chromatin [69]. In addition, remodeling complexes of the ISWI and CHD families use their HSS and DBD domains for accurate postreplicative positioning of nucleosomes in chromatin [16].

It should be noted that chromatin remodeling complexes can directly influence not only the position of nucleosomes on DNA, but also the association of other DNA-binding proteins with chromatin [70]. The ability of remodeler translocation domains to bind and induce the movement of transcription factors and transcription repressors may play a significant role in the regulation of gene transcription. Thus, it was found that the ATPase of the SWI/SNF complex in yeast is capable of interacting with the heterochromatin repressor Sir3p and removing it from nucleosome templates *in vitro* [71]. More recently, it was proven *in vivo* that the SWI/SNF complex participates in the removal of the repressive effect of Sir3p from its target genes during the activation of their expression in the M/G1 phase of the cell cycle (*Fig. 3D*) [72].

The functional role of the SWI/SNF remodeling complex in the withdrawal of Pc-driven repression has been demonstrated as relates to various organisms [73]. The positive correlation in the violation of these molecular systems during oncotransformation of cells has been studied extensively [74]. Until recently, it was believed that the role of SWI/SNF complexes in the removal of PRC complexes from chromatin may be indirect. However, recent experiments on the artificial recruitment of SWI/SNF to the Pc-repressed locus have demonstrated direct removal of PRC complexes by the SWI/SNF complex (artificial recruitment of the latter led to a decrease in the PRC level within several minutes and did not depend on the recruitment of RNA polymerase II to the studied locus) (*Fig. 4A*) [75]. The role of remodeling complexes in the removal of transcription factors from chromatin is likely much more significant than is currently known. Unfortunately, the study of this mechanism *in vivo* is an extremely complicated methodological problem. The obtained information can almost always be questioned because of the presence of indirect experimental contributors.

The noncatalytic function of remodeling complexes in the regulation of transcription

Many transcriptional chromatin remodeling complexes are characterized by the presence of a large number of subunits, in addition to the enzymatic subunit responsible for histone movement [76]. Moreover, the number of subunits in these complexes increases over the course of evolution [77]. Previously, it was believed that the noncatalytic subunits of chromatin remodeling complexes are responsible for the specificity of recruitment to chromatin. It has been shown that a decrease in the intracellular level of individual noncatalytic SWI/SNF subunits of Drosophila leads to a complete disruption of the binding of this complex to chromatin, while preserving the structural stability of its core module that contains ATPase [78]. Recently, the attitude of researchers towards the functional capabilities of the noncatalytic subunits of remodelers has changed. There are data indicating the presence of additional functions in the noncatalytic subunits of chromatin remodeling complexes.

This development appears quite logical from an evolutionary point of view. Transcriptional activation and repression are extremely dynamic and complex processes. Within these processes, many multicompo-

nent complexes replace each other at high speed in a limited space (i.e., on regulatory DNA elements). This exchange assumes a high probability of contacts between the participants and, accordingly, the possibility of positive or negative mutual regulation. Chromatin remodeling complexes, in the course of their work on a regulatory element, bring with them many additional noncatalytic subunits. It is likely that while the ATPase part of the complex performs its main catalytic activity, the remaining subunits participate in the activation/repression of the transcription process [79]. The best characterized is the association of the ATPase subunits of remodeling complexes with transcriptional repressors. During the study of the ToRC repressor complex of Drosophila, the ability of the ISWI enzymatic remodeler subunit to physically interact with the transcriptional repressor CtBP was described [80]. Moreover, the ATPase and repressor subunits in this complex were discovered to exert a reciprocal functional effect on each other: CtBP enhances the ability of ISWI to remove or insert nucleosomes, and ISWI is apparently involved in the transcriptional repression of CtBP-dependent genes (*Fig. 4B*). Another remodeler enzyme, the chromodomain-containing ATPase CHD4/Mi-2, was also shown to be able to interact with other proteins to form the NuRD complex, which represses gene transcription [81]. This repressor complex contains more subunits than the ToRC complex described above. NuRD subunits form dynamically interacting modules with the remodeling activity implemented by the CHD4/Mi-2 subunit or the histone deacetylase activity due to Rpd3 [82]. The functional role of NuRD includes controlling for both the density of nucleosomes and their level of covalent modifications at developmental enhancers [83].

Interestingly, the ATPase subunit of the SWI/SNF complex, the BRM protein, was also found to exert a repressive effect on transcription, independent of its catalytic activity [84, 85]. At the moment, it is unclear which molecular partners enable the repressive functions of BRM ATPase. However, it seems reasonable to propose a mechanism for the positive role of SWI/SNF in transcription regulation that does not depend on the ATPase activity of this complex. Approximately ten years ago, the physical interaction of the Drosophila SWI/SNF complex with the common transcription factor TFIID, mediated by its SAYP subunit, was described [79, 86]. It was shown that the SAYP subunit plays a key role in the recruitment of the SWI/SNF complex to half of its genomic targets [87]. Interaction with TAF5 allows SAYP to recruit not only the SWI/SNF remodeling complex, but also TFIID to its genomic targets, contributing to the formation of the preinitiation complex (*Fig. 4C*) [79, 88, 89]. Thus, the noncatalytic SWI/SNF subunit is a bifunctional regulator that simultaneously promotes chromatin remodeling and transcription initiation.

The transcriptional complexes that covalently modify histones

Since the inception of the "histone code" hypothesis, proteins capable of covalent modification of histones have been the subject of numerous studies [90]. For a long time, it was assumed that the set of histone modifications determines the pattern of transcriptional complexes associated with the regulatory elements of the genome (which is the concept of the "histone code"). Currently, researchers are inclined to believe that the presence of certain chromatin modifications is a sufficient condition for the recruitment of only a limited number of regulators [1]. In most cases, the binding of histone modification is only an additional factor in the recruitment of the transcriptional regulator or may not even contribute at all to its recruitment to chromatin.

The role of covalent histone modifications in the recruitment of transcriptional complexes to chromatin

Initially, the "histone code" hypothesis was actively investigated in the context of the transcription activation process. Many researchers tried to establish the histone modifications that determine the recruitment of the protein complexes stimulating transcription. In turn, researchers who studied protein complexes worked to determine the protein domains responsible for the recruitment of the complexes to the corresponding "activating" modification. It is worth noting that many of these studies proved unsuccessful. It turned out that such "activating" histone modifications are often unable to recruit transcription complexes by themselves. A striking example of such an "activating" modification with a complex history of its study is the trimethylation of histone H3 at position 4. Indeed, there is much evidence of the correlation between the presence of this modification on the promoter and the active work of the corresponding gene [91]. However, the role of this modification in the recruitment of transcriptional regulators to the promoters of the corresponding genes is not so unambiguous. Domains capable of specifically interacting with the H3K4me3 modification have been identified in various protein complexes (among which the TFIID, NURF, mSin3a – HDAC1, and SAGA complexes are especially noteworthy) [92, 93]. For the first time, a specific domain that binds the H3K4me3 modification was identified in the ING2 protein, which is part of the mSin3a – HDAC1 repressive complex [94]. However, it was shown almost immediately that

the disruption of the interaction between ING2 and the modification of histone H3K4me3 leads to a change in the functional activity of the complex (a decrease in deacetylating activity) rather than to a violation of its recruitment [95]. The study of the domain that recognizes the H3K4me3 modification in the CHD1 chromatin regulator developed in a similar fashion [96]. The specific interaction of CHD1 with this chromatin modification disrupts the functional activity of the complex but does not prevent its interaction with chromatin [97]. It should be noted that, in the case of the TFIID and NURF coregulators, a positive contribution of the protein domains recognizing the H3K4me3 modification to the recruitment of these complexes to genomic sites was demonstrated [98–100].

Apparently, the process of recruitment of protein complexes to the regulatory elements of DNA is more complicated than we had imagined earlier: it is not realized through individual protein-protein interactions (for example, between a histone modification and a separate protein domain that "reads" the modification or between a DNA-binding transcription factor and a subunit of the protein complex). The protein complexes that regulate transcription most often contain a set of different subunits, many of which carry different protein domains (i.e., domains that are capable of DNA binding, recognize histone modifications, and interact with transcription factors). It appears that several domains that are part of various subunits are involved in a single act of recruitment of a transcriptional complex to chromatin. It is the set of such DNA-protein and protein-protein interactions that are realized in a separate act of recruitment of a complex to a regulatory element that can determine the type of functional activity of the complex in a given chromatin region (*Fig. 5A*).

The role of covalent histone modifications in the spreading of compacted chromatin, which represses transcription, has been investigated much more in detail and unambiguously. The cell uses various systems to create regions of compacted chromatin to suppress unwanted gene transcription. Two active systems of chromatin compaction can be distinguished based on the Pc and HP1 proteins, both containing chromoprotein domains capable of binding specific methylated residues of histone H3 [101–103]. Interestingly, for both

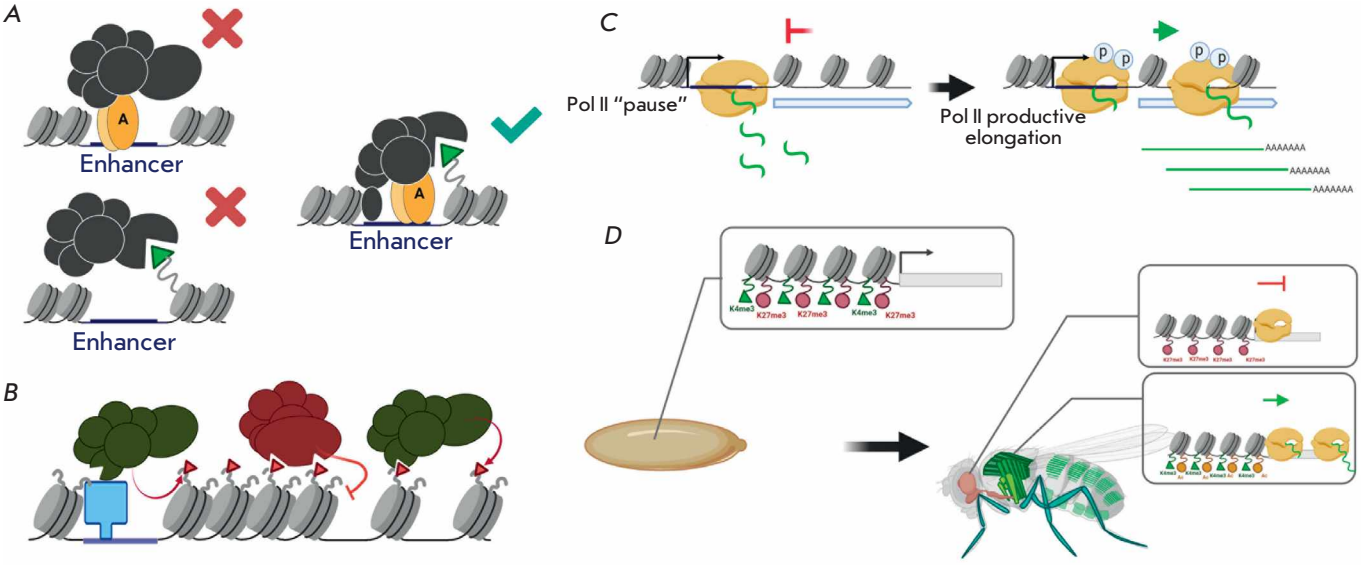


Fig. 5. (*A*) – the combinatorial nature of the recruitment of transcriptional co-regulators. The coregulator subunits often include DNA-binding motifs, domains recognizing covalent histone modifications, as well as domains of association with nuclear receptors and transcription factors. A number of protein domains can play a role in the association of a coregulator with a regulatory DNA element, as well as affect its functional activity. (*B*) – general concept of the role of histone modifications in the propagation of compacted chromatin. The initial recruitment of chromatin-compacting complexes is mediated by the DNA-binding factors. Covalent histone modifications are involved in a process of propagation of chromatin compaction around the site of the initial binding. (*C*) – The "pause" state of RNA polymerase II is characterized by the presence of short "abortive" transcripts at the promoter-proximal regions of inactive genes. (*D*) – genes with bivalent histone modifications (both, active, H3K4me3, and repressive, H3K27me3) in embryogenesis have enhanced capability of duality of action later in development (to be repressed or activated depending on the type of tissue they present). A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

chromatin compaction systems (Pc- and HP1-dependent), recognition of covalent histone modifications plays a role precisely at the stage of chromatin spreading within the chromosomal domain but not at the stage of primary recruitment of repressive complexes to DNA (which is carried out by specific DNA-binding proteins) (*Fig. 5B*). Thus, the propagation of Pc-dependent repression occurs with the participation of the PRC1 and PRC2 complexes, one of which is capable of recognizing the H3K27me3 chromatin modification, while the second introduces it. The interrelated molecular work of these complexes organizes the propagation of Pc-dependent repression around the PRE elements, which are initiators of Pc-dependent compaction [104]. Apparently, the histone modification of H3K27me3 is necessary for not only the propagation of Pc-dependent chromatin along the DNA strand, but also for the preservation of the corresponding chromatin status after the replication fork has passed through it [105]. A positive feedback loop based on the introduction of a covalent histone modification also exists in the mechanism of pericentromeric heterochromatin spreading. In this case, methyltransferase Su(var)3-9 (Suv39H in mammals) modifies histone H3 at position 9, which leads to the recruitment of the HP1 heterochromatin protein (which in turn recruits a new portion of methyltransferase to the compacted site) [106].

In the processes of activation and repression of transcription, the recognition of histone modifications is often not the primary signal that determines the recruitment of transcriptional regulators. The logical extension of the "histone" code idea was the hypothesis holding that covalent modifications of histones are necessary for the exchange of transcription complexes at regulatory sites [107]. This was facilitated by experiments that showed the existence of an active exchange of nucleosomes and associated proteins on working regulatory elements [108].

The role of covalent histone modifications in the regulation of the RNA polymerase II pause

For a long time, the recruitment of RNA polymerase II to promoters was considered the main mechanism of activation of gene transcription. Later, it became obvious that many inactive genes of multicellular organisms contain bound RNA polymerase II on their promoters [109]. The transcription of such genes is activated by stimulating the productive elongation of RNA polymerase II transcription. This mechanism of transcriptional regulation is called the "pause" of RNA polymerase II and is characterized by the presence of short "abortive" transcripts on the promoters of inactive genes (*Fig. 5C*). Currently, it is believed that this mechanism is widely used by organisms to regulate the transcription of genes that require high accuracy of induction in space and time (for example, in a certain tissue or developmental stage) [110]. The prevalence of this mechanism has made it an attractive area of research. One of the intensive areas of research on RNA polymerase II pausing was the search for the covalent histone markers associated with both the "pause" itself and the release of RNA polymerase II from this state.

For instance, the first description of bivalent nucleosomes was provided in the context of studying the "pause" of RNA polymerase II [111]. In mouse embryonic stem cells, it was found that the RNA polymerase II "pause" is present on promoters carrying the H3K27me3 modification in chromatin, which is characteristic of transcriptional repression. At the same time, RNA polymerase II was absent on promoters carrying both active H3K4me3 and repressive H3K27me3 modifications (containing bivalent nucleosomes) [112]. Later, it became clear that bivalent modifications in embryonic stem cells are mainly present on the promoters of genes whose transcription is regulated in different ways during cellular differentiation [113]. During development, these genes are activated in certain tissues (an active modification of H3K27Ac is introduced to their promoters), while in others, they remain inactive (the H3K4me3 modification is removed from their promoters, H3K27me3 is preserved, and the genes are put into a state of transcriptional "pause") (*Fig. 5D*) [114]. This concept has been supported by various data. The maintenance of the Pol II "paused" state and the transfer of promoters to this state were found to be carried out by enzymes that modify the K4 and K27 residues of histone H3. Thus, the maintenance of the "pause" state on gene promoters in mouse embryonic stem cells was associated with the activity of Lsd1-specific demethylase H3K4me3 [115]. For the JMJD3 enzyme, which is aimed at demethylation of the H3K27me3 modification, a role in the control of transcription elongation in human cells was also revealed [116]. It was shown that a decrease in the intracellular level of this demethylase leads to a decrease in the level of elongating RNA polymerase II.

There are a number of covalent histone modifications that are associated with the release of RNA polymerase II from a "pause" state and the stimulation of transcription elongation. The acetyl residues of histones are positive markers of transcription elongation. This role (to overcome the transcription pause and stimulate elongation) was discovered for the main acetyltransferase, functioning on enhancers, the CBP protein [117]. CBP was found to introduce acetylation at H3K27 of the first nucleosome in the gene's body, which is essential for the elongation of RNA polymerase II. Another acetyl modification of histones, H3K9Ac, was

associated with the release of RNA polymerase II from the "pause" state by recruitment of the SEC (super elongation complex), which contains a number of the factors necessary for transcription elongation [118]. A decrease in the level of H3K9Ac was shown to prevent the elongation of genes and to lead to an increase in the "pause" index (i.e., an increase in the ratio between the levels of RNA polymerase II on the promoter and in the gene's body).

Recently, our group has studied the kinetics of recruitment of chromatin modifiers and the appearance of covalent histone modifications in the first minutes of transcriptional activation (on a model of developmental genes, which persist in a "pause" state in *Drosophila* cells) [119]. We have studied the recruitment of two dozen transcriptional complexes, which allowed us to identify an unexpected regulatory effect. We almost did not observe an increase in the level of binding of chromatin-modifying complexes with promoters during their activation. At the same time, we found a significant increase in the level of chromatin modifications introduced by these complexes. We called this effect the "pause" of transcriptional coactivators (*Fig. 6A*). We believe that during the formation of a transcriptional "pause," not only RNA polymerase II, but also many coregulatory complexes that modify chromatin are recruited to the promoters. The signal-inducing transcription does not lead to a further increase in the level of binding of these complexes but stimulates their functional activity, leading to an increase in the level of chromatin modifications. We plan to test the prevalence of the effect of coactivator "pause" in the *Drosophila* genome in future studies.

Are covalent histone modifications actually "side targets"?

The covalent modifications of histones have long attracted the attention of researchers because of the popularity of the "histone code" hypothesis. In particular, the introduction of covalent histone modifications was described as the main molecular function for a variety of transcriptional regulators (including multisubunit complexes). Later, it was discovered that a number of histone modifications make rather modest functional contributions to the regulation of transcription and that the enzymatic functions of the regulators introducing them have completely different, nonhistone protein targets, which are of greater importance.

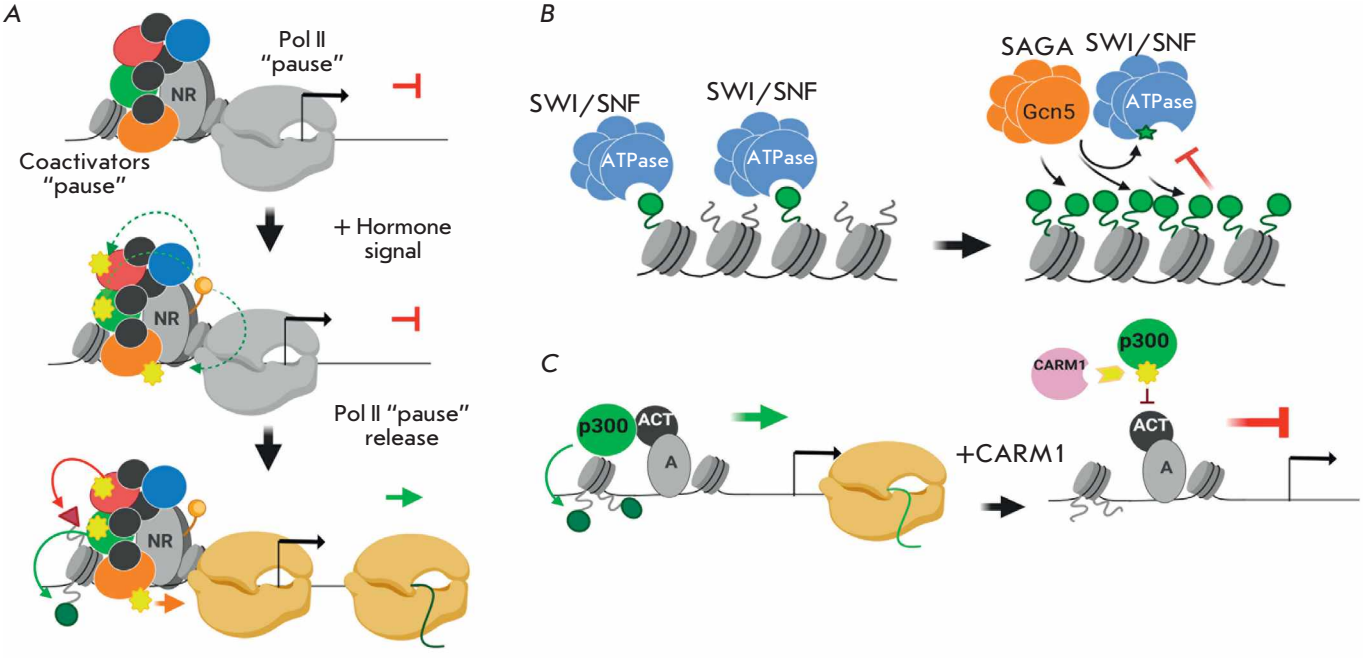


Fig. 6. (*A*) – promoter regions of genes regulated via Pol II pausing contain not only pre-bound Pol II, but also pre-bound co-activators in their inactive state. Transcriptional induction is realized with the transition of co-regulators into the functional state but not with an increase in their promoter-bound level. (*B*) – the SAGA histone acetyltransferase complex acetylates the ATPase subunit of the SWI/SNF complex, regulating its ability to bind chromatin. (*C*) – arginine methyltransferase CARM1 methylates the CBP/p300 acetyltransferase, decreasing the activity of CBP/p300 and disrupting its ability to bind transcription activators. A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

A striking example of such a chromatin modifier is the SAGA complex, which is capable of acetylating lysine residues in the histones H3 and H4. For a long time, researchers believed that the acetyl groups introduced by this complex are specific labels that are accurately "read" by other transcriptional regulators using protein "reader" domains. In particular, it was assumed that the modifications introduced by the SAGA complex are recognized by bromodomains, which are part of the SWI/SNF complex, which acquires the ability to remodel exactly acetylated histones [120]. This hypothesis was in good agreement with the joint presence of SAGA and SWI/SNF complexes at the genomic sites of various organisms [10, 121]. Over time, it became clear that the acetyl residues of histones are unlikely to be specific markers for the recruitment of any specific complexes. The point is that the functional effect of acetyl chromatin residues on transcription has a combinatorial nature that depends on the total number of modified residues but is almost indifferent to their qualitative composition [122, 123]. Deeper studies have led to the description of additional targets for acetylation by the SAGA complex. In particular, SAGA acetylates the ATPase subunit of the SWI/SNF complex, thereby regulating the strength of its binding to chromatin (*Fig. 6B*) [124].

Additional protein targets have been described for other chromatin modifiers. Thus, the arginine methyltransferase CARM1, originally described as a specific modifier of 17-arginine in histone H3, was found to methylate the arginine residues of many transcriptional regulators, modulating their functions [125, 126]. In particular, the targets of CARM1 activity are splicing factors, and methylation provokes exon skipping in mRNA [35]. Another target of CARM1 methylation is CBP/p300 acetyltransferase, which is one of the key enzymes that function on enhancers. Methylation of CBP/p300 by CARM1 decreases acetyltransferase activity and impairs its ability to bind transcriptional activators (*Fig. 6C*) [36, 127].

As we can see, a deeper study of transcriptional regulators, initially characterized as chromatin modifiers, leads to the description of their additional enzymatic targets. It is likely that further study of these additional targets will uncover a higher functional significance in comparison with target histones, which for a number of modifiers can only be "side targets." This assumption is supported by the results of some mutational studies that aimed at identifying the functional significance of individual histone modifications. It has been shown that mutations in individual chromatin modifiers have a stronger effect on the regulation of transcription than mutations in their target sites in histones, thereby demonstrating the presence of more significant targets – transcriptional regulators [128, 129]. It is likely that future studies will reveal other histone modifications that are only a by-product of the action of the chromatin modifier, achieving its main regulatory target.

FURTHER PROSPECTS AND UNRESOLVED ISSUES

The growing volume of experimental information on the mechanisms of transcriptional regulation and the activity of coregulators has not led to answers to some of the questions that were formulated earlier. Below, we will address several problems for which definite solutions have yet to be found, despite the wealth of experimental arsenals available today.

Recruiting a transcriptional regulator to chromatin: in a complex with a DNA-binding protein or sequentially?

DNA-binding proteins determine the specificity of the effect of the coregulator on gene transcription. They mediate the binding of coregulators to enhancer and promoter sequences. Until now, the general mechanism of interaction of coregulators with DNA has remained unclear. Does the sequential binding of the DNA-binding protein to the regulatory element and the subsequent recruitment of the coregulatory complex occur, or is the preformed complex recruited to the genomic sites? (*Fig. 7A*).

The concept of sequential recruitment looks questionable in the context of studies focused on the dynamics of protein binding to chromatin. It was shown that the association of any proteins with DNA only lasts a few minutes [107]. In this regard, sequential association of proteins on the regulatory element looks unlikely – there remains very little time for their functional action. The hypothesis of the simultaneous recruitment of coregulators and DNA-binding proteins has been invoked repeatedly for a long time. Nevertheless, the initial concept of sequential binding appears more widespread [67]. However, it has been shown recently that complexes of transcription factors — nuclear receptors with chromatin remodeling coregulators — are capable of interacting with chromatin, acting as "pioneer" factors [130]. Moreover, it was demonstrated that knockouts of the ATPase subunits of the SWI/SNF and ISWI coregulators significantly disrupt the binding of transcription factors in the genome of mouse embryonic stem cells (which would be impossible in the concept of sequential recruitment) [131]. All these data substantiate the model of joint recruitment of coregulators and DNA-binding factors. Biochemical isolation of transcription factors, in combination with coregulators, would be very useful to further lend credence to this concept. However, the connection between the

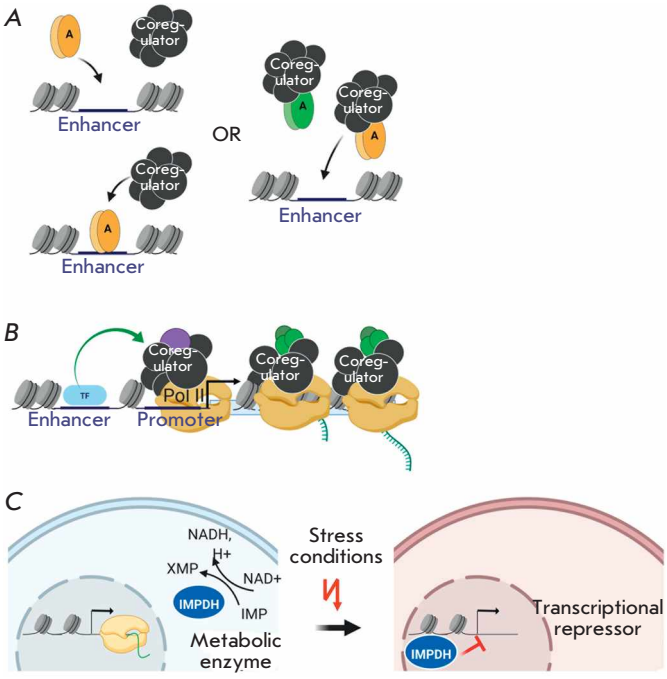


Fig. 7. (*A*) – recruitment of a transcriptional regulator to chromatin: in a combination with a DNA-binding protein or stepwise. (*B*) – changes in the subunit composition of protein complexes during the transcriptional cycle: transformation of the same complex or recruitment of a new complex. (*C*) – inosine-5′-monophosphate dehydrogenase (IMPDH), an enzyme of purine biosynthesis that works in the cytoplasm of the cell, is able to shuttle into the nucleus under stress conditions and regulate gene transcription. A more detailed description of the figures is given in the text. There are also references to the works that served as the basis for the molecular models. All illustrations were created using the app BioRender.com

transcription factor and the coregulator is a specific interaction, albeit a weak one, that is easily lost during biochemical purification. Let us hope that the recently developed techniques for studying weak protein-protein interactions (e.g., *in vivo* biotinylation of proteins by their partners) will help answer questions regarding the mechanisms underlying the interaction between coregulators and chromatin.

Changes in the subunit composition of protein complexes during transcription: transformation of the same complex or recruitment of a new complex?

Many coregulatory complexes are involved in the various stages of gene transcription. Often, in the course of the study of such complexes, researchers focus on examining the distribution and properties of the enzymatic subunits of the complex, while the behavior of the other subunits remains unexplored. Nevertheless, for a number of transcriptional complexes, it was shown that their composition is not constant and can change depending on the stage of gene transcription (*Fig. 7B*). Thus, it is known that the transcriptional coregulator SAGA exhibits acetyltransferase and deubiquitinylating activities towards histones. Both of these activities are required for SAGA to function on the gene promoter, where it promotes the initiation of transcription [132, 133]. At the same time, it is known that a component of the deubiquitinylating module of the SAGA complex, the SGF11 protein, is also associated with the CAP of newly synthesized mRNA as a part of the AMEX complex, where it is involved in mRNA export from the nucleus to the cytoplasm [134]. An interesting detail is the possibility of transition of the SAGA subunit SGF11 to the AMEX complex during transcription. Is there an independent recruitment of two separately existing complexes to the active gene? Or is there a subunit transformation of the SAGA complex, initially recruited to the promoter, during the transition of the RNA polymerase II complex into the body of the gene?

Another well-known example of a change in the subunit composition of a coregulator during transcription is the Mediator complex. The main role of this large multisubunit complex is to coordinate the recruitment of RNA polymerase to the promoter and initiate transcription [135]. However, the Mediator contains a separate four-subunit CDK8 module that possesses kinase activity and a number of additional functions. Interestingly, the interaction of the core Mediator with RNA polymerase II mutually excludes the presence of the CDK8 module. Moreover, the role of the CDK8 module in the stimulation of elongation, that is, in the latest stages of transcriptional activation, is well known [136]. It remains unclear how the module is recruited to CDK8-dependent genes in order to participate in elongation stimulation. Is this an alternative to Mediator-dependent recruitment, or is there a structural transformation of the entire Mediator complex during the transcriptional cycle?

The two examples given above are only an illustration of the challenges in the study of multisubunit complexes. There are many indirect confirmations of changes in the composition and properties of coregulatory complexes during transcription. However, there is still no direct experimental evidence of these phenomena due to a lack of convenient research methods.

Influence of nontranscriptional complexes on transcription: hierarchy of functions and determination of the leading function

The original strategy researchers employed to study the functions of proteins and protein complexes was to perform an in-depth study of one original function de-

scribed for a protein of interest. Subsequently, another direction in the study of protein properties became more prominent, in which researchers tried to identify and describe as many new functions as possible for a single protein, even when their molecular processes were sufficiently distant from each other. Therefore, a number of metabolic enzymes normally functioning in the cytoplasm of the cell were observed to shuttle into the cell nucleus under stress conditions and control gene transcription, acting as transcriptional regulators (*Fig. 7C*) [137]. Another impressive example is the ORC complex, which is responsible for recognizing the origins of replication and initiating the formation of a pre-replicative complex on DNA [138]. The ORC complex was recently shown to be involved in mRNA processing and transport from the nucleus to the cytoplasm. Many ORC subunits were shown to interact *in vivo* with processing factors, while their knockdown led to impaired mRNA transport [139, 140].

At some point, the problem of a rethinking of the available data and established views on the leading functions of some multifunctional complexes arises. It may well turn out that the initially described functional role for many regulators can only be an indirect result of their leading function, which was noticed much later. Given the exponential growth in the amount of experimental data, it is likely that we will have to go through such stages of rethinking of the hierarchy of functions for most of the known proteins. It seems to us that evolutionary research can be very helpful in this case. Obtaining information about the functional properties of proteins in non-model organisms can help trace the history of the emergence of new functions and create a hierarchy of their significance.

CONCLUSION

Eukaryotic organisms use coregulatory complexes as one of the ways to control the transcription of a certain set of genes. Thus, coregulatory transcriptional complexes may well be promising therapeutic targets for the development of drugs aimed at altering the transcription levels of a specific set of genes. Currently, there are a number of such drugs in clinical trials. The following are considered the most promising transcription coregulator targets for the development of low-molecular-weight inhibitors: the EZH2 enzymatic subunit of the PRC2 complex, the Brd4 transcription elongation coregulator, and various HDAC histone deacetylases [141–143]. The development and testing of drugs aimed at modifying the functional properties of these proteins began quite recently. Of course, the family of transcriptional regulators still harbors many other promising target proteins. ●

This study was supported with a grant of Russian Science Foundation №18-14-00219 (to N.E. Vorobyeva).

REFERENCES

1. Krasnov A.N., Mazina M.Y., Nikolenko J.V., Vorobyeva N.E. // Cell Biosci. 2016. V. 6. P. 15.
2. Rybakova K.N., Bruggeman F.J., Tomaszewska A., Moné M.J., Carlberg C., Westerhoff H.V. // PLoS Comput. Biol. 2015. V. 11. № 4. P. e1004236.
3. Wang Y., Ni T., Wang W., Liu F. // Biol. Rev. 2019. V. 94. № 1. P. 248–258.
4. Azpeitia E., Wagner A. // Front. Mol. Biosci. 2020. V. 7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7198700/.
5. Harada B.T., Hwang W.L., Deindl S., Chatterjee N., Bartholomew B., Zhuang X. // eLife. 2016. V. 5. P. e10051.
6. Yague-Sanz C., Vázquez E., Sánchez M., Antequera F., Hermand D. // Curr. Genet. 2017. V. 63. № 2. P. 187–193.
7. Rawal Y., Chereji R.V., Qiu H., Ananthakrishnan S., Govind C.K., Clark D.J., Hinnebusch A.G. // Genes Dev. 2018. V. 32. № 9–10. P. 695–710.
8. Dechassa M.L., Sabri A., Pondugula S., Kassabov S.R., Chatterjee N., Kladde M.P., Bartholomew B. // Mol. Cell. 2010. V. 38. № 4. P. 590–602.
9. Mazina M.I., Vorob'eva N.E., Krasnov A.N. // Tsitologiia. 2013. V. 55. № 4. P. 218–224.
10. Vorobyeva N.E., Mazina M.U., Golovnin A.K., Kopytova D.V., Gurskiy D.Y., Nabirochkina E.N., Georgieva S.G., Georgiev P.G., Krasnov A.N. // Nucl. Acids Res. 2013. V. 41. № 11. P. 5717–5730.
11. Chatterjee N., Sinha D., Lemma-Dechassa M., Tan S., Shogren-Knaak M.A., Bartholomew B. // Nucl. Acids Res. 2011. V. 39. № 19. P. 8378–8391.
12. Mitra D., Parnell E.J., Landon J.W., Yu Y., Stillman D.J. // Mol. Cell. Biol. 2006. V. 26. № 11. P. 4095–4110.
13. Sullivan E.K., Weirich C.S., Guyon J.R., Sif S., Kingston R.E. // Mol. Cell. Biol. 2001. V. 21. № 17. P. 5826–5837.
14. Yudkovsky N., Logie C., Hahn S., Peterson C.L. // Genes Dev. 1999. V. 13. № 18. P. 2369–2374.
15. Zhang B., Chambers K.J., Faller D.V., Wang S. // Oncogene. 2007. V. 26. № 50. P. 7153–7157.
16. McKnight J.N., Jenkins K.R., Nodelman I.M., Escobar T., Bowman G.D. // Mol. Cell. Biol. 2011. V. 31. № 23. P. 4746–4759.
17. Collins N., Poot R.A., Kukimoto I., García-Jiménez C., Dellaire G., Varga-Weisz P.D. // Nat. Genet. 2002. V. 32. № 4. P. 627–632.
18. Wysocka J., Swigut T., Xiao H., Milne T.A., Kwon S.Y., Landry J., Kauer M., Tackett A.J., Chait B.T., Badenhorst P., et al. // Nature. 2006. V. 442. № 7098. P. 86–90.
19. Eberharter A., Ferrari S., Längst G., Straub T., Imhof A., Varga-Weisz P., Wilm M., Becker P.B. // EMBO J. 2001. V. 20. № 14. P. 3781–3788.
20. Badenhorst P., Xiao H., Cherbas L., Kwon S.Y., Voas M., Rebay I., Cherbas P., Wu C. // Genes Dev. 2005. V. 19. № 21. P. 2540–2545.

21. Song H., Spichiger-Haeusermann C., Basler K. // EMBO Rep. 2009. V. 10. № 10. P. 1140–1146.
22. Siggens L., Cordeddu L., Rönnerblad M., Lennartsson A., Ekwall K. // Epigenetics Chromatin. 2015. V. 8. № 1. P. 4.
23. Smolle M., Workman J.L. // Biochim. Biophys. Acta BBA – Gene Regul. Mech. 2013. V. 1829. № 1. P. 84–97.
24. Bracken A.P., Brien G.L., Verrijzer C.P. // Genes Dev. 2019. V. 33. № 15–16. P. 936–959.
25. Krietenstein N., Wal M., Watanabe S., Park B., Peterson C.L., Pugh B.F., Korber P. // Cell. 2016. V. 167. № 3. P. 709–721.e12.
26. Udugama M., Sabri A., Bartholomew B. // Mol. Cell. Biol. 2011. V. 31. № 4. P. 662–673.
27. Willhoft O., Wigley D.B. // Curr. Opin. Struct. Biol. 2020. V. 61. P. 50–58.
28. Poli J., Gerhold C.-B., Tosi A., Hustedt N., Seeber A., Sack R., Herzog F., Pasero P., Shimada K., Hopfner K.-P., et al. // Genes Dev. 2016. V. 30. № 3. P. 337–354.
29. Lafon A., Taranum S., Pietrocola F., Dingli F., Loew D., Brahma S., Bartholomew B., Papamichos-Chronakis M. // Mol. Cell. 2015. V. 60. № 5. P. 784–796.
30. Hallson G., Hollebakken R.E., Li T., Syrzycka M., Kim I., Cotsworth S., Fitzpatrick K.A., Sinclair D.A.R., Honda B.M. // Genetics. 2012. V. 190. № 1. P. 91–100.
31. Bae H.J., Dubarry M., Jeon J., Soares L.M., Dargemont C., Kim J., Geli V., Buratowski S. // Nat. Commun. 2020. V. 11. № 1. P. 2181.
32. Tie F., Banerjee R., Saiakhova A.R., Howard B., Monteith K.E., Scacheri P.C., Cosgrove M.S., Harte P.J. // Dev. Camb. Engl. 2014. V. 141. № 5. P. 1129–1139.
33. Carbonell A., Mazo A., Serras F., Corominas M. // Mol. Biol. Cell. 2013. V. 24. № 3. P. 361–372.
34. Schurter B.T., Koh S.S., Chen D., Bunick G.J., Harp J.M., Hanson B.L., Henschen-Edman A., Mackay D.R., Stallcup M.R., Aswad D.W. // Biochemistry. 2001. V. 40. № 19. P. 5747–5756.
35. Cheng D., Côté J., Shaaban S., Bedford M.T. // Mol. Cell. 2007. V. 25. № 1. P. 71–83.
36. Bao J., Rousseaux S., Shen J., Lin K., Lu Y., Bedford M.T. // Nucl. Acids Res. 2018. V. 46. № 9. P. 4327–4343.
37. Xu W., Cho H., Kadam S., Banayo E.M., Anderson S., Yates J.R., Emerson B.M., Evans R.M. // Genes Dev. 2004. V. 18. № 2. P. 144–156.
38. Strahl B.D., Briggs S.D., Brame C.J., Caldwell J.A., Koh S.S., Ma H., Cook R.G., Shabanowitz J., Hunt D.F., Stallcup M.R., et al. // Curr. Biol. 2001. V. 11. № 12. P. 996–1000.
39. Kwak Y.T., Guo J., Prajapati S., Park K.-J., Surabhi R.M., Miller B., Gehrig P., Gaynor R.B. // Mol. Cell. 2003. V. 11. № 4. P. 1055–1066.
40. Zhang C., Robinson B.S., Xu W., Yang L., Yao B., Zhao H., Byun P.K., Jin P., Veraksa A., Moberg K.H. // Dev. Cell. 2015. V. 34. № 2. P. 168–180.
41. Le Romancer M., Treilleux I., Leconte N., Robin-Lespinasse Y., Sentis S., Bouchekioua-Bouzaghou K., Goddard S., Gobert-Gosse S., Corbo L. // Mol. Cell. 2008. V. 31. № 2. P. 212–221.
42. Tang J., Kao P.N., Herschman H.R. // J. Biol. Chem. 2000. V. 275. № 26. P. 19866–19876.
43. Sun X.-J., Wei J., Wu X.-Y., Hu M., Wang L., Wang H.-H., Zhang Q.-H., Chen S.-J., Huang Q.-H., Chen Z. // J. Biol. Chem. 2005. V. 280. № 42. P. 35261–35271.
44. Chen K., Liu J., Liu S., Xia M., Zhang X., Han D., Jiang Y., Wang C., Cao X. // Cell. 2017. V. 170. № 3. P. 492–506.e14.
45. Kizer K.O., Phatnani H.P., Shibata Y., Hall H., Greenleaf A.L., Strahl B.D. // Mol. Cell. Biol. 2005. V. 25. № 8. P. 3305–3316.
46. Govind C.K., Qiu H., Ginsburg D.S., Ruan C., Hofmeyer K., Hu C., Swaminathan V., Workman J.L., Li B., Hinnebusch A.G. // Mol. Cell. 2010. V. 39. № 2. P. 234–246.
47. Bonnet J., Wang C.-Y., Baptista T., Vincent S.D., Hsiao W.-C., Stierle M., Kao C.-F., Tora L., Devys D. // Genes Dev. 2014. V. 28. № 18. P. 1999–2012.
48. Riss A., Scheer E., Joint M., Trowitzsch S., Berger I., Tora L. // J. Biol. Chem. 2015. V. 290. № 48. P. 28997–29009.
49. Weake V.M., Workman J.L. // Trends Cell Biol. 2012. V. 22. № 4. P. 177–184.
50. Jin Q., Yu L.-R., Wang L., Zhang Z., Kasper L.H., Lee J.-E., Wang C., Brindle P.K., Dent S.Y.R., Ge K. // EMBO J. 2011. V. 30. № 2. P. 249–262.
51. Tropberger P., Pott S., Keller C., Kamieniarz-Gdula K., Caron M., Richter F., Li G., Mittler G., Liu E.T., Bühler M., et al. // Cell. 2013. V. 152. № 4. P. 859–872.
52. Han Y., Jin Y.-H., Kim Y.-J., Kang B.-Y., Choi H.-J., Kim D.-W., Yeo C.-Y., Lee K.-Y. // Biochem. Biophys. Res. Commun. 2008. V. 375. № 4. P. 576–580.
53. Bedford D.C., Kasper L.H., Fukuyama T., Brindle P.K. // Epigenetics. 2010. V. 5. № 1. P. 9–15.
54. Wang F., Marshall C.B., Ikura M. // Cell. Mol. Life Sci. CMLS. 2013. V. 70. № 21. P. 3989–4008.
55. Kim J., Hake S.B., Roeder R.G. // Mol. Cell. 2005. V. 20. № 5. P. 759–770.
56. Sun Z.-W., Allis C.D. // Nature. 2002. V. 418. № 6893. P. 104–108.
57. Van Oss S.B., Shirra M.K., Bataille A.R., Wier A.D., Yen K., Vinayachandran V., Byeon I.-J.L., Cucinotta C.E., Héroux A., Jeon J., et al. // Mol. Cell. 2016. V. 64. № 4. P. 815–825.
58. Clapier C.R., Iwasa J., Cairns B.R., Peterson C.L. // Nat. Rev. Mol. Cell Biol. 2017. V. 18. № 7. P. 407–422.
59. Mazina M.Y., Vorobyeva N.E. // Russ. J. Genet. 2016. V. 52. № 5. P. 463–472.
60. Zaret K.S., Carroll J.S. // Genes Dev. 2011. V. 25. № 21. P. 2227–2241.
61. Cirillo L.A., Lin F.R., Cuesta I., Friedman D., Jarnik M., Zaret K.S. // Mol. Cell. 2002. V. 9. № 2. P. 279–289.
62. Updike D.L., Mango S.E. // PLoS Genet. 2006. V. 2. № 9. P. e161.
63. Lupien M., Eeckhoute J., Meyer C.A., Wang Q., Zhang Y., Li W., Carroll J.S., Liu X.S., Brown M. // Cell. 2008. V. 132. № 6. P. 958–970.
64. Hurtado A., Holmes K.A., Ross-Innes C.S., Schmidt D., Carroll J.S. // Nat. Genet. 2011. V. 43. № 1. P. 27–33.
65. Kong S.L., Li G., Loh S.L., Sung W.-K., Liu E.T. // Mol. Syst. Biol. 2011. V. 7. P. 526.
66. Belandia B., Orford R.L., Hurst H.C., Parker M.G. // EMBO J. 2002. V. 21. № 15. P. 4094–4103.
67. Vicent G.P., Nacht A.S., Font-Mateu J., Castellano G., Gaveglia L., Ballaré C., Beato M. // Genes Dev. 2011. V. 25. № 8. P. 845–862.
68. Tang L., Nogales E., Ciferri C. // Prog. Biophys. Mol. Biol. 2010. V. 102. № 2. P. 122–128.
69. Torigoe S.E., Urwin D.L., Ishii H., Smith D.E., Kadonaga J.T. // Mol. Cell. 2011. V. 43. № 4. P. 638–648.
70. Li M., Hada A., Sen P., Olufemi L., Hall M.A., Smith B.Y., Forth S., McKnight J.N., Patel A., Bowman G.D., et al. // eLife. 2015. V. 4. P. e06249.
71. Manning B.J., Peterson C.L. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. № 50. P. 17827–17832.
72. Rege M., Feldman J.L., Adkins N.L., Peterson C.L. // bioRxiv. 2020. P. 2020.03.24.006205.

73. Kia S.K., Gorski M.M., Giannakopoulos S., Verrijzer C.P. // Mol. Cell. Biol. 2008. V. 28. № 10. P. 3457–3464.
74. Kim K.H., Kim W., Howard T.P., Vazquez F., Tsherniak A., Wu J.N., Wang W., Haswell J.R., Walensky L.D., Hahn W.C., et al. // Nat. Med. 2015. V. 21. № 12. P. 1491–1496.
75. Kadoch C., Williams R.T., Calarco J.P., Miller E.L., Weber C.M., Braun S.M.G., Pulice J.L., Chory E.J., Crabtree G.R. // Nat. Genet. 2017. V. 49. № 2. P. 213–222.
76. Längst G., Manelyte L. // Genes. 2015. V. 6. № 2. P. 299–324.
77. Kadoch C., Crabtree G.R. // Sci. Adv. 2015. V. 1. № 5. P. e1500447.
78. Moshkin Y.M., Mohrmann L., van Ijcken W.F.J., Verrijzer C.P. // Mol. Cell. Biol. 2007. V. 27. № 2. P. 651–661.
79. Vorobyeva N.E., Soshnikova N.V., Nikolenko J.V., Kuzmina J.L., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 27. P. 11049–11054.
80. Emelyanov A.V., Vershilova E., Ignatyeva M.A., Pokrovsky D.K., Lu X., Konev A.Y., Fyodorov D.V. // Genes Dev. 2012. V. 26. № 6. P. 603–614.
81. Torrado M., Low J.K.K., Silva A.P.G., Schmidberger J.W., Sana M., Sharifi Tabar M., Isilak M.E., Winning C.S., Kwong C., Bedward M.J., et al. // FEBS J. 2017. V. 284. № 24. P. 4216–4232.
82. Zhang W., Aubert A., Gomez de Segura J.M., Karuppasamy M., Basu S., Murthy A.S., Diamante A., Drury T.A., Balmer J., Cramard J., et al. // J. Mol. Biol. 2016. V. 428. № 14. P. 2931–2942.
83. Bornelöv S., Reynolds N., Xenophontos M., Gharbi S., Johnstone E., Floyd R., Ralser M., Signolet J., Loos R., Dietmann S., et al. // Mol. Cell. 2018. V. 71. № 1. P. 56–72.e4.
84. Kwok R.S., Li Y.H., Lei A.J., Edery I., Chiu J.C. // PLoS Genet. 2015. V. 11. № 7. P. e1005307.
85. Jordán-Pla A., Yu S., Waldholm J., Källman T., Östlund Farrants A.-K., Visa N. // BMC Genomics. 2018. V. 19. № 1. P. 367.
86. Vorobyeva N.E., Soshnikova N.V., Kuzmina J.L., Kopantseva M.R., Nikolenko J.V., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. // Cell Cycle. 2009. V. 8. № 14. P. 2152–2156.
87. Moshkin Y.M., Chalkley G.E., Kan T.W., Reddy B.A., Ozgur Z., van Ijcken W.F.J., Dekkers D.H.W., Demmers J.A., Travers A.A., Verrijzer C.P. // Mol. Cell. Biol. 2012. V. 32. № 3. P. 675–688.
88. Vorobyeva N.E., Nikolenko J.V., Krasnov A.N., Kuzmina J.L., Panov V.V., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. // Cell Cycle Georget. Tex. 2011. V. 10. № 11. P. 1821–1827.
89. Panov V.V., Kuzmina J.L., Doronin S.A., Kopantseva M.R., Nabirochkina E.N., Georgieva S.G., Vorobyeva N.E., Shidlovskii Y.V. // Nucl. Acids Res. 2012. V. 40. № 6. P. 2445–2453.
90. Jenuwein T., Allis C.D. // Science. 2001. V. 293. № 5532. P. 1074–1080.
91. Howe F.S., Fischl H., Murray S.C., Mellor J. // BioEssays. 2017. V. 39. № 1. P. e201600095.
92. Musselman C.A., Lalonde M.-E., Côté J., Kutateladze T.G. // Nat. Struct. Mol. Biol. 2012. V. 19. № 12. P. 1218–1227.
93. Yun M., Wu J., Workman J.L., Li B. // Cell Res. 2011. V. 21. № 4. P. 564–578.
94. Peña P.V., Davrazou F., Shi X., Walter K.L., Verkhusha V.V., Gozani O., Zhao R., Kutateladze T.G. // Nature. 2006. V. 442. № 7098. P. 100–103.
95. Shi X., Hong T., Walter K.L., Ewalt M., Michishita E., Hung T., Carney D., Peña P., Lan F., Kaadige M.R., et al. // Nature. 2006. V. 442. № 7098. P. 96–99.
96. Sims R.J., Chen C.-F., Santos-Rosa H., Kouzarides T., Patel S.S., Reinberg D. // J. Biol. Chem. 2005. V. 280. № 51. P. 41789–41792.
97. Morettini S., Tribus M., Zeilner A., Sebald J., Campo-Fernandez B., Scheran G., Wörle H., Podhraski V., Fyodorov D.V., Lusser A. // Nucl. Acids Res. 2011. V. 39. № 8. P. 3103–3115.
98. Lauberth S.M., Nakayama T., Wu X., Ferris A.L., Tang Z., Hughes S.H., Roeder R.G. // Cell. 2013. V. 152. № 5. P. 1021–1036.
99. Li H., Ilin S., Wang W., Duncan E.M., Wysocka J., Allis C.D., Patel D.J. // Nature. 2006. V. 442. № 7098. P. 91–95.
100. Li Y., Schulz V.P., Deng C., Li G., Shen Y., Tusi B.K., Ma G., Stees J., Qiu Y., Steiner L.A., et al. // Nucl. Acids Res. 2016. V. 44. № 15. P. 7173–7188.
101. Beisel C., Paro R. // Nat. Rev. Genet. 2011. V. 12. № 2. P. 123–135.
102. Kassis J.A., Brown J.L. // Advances in Genetics / Eds Friedmann T., Dunlap J.C., Goodwin S.F. Acad. Press, 2013. V. 81. P. 83–118. http://www.sciencedirect.com/science/article/pii/B9780124076778000038.
103. Saksouk N., Simboeck E., Déjardin J. // Epigenetics Chromatin. 2015. V. 8. № 1. P. 3.
104. Kahn T.G., Dorafshan E., Schultheis D., Zare A., Stenberg P., Reim I., Pirrotta V., Schwartz Y.B. // Nucl. Acids Res. 2016. V. 44. № 21. P. 10132–10149.
105. Francis N.J., Follmer N.E., Simon M.D., Aghia G., Butler J.D. // Cell. 2009. V. 137. № 1. P. 110–122.
106. Müller-Ott K., Erdel F., Matveeva A., Mallm J.-P., Rademacher A., Hahn M., Bauer C., Zhang Q., Kaltofen S., Schotta G., et al. // Mol. Syst. Biol. 2014. V. 10. № 8. P. 746.
107. Coulon A., Chow C.C., Singer R.H., Larson D.R. // Nat. Rev. Genet. 2013. V. 14. № 8. P. 572–584.
108. Deal R.B., Henikoff J.G., Henikoff S. // Science. 2010. V. 328. № 5982. P. 1161–1164.
109. Adelman K., Lis J.T. // Nat. Rev. Genet. 2012. V. 13. № 10. P. 720–731.
110. Gaertner B., Johnston J., Chen K., Wallaschek N., Paulson A., Garruss A.S., Gaudenz K., De Kumar B., Krumlauf R., Zeitlinger J. // Cell Rep. 2012. V. 2. № 6. P. 1670–1683.
111. Min I.M., Waterfall J.J., Core L.J., Munroe R.J., Schimenti J., Lis J.T. // Genes Dev. 2011. V. 25. № 7. P. 742–754.
112. Vastenhouw N.L., Schier A.F. // Curr. Opin. Cell Biol. 2012. V. 24. № 3. P. 374–386.
113. Kuroda M.I., Kang H., De S., Kassis J.A. // Annu. Rev. Biochem. 2020. V. 89. P. 235–253.
114. Atlasi Y., Stunnenberg H.G. // Nat. Rev. Genet. 2017. V. 18. № 11. P. 643–658.
115. Kim H.J., Kim T., Oldfield A.J., Yang P. // bioRxiv. 2020. P. 2020.10.13.338103.
116. Chen S., Ma J., Wu F., Xiong L., Ma H., Xu W., Lv R., Li X., Villen J., Gygi S.P., et al. // Genes Dev. 2012. V. 26. № 12. P. 1364–1375.
117. Boija A., Mahat D.B., Zare A., Holmqvist P.-H., Philip P., Meyers D.J., Cole P.A., Lis J.T., Stenberg P., Mannervik M. // Mol. Cell. 2017. V. 68. № 3. P. 491–503.e5.
118. Gates L.A., Shi J., Rohira A.D., Feng Q., Zhu B., Bedford M.T., Sagum C.A., Jung S.Y., Qin J., Tsai M.-J., et al. // J. Biol. Chem. 2017. V. 292. № 35. P. 14456–14472.
119. Mazina M.Yu., Kovalenko E.V., Derevyanko P.K., Nikolenko J.V., Krasnov A.N., Vorobyeva N.E. // Biochim. Biophys. Acta BBA – Gene Regul. Mech. 2018. V. 1861. № 2. P. 178–189.
120. Chandy M., Gutiérrez J.L., Prochasson P., Workman J.L.

// Eukaryot. Cell. 2006. V. 5. № 10. P. 1738–1747.
121. Qiu H., Chereji R.V., Hu C., Cole H.A., Rawal Y., Clark D.J., Hinnebusch A.G. // Genome Res. 2016. V. 26. № 2. P. 211–225.
122. Dion M.F., Altschuler S.J., Wu L.F., Rando O.J. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. № 15. P. 5501–5506.
123. Shahbazian M.D., Grunstein M. // Annu. Rev. Biochem. 2007. V. 76. № 1. P. 75–100.
124. Kim J.-H., Saraf A., Florens L., Washburn M., Workman J.L. // Genes Dev. 2010. V. 24. № 24. P. 2766–2771.
125. Daujat S., Bauer U.-M., Shah V., Turner B., Berger S., Kouzarides T. // Curr. Biol. 2002. V. 12. № 24. P. 2090–2097.
126. Ma H., Baumann C.T., Li H., Strahl B.D., Rice R., Jelinek M.A., Aswad D.W., Allis C.D., Hager G.L., Stallcup M.R. // Curr. Biol. 2001. V. 11. № 24. P. 1981–1985.
127. Chevillard-Briet M., Trouche D., Vandel L. // EMBO J. 2002. V. 21. № 20. P. 5457–5466.
128. Dorafshan E., Kahn T.G., Glotov A., Savitsky M., Walther M., Reuter G., Schwartz Y.B. // EMBO Rep. 2019. V. 20. № 4. P. e46762.
129. Hödl M., Basler K. // Curr. Biol. 2012. V. 22. № 23. P. 2253–2257.
130. Swinstead E.E., Paakinaho V., Presman D.M., Hager G.L. // BioEssays. 2016. V. 38. № 11. P. 1150–1157.
131. Barisic D., Stadler M.B., Iurlaro M., Schübeler D. // Nature. 2019. V. 569. № 7754. P. 136–140.
132. Baptista T., Grünberg S., Minoungou N., Koster M.J.E., Timmers H.T.M., Hahn S., Devys D., Tora L. // Mol. Cell. 2017. V. 68. № 1. P. 130–143.e5.
133. Chandrasekharan M.B., Huang F., Sun Z.-W. // Epigenetics. 2010. V. 5. № 6. P. 460–468.
134. Gurskiy D., Orlova A., Vorobyeva N., Nabirochkina E., Krasnov A., Shidlovskii Y., Georgieva S., Kopytova D. // Nucl. Acids Res. 2012. V. 40. № 21. P. 10689–10700.
135. Allen B.L., Taatjes D.J. // Nat. Rev. Mol. Cell Biol. 2015. V. 16. № 3. P. 155–166.
136. Fant C.B., Taatjes D.J. // Transcription. 2018. https://www.tandfonline.com/doi/abs/10.1080/21541264.2018.1556915.
137. van der Knaap J.A., Verrijzer C.P. // Genes Dev. 2016. V. 30. № 21. P. 2345–2369.
138. Hoggard T., Fox C.A. //The Initiation of DNA Replication in Eukaryotes / Ed. Kaplan D.L. Cham. Springer Internat. Publ. 2016. P. 159–188. https://doi.org/10.1007/978-3-319-24696-3_9.
139. Kopytova D., Popova V., Kurshakova M., Shidlovskii Y., Nabirochkina E., Brechalov A., Georgiev G., Georgieva S. // Nucl. Acids Res. 2016. V. 44. № 10. P. 4920–4933.
140. Popova V.V., Brechalov A.V., Georgieva S.G., Kopytova D.V. // Nucleus. 2018. V. 9. № 1. P. 460–473.
141. Fioravanti R., Stazi G., Zwergel C., Valente S., Mai A. // Chem. Rec. 2018. V. 18. № 12. P. 1818–1832.
142. Laubach J.P., San-Miguel J.F., Hungria V., Hou J., Moreau P., Lonial S., Lee J.H., Einsele H., Alsina M., Richardson P.G. // Expert Rev. Hematol. 2017. V. 10. № 3. P. 229–237.
143. Waring M.J., Chen H., Rabow A.A., Walker G., Bobby R., Boiko S., Bradbury R.H., Callis R., Clark E., Dale I., et al. // Nat. Chem. Biol. 2016. V. 12. № 12. P. 1097–1104.

CTCF As an Example of DNA-Binding Transcription Factors Containing Clusters of C2H2-Type Zinc Fingers

O. G. Maksimenko[1,2*], D. V. Fursenko[1], E. V. Belova[1,2], P. G. Georgiev[1*]
[1]Institute of Gene Biology RAS, Moscow, 119334 Russia
[2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology RAS, Moscow, 119334 Russia
*E-mail: maksog@mail.ru; georgiev_p@mail.ru
Received September 17, 2020; in final form, November 12, 2020
DOI: 10.32607/actanaturae.11206

ABSTRACT **In mammals, most of the boundaries of topologically associating domains and all well-studied insulators are rich in binding sites for the CTCF protein. According to existing experimental data, CTCF is a key factor in the organization of the architecture of mammalian chromosomes. A characteristic feature of the CTCF is that the central part of the protein contains a cluster consisting of eleven domains of C2H2-type zinc fingers, five of which specifically bind to a long DNA sequence conserved in most animals. The class of transcription factors that carry a cluster of C2H2-type zinc fingers consisting of five or more domains (C2H2 proteins) is widely represented in all groups of animals. The functions of most C2H2 proteins still remain unknown. This review presents data on the structure and possible functions of these proteins, using the example of the vertebrate CTCF protein and several well- characterized C2H2 proteins in Drosophila and mammals.**
KEYWORDS **C2H2-type zinc fingers, architectural proteins, transcription regulation, insulators, TAD, enhancers, promoters, CTCF.**
ABBREVIATIONS **kbp – kilobase pairs; TAD – topologically associating domain; CTCF – CCCTC-binding factor; C2H2 – a domain consisting of two cysteine and two histidine residues coordinated to a zinc ion.**

INTRODUCTION

Cell differentiation in higher eukaryotes has led to significant complication in the regulation of gene expression. Cell specialization is determined by transcription factor repertoires assembling on regulatory elements of the genome. The genes responsible for cell differentiation are usually regulated by enhancers, each of which activates a promoter in a particular group of cells for a specific time interval [1–3]. In some cases, transcription of the developmental genes is regulated by several dozens of enhancers; the distance between some of these enhancers and the regulated promoter can reach up to several hundred kilobase pairs.

The ability of enhancers to perform long-range stimulation of promoters has led to the assumption that there may be some specialized transcription domains within which contacts between enhancers and promoters occur more efficiently [4]. It was believed that at the boundaries of transcription domains there are special regulatory elements capable of blocking interactions between enhancers and promoters [5, 6]. The most common opinion was that domain boundaries interact either with each other or with the nuclear structures bound to the nuclear envelope. Indeed, regulatory elements with the predicted properties were found first in Drosophila and then in mammals; these elements became known as insulators [7]. The

two main properties of insulators have been described using the model systems in transgenic Drosophila lines: their ability to block the enhancer–promoter contacts and that to prevent repression of transgene expression during its integration into the heterochromatin regions within the genome [5, 6].

The emergence of methods for genome-wide identification of contacts between chromatin regions *in vivo*, and high-resolution microscopy [8–11], took the study of the spatial organization of the genome to a completely new level. It turns out that the chromosomes of all eukaryotes are organized into topologically associating domains (TADs), which are formed through predominant interaction between the ends or boundaries of the domains [12–15]. In this case, contacts within a TAD form much more efficiently than contacts between sequences located in adjacent TADs.

The discovery of TADs gave grounds to assume that their boundaries correspond to the insulators that restrict independent regulatory domains [16–18]. However, studies carried out on single cells have shown that TAD boundaries form as a set of preferred contacts and are not strict physical barriers blocking any trans-interactions between regulatory elements located in different TADs [12, 14, 19, 20]. Most of the characterized insulators are located within the same TAD. The improvement in the resolution of contact maps within the TAD has led to the discovery of subdomains, which usually correspond to local contacts between regulatory elements [19].

CTCF AS THE BEST STUDIED PROTEIN WITH A C2H2 ZINC FINGER CLUSTER IN MAMMALS

The vertebrate protein CTCF (CCCTC binding factor), which has been well-studied in humans and mice [21, 22], is expressed at all ontogenetic stages in all cell types and is required during embryogenesis. Depending on the context, CTCF can act as a transcriptional activator or repressor. It is involved in the inactivation of one of the X chromosomes in mammals, it regulates alternative splicing of pre-mRNA in some genes, controls imprinting, participates in recombination and repair, and is responsible for the activity of enhancers, promoters, and insulators. However, the key role played by vertebrate CTCF in the chromosomal architecture is what has been described most thoroughly [23–25]. Mammalian genomes contain from 40,000 to 80,000 CTCF binding sites, with over 5,000 sites being conserved in different species and cell lines [21, 26]. Approximately 50%, 15%, and 35% of the CTCF binding sites are located, respectively, in intergenic regions, near promoters, and within gene bodies (30% residing in introns and 5% residing in exons) [27]. Mammalian CTCF consists of non-structured terminal regions and eleven zinc fingers residing in the central part of the protein; the first ten zinc fingers are C2H2-type, and the last one is C2HC-type. It is worth noting that proteins containing one or, less frequently, several clusters of C2H2 zinc finger domains constitute a significant portion of all the C2H2 zinc finger proteins [28]. The classical C2H2 domain has the consensus sequence $CX_{2-4}CX_{12}HX_{2-8}H$. In the presence of a zinc ion, this sequence folds to form a ββα structure, where zinc is tetrahedrally coordinated by two cysteine residues at one end of the β-sheet and two histidine residues at the C-ends of α-helices. The structure is stabilized by hydrophobic bonds. In the canonical complex, the α-helical sections of tandem C2H2 zinc fingers are located in the major groove of DNA. The high-affinity specific binding is ensured by specific interactions with nitrogenous bases and nonspecific contacts with the phosphate backbone of DNA. For any DNA triplet, it is possible to choose a C2H2 domain carrying the desired amino acids at key positions of the α-helix and specifically recognizing this triplet [29–31]. Therefore, just within a few years after the first description of the structure of the C2H2 domains bound to DNA, chimeric proteins consisting of a C2H2 domain cluster and the FokI domain introducing double-strand breaks in the DNA sequence started being actively used as site-specific endonucleases for targeted genome editing [32, 33].

In proteins containing a C2H2-type zinc finger cluster, short 5-aa linkers residing between the domains possess the consensus sequence TGEKP and are a characteristic feature of DNA-binding C2H2 proteins [34]. The linkers are critical in terms of the affinity and specificity of DNA binding; mutations in them may cause a loss of the protein's function *in vivo* [35, 36]. It is believed that each amino acid residue within the linker plays its own role in the interaction with DNA. Flexible in its unbound state, the protein structure, consisting of several C2H2 domains, "latches itself" as soon as it binds to the correct DNA sequence. The OH group of the first threonine residue T1 (or serine residue) forms a hydrogen bond with the amide group of glutamic acid E3; glycine G2 ensures the flexibility of the main chain required for latching. Glutamic acid E3 can contribute to the stabilization of the contacts between the zinc fingers. The lysine residue K4 (or arginine residue) is in contact with the DNA phosphate backbone. Proline residue P5 probably strengthens the bond between the linker and the following zinc finger; it also immobilizes the following conserved phenylalanine or tyrosine residue, whose aromatic ring lies on the N-end of the α-helix [37]. TGEKP-like linkers also connect the DNA-binding C2H2 domains of human CTCF *(Fig. 1)*.

The conformational changes in the DNA structure introduced by the C2H2 domains during binding limit the potential number of C2H2 domains connected by short linkers and capable of cooperatively interacting with DNA, and, therefore, limit the length of the canonical binding site [37]. This is probably why only four or five C2H2 domains are involved in the interaction and specific recognition of a 12–15 bp long DNA site in most proteins. Studies with artificial C2H2 clusters have shown that the specificity of protein binding to DNA increases when several short DNA-recognizing C2H2-domain clusters are connected by longer non-canonical linkers [28]. Therefore, it can be assumed that proteins carrying a large number of C2H2 domains in a cluster can specifically recognize different DNA sequences.

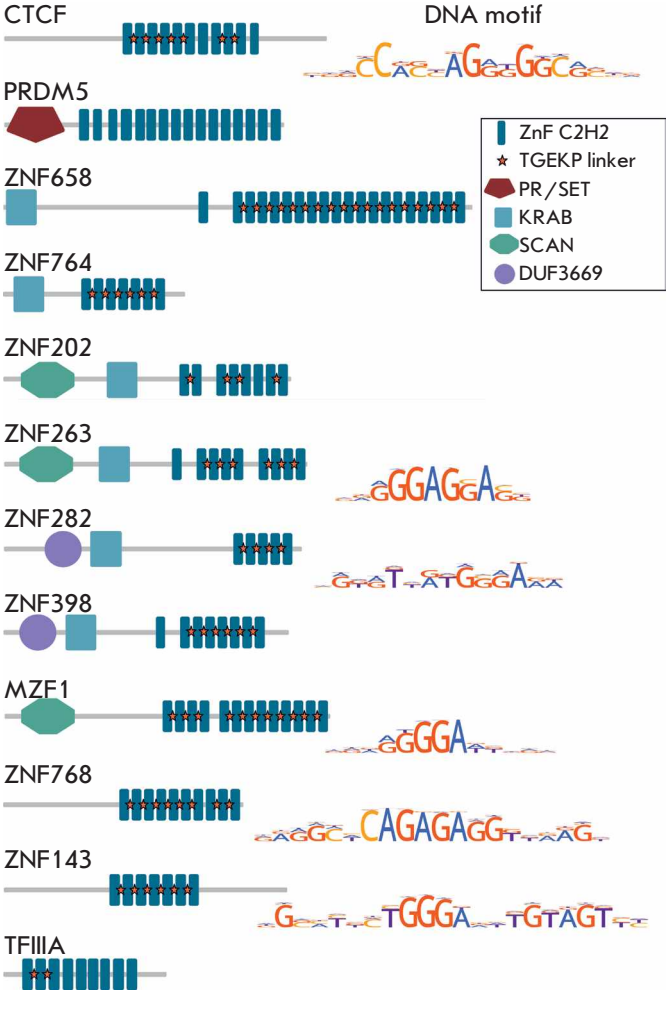


Fig. 1. C2H2 proteins of vertebrates with architectural functions. The domain organization of the described proteins and the known binding motifs are shown

In human CTCF, the C2H2 domains **3–7** are responsible for specific binding to the 15-bp consensus motif (*Fig. 1*) [38]. The C2H2 domain **8** lies outside the major groove and is not involved in the recognition of DNA nitrogenous bases; therefore, it can act as a bridge connecting the C2H2-domains **3–7** recognizing the key motif with the C2H2 domains **9–11**, which can specifically bind to an additional DNA motif that is found in approximately 15% of CTCF binding sites [39, 40]. The C2H2 domains **1–2** can also bind to a non-conserved DNA sequence [39]. Thus, different combinations of C2H2 domains of CTCF can bind to a broad range of motifs with different levels of efficiency [41, 42].

It has been shown *in vitro* that CTCF–DNA binding is inhibited by cytosine methylation at position 2 in the consensus site, whereas cytosine methylation at position 12 has almost no effect. The cytosine at position 2 is recognized by the aspartic acid residue, which prefers the unmethylated base. At position 12, the cytosine is recognized by a glutamic acid residue, with the binding affinity slightly increasing in the case of a methylated base. [38]. Moreover, an important role in methyl group recognition is played by the arginine residue that forms the 5-methylcytosine–arginine–guanine triad in a complex with DNA; this triad is found in all C2H2 protein complexes with methylated DNA [43, 44].

Cytosine methylation at binding sites can enhance, weaken, or completely inhibit the binding of C2H2 proteins to DNA; i.e., it is a global mechanism of regulation of the promoter, enhancer, and insulator activities [45]. The most striking example of the role played by the methylation of binding sites for C2H2 proteins is the participation of CTCF in genomic imprinting, an epigenetic mechanism for regulating the expression of alleles of the same gene depending on their parental origin (male or female) [46]. Imprinting occurs with the participation of special regulatory elements known as differentially methylated regions (DMRs), which often contain CTCF binding sites. Imprinting has been described most thoroughly for the *Igf2* and *H19* genes, which are activated by a group of adjacent enhancers. The DMR acting as an insulator resides between the *Igf2* gene and the enhancers; it consists of four CTCF binding sites carrying a cytosine residue at position 2. DMR methylation is maintained only in the paternal *Igf2/H19* locus, resulting in a loss of CTCF binding and activation of the *Igf2* gene. Meanwhile, in the maternal locus, CTCF binds to the corresponding sites in the DMR, thus inhibiting the interaction between the enhancers and the *Igf2* gene. Methyla-

tion of the binding sites for transcription factors (and C2H2 proteins in particular) can also be involved in the global inactivation of transcription in one of the two X chromosomes in mammals [47].

C2H2-domain clusters can participate in specific and non-specific interactions with RNA [48, 49]. Specific interaction between the TFIIIA protein and 5S RNA has been the one studied most thoroughly. It was shown that the C2H2 domains **1**–**3**, **5**, and **7**–**9** bind to DNA motifs in the promoter region of the 5S RNA gene, while the C2H2 domains **4**, **5**, and **6** interact with 5S RNA. Therefore, the C2H2 domains **4** and **6** act as linkers broadening the TFIIIA protein–DNA binding capacity. At the same time, specific interaction of these C2H2 domains with newly synthesized 5S RNA is necessary for its stabilization during export from the nucleus to the cytoplasm, prior to ribosome assembly.

Two C2H2 domains, **1** and **10**, are responsible for the nonspecific interaction between CTCF and a broad range of RNAs [50, 51]. Interestingly, the disruption of the C2H2-domain structure caused by mutation in histidine does not affect RNA binding. This finding suggests that individual amino acids in the C2H2 domains play an important role in RNA binding, rather than the structure of the zinc finger itself. There are experimental data showing that interaction between CTCF and RNA may cause protein multimerization, but the mechanism of this process remains unknown [50, 52]. Since a large number of CTCF sites reside inside the introns of genes, it can be expected that CTCF is involved in the regulation of pre-mRNA splicing and termination (the processes running concomitantly with transcription) by non-specifically binding to RNA. For example, CTCF can slow down the movement of RNA polymerase II, leading to selection of either an alternative exon during splicing [53, 54] or an alternative polyadenylation signal during transcription termination [55]. A domain capable of interacting with RNA polymerase II has been mapped to the C-terminal domain of the CTCF protein, which can also be involved in the slowing-down of the movement of RNA polymerase II when passing through the CTCF binding sites [56].

A large body of experimental data shows that individual C2H2 domains or their clusters are involved in protein–protein interactions [34]. However, the detailed mechanisms behind these processes and their specificity remain poorly studied. C2H2 domains often interact with the complexes that are involved in chromatin remodeling and histone modification. According to data obtained through mutational analysis, any amino acids within the C2H2 domains and linkers connecting them can participate in these interactions (unlike during DNA binding). Therefore, it can be assumed that, in some cases, even the C2H2 domains associated with DNA can participate in the recruitment of regulatory complexes to chromatin.

A cluster of C2H2 domains is the only conserved part of the CTCF protein that shares high homology in most vertebrates, insects, and some nematodes [57–59]. The CTCF protein is not found in plants, yeast, or roundworms. The distribution of CTCF binding sites in the genome is also characterized by a certain degree of conservatism. In particular, CTCF binding sites are found at the boundaries of the regulatory domains of the homeotic genes in mammals, fish, and Drosophila [60, 61], where CTCF performs an insulator function and delimits the regions where enhancers residing in the adjacent domains perform their function [62–66]. It is worth mentioning that the CTCF binding sites are located in the repetitive elements of mammalian genomes, which could be the starting point for the evolutionary expansion of the CTCF binding sites in the intergenic regions where TAD boundaries are located [26, 67].

Despite the absence of homologous regions, the N-terminal domains of the CTCF proteins in nine animal species belonging to different classes are represented by unstructured homodimerization domains [68]. Deletion of the dimerized domain within Drosophila CTCF significantly reduces the functional activity of the mutant CTCF [69]. It was discovered in mouse embryonic stem cells that the N-terminal domain is involved in the specific binding of CTCF to the respective sites [70]. A YxF motif was found between the N-terminal homodimerization domain and the C2H2 cluster, which is necessary for interaction with the SA2–Scc1 cohesin subcomplex [71]. A similar motif was also found in the CTCF of other animal species. Therefore, although there is no significant homology, the N-terminal domains of the CTCF proteins in different species share characteristic structural features. A region interacted with the SA2 subunit of the cohesin complex was previously mapped on the C-end of the CTCF protein *in vitro* [72]; however, a more recent study has failed to confirm this finding [71].

The roles played by C2H2 proteins largely depend on the proteins with which they interact. More than 90 potential CTCF partner proteins have been identified [73, 74]. However, the mechanisms and specificity of these interactions remain disputable. Most protein–protein interactions are found within the cluster of C2H2 domains and in the unstructured C-terminus of the CTCF protein. Many different C2H2 proteins can potentially interact with the same protein complexes through the C2H2 domains. CTCF was shown to interact directly with the catalytic subunit Brg1

of the SWI/SNF chromatin remodeling complex [74] and the general transcription factor II-I (TFII-I) [75]. Therefore, the most probable function of CTCF in the promoter regions of actively transcribed genes is participation in the formation of open chromatin regions through the recruitment of the SWI / SNF complex, which increases the mobility of nucleosomes. CTCF can also be involved in stabilization on promoters of the TFIID complex (TFII-I being a part of this complex). When CTCF is inactivated, the expression level drops significantly only in the genes whose promoter regions contain CTCF binding sites [76]. Thus, one of the key functions of CTCF consists in organizing active promoters. Interestingly, like many other C2H2 proteins, CTCF contains regions enriched in proline and acidic amino acids, which is typical of the transcription activators recruiting transcription complexes to chromatin.

A domain interacting with DEAD box RNA helicases was identified in the C-terminus of CTCF [74, 77], which may be related to the potential significant participation of CTCF in the regulation of splicing and transcription termination. To perform these functions, the found interaction of CTCF with topoisomerase II (Top2) is probably also needed [78]. Top2 regulates chromatin topology by introducing ATP-dependent double-strand breaks into DNA. The Top2 protein has been found in approximately half of all CTCF binding sites [78]. Top2 activity is most often observed in the close vicinity of CTCF binding sites [79]. It is thought that Top2 is recruited to the open chromatin regions that form at CTCF sites, and that direct protein-protein interactions enhance this process. Possibly, CTCF helps recruit Top2 to the introns and 3'-ends of genes, which might be required during gene transcription.

The activity of C2H2 proteins is regulated by various post-translational modifications. Phosphorylation of C2H2 proteins at the linkers between the C2H2 domains, which occurs during mitosis and reduces the efficiency of protein binding to chromatin, has been studied quite thoroughly [80–83]. C2H2 proteins can also undergo further modifications, such as ubiquitination, SUMOylation, and poly-ADP-ribosylation [84]. The ribosylation site resides at the N-terminus of CTCF [85]; this modification can affect protein dimerization and its binding to the cohesin complex. Poly-ADP-ribosylation affects the localization of the CTCF protein in nuclear compartments, chromatin binding, and transcription regulation [85–87]. Interestingly, the N-terminus of human CTCF interacts with the C-terminus of nucleophosmin 1 (NPM1), which can be responsible for CTCF localization within the cell [88]. Sites for covalent attachment of the SUMO protein through lysine were found in the C-terminal domains of the CTCF protein [89]. The Pc2 protein belonging to the Polycomb group of transcriptional repressors was identified as a SUMO E3 ligase for CTCF. Within cell nuclei, CTCF and Pc2 are found in bodies enriched in Polycomb-group proteins.

It is assumed that by interacting with various proteins and forming homopolymers, SUMO catalyzes the formation of dense intranuclear protein structures (bodies) that can perform many functions, including being a source of spare proteins during chromatin formation on newly synthesized DNA during replication [90, 91]. SUMOylation of CTCF on chromatin can also regulate the recruitment of transcriptional complexes to chromatin, thus changing the properties of CTCF during the activation or repression of gene transcription.

As a member of the C2H2 protein family, CTCF has typical structural features: it contains a cluster of zinc fingers that provides specific binding to genomic targets and interacts with RNA and proteins, as well as terminal domains that are involved in the organization of the architecture of chromosomes and the recruitment of various regulatory complexes to chromatin.

CTCF IN ORGANIZATION OF THE CHROMOSOME ARCHITECTURE AND INSULATION IN VERTEBRATES

The CTCF protein was initially considered as the main vertebrate insulator protein [92]. The first vertebrate insulator was reported to be located at the boundary of the heterochromatin region and the chicken β-globin gene cluster [93, 94]. The insulator, with its core being 275-bp long, was mapped in the DNase 1 hypersensitive site and was therefore named HS4 [95]. In transgenic cell model systems, two copies of the HS4 insulator can effectively block enhancer activity and protect transgene expression from repression by surrounding chromatin. In addition to the binding site for the CTCF protein, the HS4 insulator was found to contain binding sites for USF1/USF2 proteins [96] and three binding sites for the VEZF1 (vascular endothelial zinc finger 1) protein [97]. It has been demonstrated that CTCF is required to block enhancers and recruit USF1/USF2 proteins, which in turn, recruit the complexes responsible for chromatin remodeling and histone modification. As a result, the nucleosomes around the HS4 insulator are enriched by nucleosome modifications associated with active chromatin (histone H3 methylated at lysine 4 and acetylated histones H3 and H4).

The VEZF1 protein contains a cluster consisting of six C2H2 domains, and it predominantly binds to active promoters [98]. Inactivation of the VEZF1 binding sites on the HS4 insulator in transgenic cell

lines enhances DNA methylation at the promoter of a reporter gene [97]. It is assumed that VEZF1 recruits a complex that performs DNA demethylation, thereby facilitating the recruitment of transcription factors (which cannot efficiently bind to methylated sites) to the HS4 insulator and the adjacent regulatory elements. Thus, the HS4 insulator is a combination of the binding sites of at least two C2H2 proteins that function in close cooperation with each other.

Despite numerous examples illustrating the key role of CTCF sites in the organization of the boundaries of regulatory domains and the insulation of enhancers [23], the question remains as to what role is played by other unknown proteins whose binding to a particular regulatory element depends on the presence of CTCF. For example, in mammals, a large number of CTCF-dependent insulators block the spread of Polycomb-dependent heterochromatin, which is associated with H3K27me3 enrichment of long chromatin regions. However, inactivation of CTCF does not cause the propagation of the H3K27me3 modification into transcriptionally active regions, which suggests that other proteins are present at the boundaries that block the spreading of repressive chromatin and thereby mask the absence of CTCF [76]. Therefore, CTCF-dependent insulators, the boundaries of the regulatory domain, and TADs are most likely to consist of CTCF sites, in combination with the binding sites for other transcription factors (including C2H2 proteins that have not been described yet).

In our current understanding, which is supported by plenty of experimental data, mammalian CTCF forms chromatin loops, in cooperation with the cohesin complex, and defines the boundaries of most TADs [19, 99]. The cohesin complex is involved in mitosis, meiosis, and the regulation of gene expression [100, 101]. This complex consists of the SMC1, SMC3, and SCC1 (Rad21) proteins, forming a ring structure and binding to the fourth subunit that exists as two isoforms, STAG1 (SA1) and STAG2 (SA2), through SCC1. It has been hypothesized that SA1 and SA2 can determine the location of the cohesin complex at different chromatin sites. The NIPBL/MAU2 and WAPL complexes catalyze the ATP-dependent binding of the cohesin complex to chromatin and its subsequent dissociation, respectively [100].

Depending on the antibodies used and the cell line under study, the colocalization of CTCF and cohesin sites varies from 40% to 95% [102–104]. Inactivation of CTCF leads to a redistribution of the cohesin complexes from the CTCF binding sites to the promoters of active genes, accompanied by partial destruction of TADs [76]. Inactivation of the subunits of the cohesin complex or the Nipbl protein [105, 106], which ensures the recruitment of the cohesin complex to chromatin, leads to an almost complete disappearance of TADs. On the contrary, inactivation of the factors that negatively affect cohesin binding to chromatin stabilizes TADs and the long-range interactions in chromatin [106]. Finally, mutations and deletions in CTCF, disrupting its interaction with the cohesin complex, also significantly disturb the formation of long-range contacts and TADs [71, 104]. The Smc1 and Smc3 subunits contain ATPase domains, and the energy of ATP cleavage is required at the stages of binding and dissociation of the cohesin complex [107, 108]. Mutations in the subunits of the cohesin complex, which disrupt ATP hydrolysis, affect long-range contacts and the formation of TAD in chromosomes [109].

CTCF sites at the TAD boundaries usually have a convergent orientation [8, 110]. The convergent orientation of CTCF motifs was shown to define which pairs of CTCF sites preferentially stabilize DNA loops [8, 110–112]. A loop extrusion model has been proposed to explain why chromatin loops preferentially form between CTCF sites with a convergent orientation. According to this model, after being loaded onto chromatin, the cohesin complex triggers DNA extrusion and chromatin loop formation. CTCF can inhibit the movement of the cohesin complex only if its N-terminal domain, which interacts with the SA2–SCC1 subcomplex [71], is correctly oriented relative to the cohesin sliding complex.

The model postulates that the cohesin complex can induce chromatin extrusion and chromatin loop formation, either actively (using ATP energy) or passively. Indeed, *in vitro* studies have shown that in the presence of Nipbl and ATP molecules, the cohesin complex binds to DNA and slides along, causing loop formation [113], even if DNA is nucleosome-bound [114]. Cohesin can also bypass small nucleosome-sized protein complexes, but it is unable to overcome complexes larger than 13 nm in diameter; such complexes with a motor function can move cohesin themselves [115]. Therefore, the convergent CTCF sites limit the extrusion regions of chromatin loops, while the loop formation is performed by molecular motors.

According to the polymer model, TAD formation strongly depends on the physical properties of chromatin, which tends to form domains of the same type. This model has support in Drosophila, where TADs form through electrostatic inter-nucleosomal interactions. As a result, the TADs boundaries are predominantly composed of long, highly transcribed open chromatin regions and the inner regions of TADs are denser chromatin structures [13, 19, 116, 117]. In this model, the role of CTCF is to recruit cohesin complexes, stabilizing the interactions between chromatin

sites that are already in close vicinity to each other. However, this model does not explain why chromatin loops in mammals predominantly form only convergently oriented CTCF binding sites.

The experimental data [107, 113, 118] show that the size of a chromatin loop is independent of the time of cohesin–DNA binding but depends on the barriers limiting its sliding (similar to CTCF). CTCF binds dynamically to chromatin, which is consistent with the heterogeneity of the TAD boundaries observed in studies of single cells [20]. The CTCF binding sites at the TAD boundaries are usually represented by clusters, which probably ensure CTCF binding to the genomic targets for a longer time [119].

According to the loop extrusion model, the cohesin complexes are only transiently blocked at a certain CTCF site and can continue to pull chromatin after crossing the block created by CTCF or after CTCF leaves chromatin [20]. Inactivation of Wapl stabilizes the binding of the cohesin complexes to chromatin; the size of chromatin loops increases, which is explained by longer residence time of the cohesin complex on chromatin [106, 120, 121].

Mitosis is characterized by chromosome condensation associated with large-scale chromatin changes and the loss of binding of some transcription factors to DNA. During mitotic prophase, most of the cohesin leaves the chromosomes (except for cohesin associated with centromeres). In anaphase, cohesin dissociation caused by separase promotes the segregation of sister chromatids [101]. The structure of TADs on compact mitotic chromosomes is lost almost completely but is quickly restored by the mid-G1 stage [122]. The data on the binding of CTCFs to the respective sites on mitotic chromosomes are inconsistent. According to some estimates, CTCF remains on 18.6% of the sites [122]; however, the binding of CTCF to its sites is mostly lost, since there is phosphorylation of linkers between the C2H2 domains [123]. It is possible that by leaving its binding sites, CTCF contributes to a more efficient removal of the cohesin complexes from mitotic chromosomes. However, CTCF binding sites are rapidly restored after mitosis, which may result from the association between free CTCF and condensed chromosomes during mitosis [123]. It remains an open question how the effective restoration of CTCF binding to the corresponding sites after mitosis occurs. It is most likely that other transcription factors remain associated with mitotic chromosomes and maintain the partially open chromatin state (act as bookmarks), which facilitates CTCF binding to the respective sites after mitosis. As a result, both the CTCF binding profile and the structure of TADs on duplicated chromosomes are rapidly restored after DNA replication. It can be assumed that excess CTCF accumulates in specialized nuclear compartments (bodies) stabilized by SUMO [89]. During replication, the excess CTCF binds to an increasing number of sites on duplicating chromosomes.

ARCHITECTURAL FUNCTIONS OF OTHER VERTEBRATE C2H2 PROTEINS

Studies focused on designing artificial C2H2-type zinc fingers that ensure specific interaction with a particular genomic target have shown that the specificity of DNA binding increases dramatically for the cluster consisting of five properly organized zinc fingers. Therefore, in this section we would like to discuss proteins with this structural organization (with at least five C2H2-type domains separated by a typical 6-bp linker sequence) as the most promising architectural C2H2 proteins.

Other C2H2 proteins remain relatively less well studied than CTCF [124, 125]. The key problems in studying this class of proteins are associated with a significant overlap of functions between different C2H2 proteins and the lack of high-quality specific antibodies against these proteins, which would make it possible to perform genome-wide studies to identify the binding sites for C2H2 proteins and their role in the maintenance of long-range contacts between regulatory elements and the formation of the chromosome architecture. Two studies [126, 127] have focused on binding sites for the 60 and 221 C2H2 proteins with GFP or HA epitope tags in HEK293T cells. The binding sites for the same C2H2 proteins studied in both publications were found to overlap only slightly [128]. It should be noted that in these studies, expression of tagged C2H2 proteins occurred in the presence of the endogenous C2H2 proteins; so, most actual binding sites were occupied by the native protein, while the tagged protein was bound (mostly non-specifically) to the domains within open chromatin regions. In the near future, the use of the CRISPR/Cas9 genome editing tool could make it possible to replace endogenous genes with modified ones expressing tagged variants of C2H2 proteins, which will simplify genome-wide studies.

Mammalian genomes are enriched in various types of repetitive sequences of differing nature, including mobile elements and retroviruses [129]. Most of the studied C2H2 proteins, including CTCF, have binding sites in different repetitive sequences [130–133]. There are many examples of repeating sequences becoming part of genetic regulatory networks and TAD boundaries [134], thereby significantly expanding the possibilities of fine-tuning gene expression during evolution.

Approximately half of all C2H2 proteins carry another domain at their N-terminus. The two most evolutionarily ancient domains that are found in all eukaryotes include the PR/SET domain (e.g., Prdm5 protein (*Fig. 1*)), which typically exhibits methyltransferase activity [135], and the BTB domain that forms dimers and recruits transcription regulators to the genomic targets [136]. One of the most numerous groups of C2H2 proteins in mammals carries the KRAB domain at their N-terminus (e.g., ZNF658 and ZNF764 proteins (*Fig. 1*)). It is believed that this domain has become widespread in mammalian C2H2 proteins, due to its repressor function with respect to mobile elements. However, in parallel with the evolution of gene regulatory systems into which mobile elements are integrated, KRAB-C2H2 proteins acquire new functions in the regulation of host gene expression [130, 131]. Some of these C2H2 proteins containing the KRAB domain carry an additional domain, SCAN (e.g., ZNF202 and ZNF263 proteins) or DUF3669 (e.g., ZNF282 and ZNF398 proteins), at their N-terminus [137–139]. Some C2H2 proteins carry only the SCAN domain (e.g., MZF1) and derive from proteins that have lost their KRAB domain. It can be assumed that some functions of the SCAN-C2H2 and DUF3669-C2H2 proteins are associated with the ability of SCAN and DUF3669 to form homo- and heterodimers [131, 137].

The most thoroughly described functions of C2H2 proteins are the formation of an open chromatin region on promoters and recruitment of transcription complexes for transcriptional activation or repression. The ZNF658 protein binds to the regulatory element residing next to 3525 promoters and participates in the activation of the expression of the rRNA genes transcribed by RNA polymerase I [140, 141]. The ZNF764 protein is ubiquitously expressed; it is involved in the regulation of glucocorticoid, androgen, and thyroid hormones activity [142]. It is interesting that the binding sites, which predominantly reside in intergenic regions (60%) and introns (31%), significantly (37%) co-localize with the binding sites for glucocorticoid receptors (GRs) [143]. It has been experimentally proved that the KRAB domain of the ZNF764 protein directly interacts with the LBD domain of a GR, suggesting that these proteins bind cooperatively to the regulatory regions. The protein-binding sites of ZNF202 [126, 144], ZNF263 [145], MZF1 [146], ZNF768 [133], and Prdm5 [147] are predominantly located in the promoter regions of genes, indicating that these factors may contribute to the activation or repression of transcription, and promoter architecture.

The N-terminus of ZNF768 (*Fig. 1*) contains 15 heptad repeats that are similar to the C-terminal domain of RNA polymerase II [133] and are presumably involved in the recruitment of the transcription elongation complex to promoters.

It has been demonstrated that MZF1 can form heterodimers with other SCAN-containing proteins (ZNF24, ZNF174, and ZNF202) through the SCAN domain [148, 149]. The ZNF282 and ZNF398 proteins form homo- and heterodimers through the DUF3669 domain [150] and can bind to the promoters in a combinatorial manner [126]. The Prdm5 protein contains an N-terminal PR/SET domain that has lost its methylation ability and is possibly involved in protein–protein interaction [151, 152].

The protein ZNF143 (*Fig. 1*), which is crucial to the embryonic development of mammals, has been characterized thoroughly [153]. Its central part contains a cluster consisting of seven C2H2 domains. The N-terminal domain contains three 15-aa repeats separated by 10-aa to 12-aa spacers [154]. The C-terminal domain is enriched in acidic amino acids, which is typical for transcriptional activators. The ZNF143 binding sites reside within a region of approximately 2,000 promoters regulated by RNA polymerases II and III [155–158]. The functional activity of ZNF143 near the promoters is related to the formation of open chromatin regions and its involvement in the recruitment of transcription activation complexes [159–161]. The ZNF143 protein has two partially overlapping consensus binding sites with the same core CCCAGA sequence [155], which can be explained by the different contributions of individual C2H2 domains to the recognition of two site variants. Genome-wide studies have shown that the ZNF143 protein may be involved in the formation of chromatin loops between enhancers and promoters [155, 156, 162-164].

A relatively large percentage of the binding sites of the Prdm5 and ZNF143 proteins colocalize with CTCF [143, 152, 163]. The Prdm5 protein has been found in association with cohesin and CTCF [152]. It was shown for HEK293T cells that inactivation of ZNF143 disrupts some CTCF-dependent chromatin loops [163]. However, there are no experimental data proving that ZNF143 (unlike CTCF) can be involved in the localization of the cohesin complex on chromatin.

Another example of the structural function of C2H2 proteins was observed when studying the chromatin architecture organized by the TFIIIC complex. It was found that the binding sites of the TFIIIC complex colocalize with condensins and can act as boundaries between active chromatin and heterochromatin, as well as maintain distant interactions and participate in the formation of the chromosome architecture [165]. Interestingly, the binding sites for Prdm5, CTCF/cohesin, and ZNF143 proteins are adjacent to or colo-

calize with the TFIIIC binding regions [152, 155, 166], suggesting that these proteins are cooperatively involved in the organization of TFIIIC-dependent regulatory elements. Furthermore, the Prdm5 protein has been isolated in complex with TFIIIC, suggesting that Prdm5 participates in the recruitment of the TFIIIC complex to chromatin [152].

The above-mentioned TFIIIA is the second (after CTCF) best-studied C2H2 protein, which binds to Pol III-dependent promoters of the genes encoding 5S rRNA in all eukaryotes [167]. TFIIIA consists of nine C2H2 domains and a C-terminal activation domain called TAS (Transcription Activating Signal). The protein binds to a regulatory element called ICR, which is located in the transcribed parts of the genes. A structural analysis revealed that the C2H2 domains **1**–**3** and **7**–**9** bind to two regions (the C and A boxes) of the ICR element; the central C2H2 domains bind specifically to 5S RNA [167]. The lack of homology in the amino acid sequence of TFIIIA proteins from different species suggests a parallel evolution of the DNA sequences of the promoters, 5S rRNAs, and C2H2 domains, which are involved in the specific binding of DNA and RNA. The TFIIIA protein determines open chromatin on the promoter, while the TAS domain is involved in the recruitment and stable binding of the TFIIIB complex to the promoter [168].

The existing data show that many C2H2 proteins are involved in the formation of active promoters, as well as the recruitment of transcription factors and complexes to regulatory elements. It is obvious that many C2H2 proteins duplicate each other's functions, which makes it difficult to prove their role in the global organization of the chromosome architecture and regulation of transcription.

C2H2 PROTEINS IN DROSOPHILA: DIFFERENT STRUCTURES BUT SIMILAR PROPERTIES

Approximately 170 proteins with clusters consisting of at least five C2H2 domains have been found in the Drosophila genome. However, data on the distribution of C2H2 binding sites in the genome and their functional role in the regulation of gene transcription and the chromosomal architecture have been obtained for only a few of these proteins (*Fig. 2*). The best studied C2H2 proteins include the first protein with insulator properties described in higher eukaryotes, Su(Hw), and an homolog of mammalian CTCF, dCTCF [22, 24, 169, 170]. Both insulator proteins have a similar structure: they contain unstructured terminal domains and a central cluster consisting of 11 (dCTCF) or 12 (Su(Hw)) C2H2 domains. The N-terminus of the dCTCF protein contains an unstructured globular domain capable of forming tetrameric com-

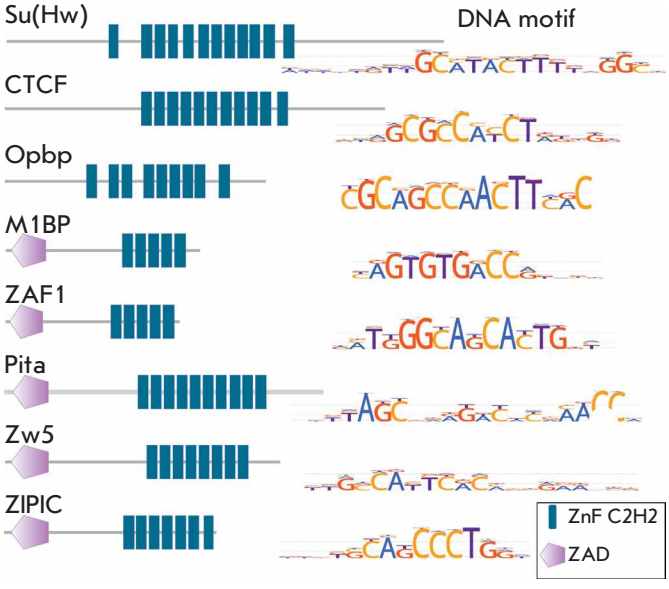


Fig.2. Drosophila C2H2 proteins with architectural functions. The domain organization of the known architectural proteins of Drosophila and their binding motifs are shown

plexes [68, 69], and a potential site of interaction with the cohesin complex, which has homology with the human YxF motif of CTCF that interacts with the SA2-SCC1 complex [71]. An interesting structural feature of another studied C2H2 protein, Opbp [171], is the presence at an N-terminus of an atypical zinc finger capable of homodimerization (*Fig. 2*). Opbp has also a cluster consisting of five C2H2 domains responsible for specific binding to DNA and an additional four C2H2 domains that can interact with RNA and proteins.

The remaining five C2H2 proteins (M1BP, ZAF1, Pita, Zw5, and ZIPIC) belong to a large group of ZAD-containing proteins. The ZAD (zinc-finger-associated domain) was found in 98 Drosophila proteins; approximately 70 of these proteins contain five or more C2H2 domains [172, 173]. The genes encoding ZAD-C2H2 proteins are typically arranged in clusters and, like mammalian KRAB-C2H2 proteins [174, 175], actively evolve as a result of multiple duplications of the original gene copies. The ZAD structure is formed by two pairs of cysteine residues coordinated to the zinc ion [176]. The N-terminal portion of the domain is a globular structure; the C-terminal stem structure is formed by a long α-helix. ZAD domains are capable of homodimerization with the formation of an antiparallel dimmer [176, 177]. Mutations in the genes

encoding the Pita and Zw5 proteins are lethal, which suggests an important role for some representatives of C2H2 proteins in the development of Drosophila [178, 179]. Inactivation of the Su(Hw) protein impairs gonad development, resulting in female sterility [180]. Like in mammals, the Drosophila CTCF protein is involved in the regulation of *hox* gene expression [181, 182]. Although the Drosophila genome was found to contain only ~ 40 Opbp binding sites, inactivation of this protein causes pupal mortality [171].

All the investigated C2H2 proteins bind to long (12–15 bp) DNA motifs via four or five C2H2 domains organized as a cluster (*Fig. 2*) [171, 183–186]. Except for Su(Hw), the binding sites of C2H2 proteins predominantly reside in the promoter regions of active genes and introns [177, 184–190]. The most illustrative example of a protein of this class is M1BP, which binds to the promoters of more than 2,000 genes [185] and, according to experimental data [191], participates in the formation of active promoters. The Opbp protein also binds exclusively to gene promoters, as it is colocalized with M1BP in about half of them [171]. Unlike M1BP and Opbp, which are involved in transcription activation, Su(Hw) binds to the promoters of a large group of neuronal genes and represses their transcription in female gonads during the early stages of Drosophila development [192, 193].

The role played by C2H2 proteins in the formation of long-range contacts and inhibiting enhancer activity was analyzed in transgenic Drosophila lines. *In vivo*, C2H2 proteins efficiently interact with artificial DNA fragments, each containing four to five binding sites [177, 184, 188, 194]. In transgenic lines, the activity of an enhancer surrounded by binding sites for the same C2H2 protein is substantially blocked. However, the enhancer activity is restored by the removal of either of the two binding sites of the C2H2 protein, which proves that the interaction between the C2H2 proteins plays a crucial role in the formation of the chromatin loop, resulting in steric isolation of the enhancer. In the transgenic model system, the C2H2 protein binding sites can bring the yeast GAL4 activator and the reporter gene promoter closer together, thus activating transcription [177, 184, 195]. At the same time, combinations of binding sites for different C2H2 proteins cannot bring the GAL4 activator closer to the promoter [177, 195], which can explain the importance of preferential homodimerization of C2H2 proteins in providing specific distant interactions between genomic elements. For example, it has been shown that the ability of the ZAF1 and ZIPIC proteins to maintain distant interactions is determined by their ZAD domains [177, 184]. Therefore, domains capable of forming homodimers seem to play an important role in the organization of specific long-range contacts between the regulatory elements in chromatin.

The role of C2H2 proteins in the organization of the boundaries of regulatory domains can be most clearly demonstrated by the example of the bithorax complex (BX-C), which includes three homeotic genes, Ubx, abd-A, and Abd-B [196, 197]. The regulatory region of the BX-C is divided into nine independent domains, each activating the transcription of one of the three homeotic genes during the development. Several domain boundaries are characterized in detail and mapped as minimal fragments that can function as effective insulators in transgenic model systems [198–202]. Each characterized boundary contains different combinations of the binding sites of the Pita, dCTCF, and Su(Hw) proteins required to ensure its functional activity [65, 66]. The boundaries can be replaced with repeats consisting of four to five binding sites for each C2H2 protein. Thus, despite the differences in their structural organization, the Su(Hw), Pita, and dCTCF proteins have similar functions and work in cooperation in such processes as the organization of regulatory domain boundaries [66, 203, 204].

Unlike in mammals, the boundaries of most TADs in Drosophila coincide with clusters of housekeeping genes [205, 206]. Thus, the M1BP protein (whose binding sites reside in many promoters of housekeeping genes) is most often found at the TAD boundaries, while the binding sites for other characterized C2H2 proteins usually reside inside the TADs. In embryos and embryonic cell lines, the dCTCF protein, despite its cohesin-binding motif, is rarely found at the boundaries of TADs, while from 40% to 60% of dCTCF sites colocalize with cohesin complexes on chromatin [205–207]. Binding of the bulk of cohesin is observed in the open chromatin zones of actively transcribed promoters [208]; therefore, it cannot be ruled out that C2H2 proteins play a direct or an indirect role (organization of open chromatin regions) in the recruitment of cohesin complexes. Interestingly, most TAD boundaries in a BG3 cell culture derived from Drosophila neural tissues coincide with dCTCF binding sites [207]. Therefore, the TAD boundaries in Drosophila can be changed during cell differentiation.

It is most likely that the TAD boundaries are fixed due to interaction between the protein complexes flanking TADs. In addition, for Drosophila, the existence of a mechanism of TAD formation has been demonstrated, and it is due to the electrostatic internucleosomal interactions that make transcriptionally active sites act as TAD boundaries [116].

CONCLUSIONS

At present, the C2H2 proteins of higher eukaryotes remain the least studied class of transcription factors. The well-studied mammalian CTCF protein provides general insight into the properties, partners, and functions of this class of transcription factors. CTCF is probably the ancestor of the entire class of C2H2 proteins, which in the course of evolution could acquire new domains and bind to new DNA sequences. In this context, it is interesting that CTCF in both Drosophila and mammals is involved in the organization of the boundaries of the transcriptional domains of homeotic genes. Drawing on existing information, it can be concluded that C2H2 proteins in mammals and Drosophila are often involved in the organization of active promoters. By interacting with nucleosome remodeling complexes, C2H2 proteins can form open chromatin and become simultaneously involved in the recruitment of major transcription factors to promoters. Many well-studied regulatory elements (promoters and insulators in particular) carry combinations of binding sites for C2H2 proteins which function cooperatively in their interaction with chromatin. Some C2H2 proteins, including CTCF, have been found to contain N-terminal homodimerization domains that may be involved in the organization of specific long-range contacts. The motif interacting with the cohesin complex has been identified only in the CTCF protein. However, C2H2 proteins can probably interact with other surfaces in cohesin and condensin complexes, which is consistent with the localization of these complexes on active promoters.

It is believed that different mechanisms are responsible for TAD boundary formation and long-range contacts in mammals and Drosophila. However, there still remains an open question as to whether the mammalian cohesin complex can cause intensive chromatin loop extrusion during the formation of TADs and long-range interactions between the regulatory elements. It is also unclear why other higher eukaryotes do not have a similar mechanism, although the cohesin complex is highly conserved. Interestingly, the genome of danio fish contains neither CTCF nor the cohesin complex at most TAD boundaries [209], despite the fact that CTCF in danio and humans shows 86% homology. On the other hand, CTCF is found at the TAD boundaries in Drosophila neural cells [207]. It can be assumed that the mechanisms of formation of TADs are actually much more universal than it appears at present. C2H2 proteins such as Prdm5 and ZNF143 can stabilize CTCF binding to mammalian TAD boundaries and be involved in long-range interactions. Drosophila C2H2 proteins, by binding in various combinations to insulators (for example, as part of BX-C), allow two identical copies of the insulator to maintain super-long-range interactions, which is similar to the formation of the boundaries of a new TAD. In mammals, the TAD boundaries usually contain the most evolutionarily conserved clusters of CTCF sites [119]. It can be assumed that at the early stages of vertebrate evolution, multiplied copies of one or several types of mobile elements containing CTCF binding sites, in combination with the sites of other C2H2 proteins, formed long-range interactions, and some of them have given rise to TADs. Therefore, despite the considerable progress achieved in studying the spatial organization of the genome and, in particular, the architectural role of CTCF, many questions remain unanswered due to the lack of data on the other participants necessary for the formation of the nucleus architecture. ●

This work was supported by the Russian Science Foundation (project No. 19-74-30026).

REFERENCES

1. Spitz F., Furlong E.E. // Nature Reviews Genetics. 2012. V. 13. № 9. P. 613–626.
2. Levine M., Cattoglio C., Tjian R. // Cell. 2014. V. 157. № 1. P. 13–25.
3. Zabidi M.A., Stark A. // Trends Genet. 2016. V. 32. № 12. P. 801–814.
4. Furlong E.E.M., Levine M. // Science. 2018. V. 361. № 6409. P. 1341–1345.
5. Geyer P.K., Clark I. // Cell. Mol. Life Sci. 2002. V. 59. № 12. P. 2112–2127.
6. West A.G., Gaszner M., Felsenfeld G. // Genes Dev. 2002. V. 16. № 3. P. 271–288.
7. Gerasimova T.I., Corces V.G. // Annu. Rev. Genet. 2001. V. 35. P. 193–208.
8. Rao S.S., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., et al. // Cell. 2014. V. 159. № 7. P. 1665–1680.
9. Boettiger A., Murphy S. // Trends Genet. 2020. V. 36. № 4. P. 273–287.
10. Boettiger A.N., Bintu B., Moffitt J.R., Wang S., Beliveau B.J., Fudenberg G., Imakaev M., Mirny L.A., Wu C.T., Zhuang X. // Nature. 2016. V. 529. № 7586. P. 418–422.
11. Dekker J., Misteli T. // Cold Spring Harbor Perspectives Biol. 2015. V. 7. № 10. P. a019356.
12. Sikorska N., Sexton T. // J. Mol. Biol. 2020. V. 432. № 3. P. 653–664.
13. Hansen A.S., Cattoglio C., Darzacq X., Tjian R. // Nucleus. 2018. V. 9. № 1. P. 20–32.
14. Luppino J.M., Park D.S., Nguyen S.C., Lan Y., Xu Z., Yunker R., Joyce E.F. // Nat. Genet. 2020. V. 52. № 8. P. 840–848.
15. Zheng H., Xie W. // Nat. Rev. Mol. Cell Biol. 2019. V. 20. № 9. P. 535–550.
16. Sexton T., Yaffe E., Kenigsberg E., Bantignies F., Leblanc B., Hoichman M., Parrinello H., Tanay A., Cavalli G. // Cell. 2012. V. 148. № 3. P. 458–472.
17. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. // Nature. 2012. V. 485. № 7398. P. 376–380.
18. Nora E.P., Lajoie B.R., Schulz E.G., Giorgetti L., Okamoto I., Servant N., Piolot T., van Berkum N.L., Meisig J., Sedat J., et al. // Nature. 2012. V. 485. № 7398. P. 381–385.
19. Szabo Q., Bantignies F., Cavalli G. // Science Advances. 2019. V. 5. № 4. P. eaaw1668.
20. Chang L.H., Ghosh S., Noordermeer D. // J. Mol. Biol. 2020. V. 432. № 3. P. 643–652.
21. Arzate-Mejia R.G., Recillas-Targa F., Corces V.G. // Development. 2018. V. 145. № 6. P. dev137729.
22. Ali T., Renkawitz R., Bartkuhn M. // Curr. Opin. Genetics Dev. 2016. V. 37. P. 17–26.
23. Braccioli L., de Wit E. // Essays Biochem. 2019. V. 63. № 1. P. 157–165.
24. Chen D., Lei E.P. // Curr. Opin. Cell Biol. 2019. V. 58. P. 61–68.
25. Merkenschlager M., Nora E.P. // Annu. Rev. Genomics Hum. Genet. 2016. V. 17. P. 17–43.
26. Schmidt D., Schwalie P.C., Wilson M.D., Ballester B., Goncalves A., Kutter C., Brown G.D., Marshall A., Flicek P., Odom D.T. // Cell. 2012. V. 148. № 1–2. P. 335–348.
27. Chen H., Tian Y., Shu W., Bo X., Wang S. // PLoS One. 2012. V. 7. № 7. P. e41374.
28. Klug A. // Quarterly Rev. Biophys. 2010. V. 43. № 1. P. 1–21.
29. Persikov A.V., Wetzel J.L., Rowland E.F., Oakes B.L., Xu D.J., Singh M., Noyes M.B. // Nucleic Acids Res. 2015. V. 43. № 3. P. 1965–1984.
30. Garton M., Najafabadi H.S., Schmitges F.W., Radovani E., Hughes T.R., Kim P.M. // Nucleic Acids Res. 2015. V. 43. № 19. P. 9147–9157.
31. Persikov A.V., Singh M. // Nucleic Acids Res. 2014. V. 42. № 1. P. 97–108.
32. Durai S., Mani M., Kandavelou K., Wu J., Porteus M.H., Chandrasegaran S. // Nucleic Acids Res. 2005. V. 33. № 18. P. 5978–5990.
33. Kim Y.G., Cha J., Chandrasegaran S. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. № 3. P. 1156–1160.
34. Brayer K.J., Segal D.J. // Cell Biochem. Biophys. 2008. V. 50. № 3. P. 111–131.
35. Ryan R.F., Darby M.K. // Nucleic Acids Res. 1998. V. 26. № 3. P. 703–709.
36. Crozatier M., Kongsuwan K., Ferrer P., Merriam J.R., Lengyel J.A., Vincent A. // Genetics. 1992. V. 131. № 4. P. 905–916.
37. Wolfe S.A., Nekludova L., Pabo C.O. // Annu. Rev. Biophys. Biomol. Structure. 2000. V. 29. P. 183–212.
38. Hashimoto H., Wang D., Horton J.R., Zhang X., Corces V.G., Cheng X. // Mol. Cell. 2017. V. 66. № 5. P. 711–720 e713.
39. Nakahashi H., Kwon K.R., Resch W., Vian L., Dose M., Stavreva D., Hakim O., Pruett N., Nelson S., Yamane A., et al. // Cell Reports. 2013. V. 3. № 5. P. 1678–1689.
40. Xiao T., Wongtrakoongate P., Trainor C., Felsenfeld G. // Cell Reports. 2015. V. 12. № 10. P. 1704–1714.
41. Xu D., Ma R., Zhang J., Liu Z., Wu B., Peng J., Zhai Y., Gong Q., Shi Y., Wu J., et al. // J. Phys. Chem. Lett. 2018. V. 9. № 14. P. 4020–4028.
42. Yin M., Wang J., Wang M., Li X., Zhang M., Wu Q., Wang Y. // Cell Research. 2017. V. 27. № 11. P. 1365–1377.
43. Liu Y., Zhang X., Blumenthal R.M., Cheng X. // Trends Biochem. Sci. 2013. V. 38. № 4. P. 177–183.
44. Hudson N.O., Buck-Koehntop B.A. // Molecules. 2018. V. 23. № 10. P. 2555.
45. Ren G., Zhao K. // Cell. Biosci. 2019. V. 9. P. 83.
46. Noordermeer D., Feil R. // Curr. Opin. Genet. Dev. 2020. V. 61. P. 17–24.
47. Heard E., Disteche C.M. // Genes Dev. 2006. V. 20. № 14. P. 1848–1867.
48. Iuchi S. // Cell. Mol. Life Sci. 2001. V. 58. № 4. P. 625–635.
49. Hall T.M. // Curr. Opin. Struct. Biol. 2005. V. 15. № 3. P. 367–373.
50. Saldana-Meyer R., Gonzalez-Buendia E., Guerrero G., Narendra V., Bonasio R., Recillas-Targa F., Reinberg D. // Genes Dev. 2014. V. 28. № 7. P. 723–734.
51. Kung J.T., Kesner B., An J.Y., Ahn J.Y., Cifuentes-Rojas C., Colognori D., Jeon Y., Szanto A., del Rosario B.C., Pinter S.F., et al. // Mol. Cell. 2015. V. 57. № 2. P. 361–375.
52. Hansen A.S., Hsieh T.S., Cattoglio C., Pustova I., Saldana-Meyer R., Reinberg D., Darzacq X., Tjian R. // Mol. Cell. 2019. V. 76. № 3. P. 395–411 e313.
53. Shukla S., Kavak E., Gregory M., Imashimizu M., Shutinoski B., Kashlev M., Oberdoerffer P., Sandberg R., Oberdoerffer S. // Nature. 2011. V. 479. № 7371. P. 74–79.
54. Marina R.J., Sturgill D., Bailly M.A., Thenoz M., Varma G., Prigge M.F., Nanan K.K., Shukla S., Haque N., Oberdoerffer S. // EMBO. 2016. V. 35. № 3. P. 335–355.

55. Nanavaty V., Abrash E.W., Hong C., Park S., Fink E.E., Li Z., Sweet T.J., Bhasin J.M., Singuri S., Lee B.H., et al. // Mol. Cell. 2020. V. 78. № 4. P. 752–764 e756.
56. Chernukhin I., Shamsuddin S., Kang S.Y., Bergstrom R., Kwon Y.W., Yu W., Whitehead J., Mukhopadhyay R., Docquier F., Farrar D., et al. // Mol. Cell Biol. 2007. V. 27. № 5. P. 1631–1648.
57. Heger P., Marin B., Schierenberg E. // BMC Mol. Biol. 2009. V. 10. P. 84.
58. Heger P., Marin B., Bartkuhn M., Schierenberg E., Wiehe T. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 43. P. 17507–17512.
59. Moon H., Filippova G., Loukinov D., Pugacheva E., Chen Q., Smith S.T., Munhall A., Grewe B., Bartkuhn M., Arnold R., et al. // EMBO Reports. 2005. V. 6. № 2. P. 165–170.
60. Holohan E.E., Kwong C., Adryan B., Bartkuhn M., Herold M., Renkawitz R., Russell S., White R. // PLoS Genet. 2007. V. 3. № 7. P. e112.
61. Kadota M., Hara Y., Tanaka K., Takagi W., Tanegashima C., Nishimura O., Kuraku S. // Sci. Rep. 2017. V. 7. № 1. P. 4957.
62. Narendra V., Rocha P.P., An D., Raviram R., Skok J.A., Mazzoni E.O., Reinberg D. // Science. 2015. V. 347. № 6225. P. 1017–1021.
63. Savitsky M., Kim M., Kravchuk O., Schwartz Y.B. // Genetics. 2016. V. 202. № 2. P. 601–617.
64. Luo H., Wang F., Zha J., Li H., Yan B., Du Q., Yang F., Sobh A., Vulpe C., Drusbosky L., et al. // Blood. 2018. V. 132. № 8. P. 837–848.
65. Kyrchanova O., Zolotarev N., Mogila V., Maksimenko O., Schedl P., Georgiev P. // Development. 2017. V. 144. № 14. P. 2663–2672.
66. Kyrchanova O., Maksimenko O., Ibragimov A., Sokolov V., Postika N., Lukyanova M., Schedl P., Georgiev P. // Sci. Adv. 2020. V. 6. № 13. P. eaaz3152.
67. Schwalie P.C., Ward M.C., Cain C.E., Faure A.J., Gilad Y., Odom D.T., Flicek P. // Genome Biol. 2013. V. 14. № 12. P. R148.
68. Bonchuk A., Kamalyan S., Mariasina S., Boyko K., Popov V., Maksimenko O., Georgiev P. // Sci. Rep. 2020. V. 10. № 1. P. 2677.
69. Bonchuk A., Maksimenko O., Kyrchanova O., Ivlieva T., Mogila V., Deshpande G., Wolle D., Schedl P., Georgiev P. // BMC Biol. 2015. V. 13. P. 63.
70. Nishana M., Ha C., Rodriguez-Hernaez J., Ranjbaran A., Chio E., Nora E.P., Badri S.B., Kloetgen A., Bruneau B.G., Tsirigos A., et al. // Genome Biol. 2020. V. 21. № 1. P. 108.
71. Li Y., Haarhuis J.H.I., Sedeno Cacciatore A., Oldenkamp R., van Ruiten M.S., Willems L., Teunissen H., Muir K.W., de Wit E., Rowland B.D., et al. // Nature. 2020. V. 578. № 7795. P. 472–476.
72. Xiao T., Wallace J., Felsenfeld G. // Mol. Cell Biol. 2011. V. 31. № 11. P. 2174–2183.
73. Zlatanova J., Caiafa P. // J. Cell Sci. 2009. V. 122. № Pt 9. P. 1275–1284.
74. Marino M.M., Rega C., Russo R., Valletta M., Gentile M.T., Esposito S., Baglivo I., De Feis I., Angelini C., Xiao T., et al. // J. Biol. Chem. 2019. V. 294. № 3. P. 861–873.
75. Pena-Hernandez R., Marques M., Hilmi K., Zhao T., Saad A., Alaoui-Jamali M.A., del Rincon S.V., Ashworth T., Roy A.L., Emerson B.M., et al. // Proc. Natl. Acad. Sci. USA. 2015. V. 112. № 7. P. E677–686.
76. Nora E.P., Goloborodko A., Valton A.L., Gibcus J.H., Uebersohn A., Abdennur N., Dekker J., Mirny L.A., Bruneau B.G. // Cell. 2017. V. 169. № 5. P. 930–944 e922.
77. Yao H., Brick K., Evrard Y., Xiao T., Camerini-Otero R.D., Felsenfeld G. // Genes Dev. 2010. V. 24. № 22. P. 2543–2555.
78. Uuskula-Reimand L., Hou H., Samavarchi-Tehrani P., Rudan M.V., Liang M., Medina-Rivera A., Mohammed H., Schmidt D., Schwalie P., Young E.J., et al. // Genome Biol. 2016. V. 17. № 1. P. 182.
79. Gittens W.H., Johnson D.J., Allison R.M., Cooper T.J., Thomas H., Neale M.J. // Nat. Commun. 2019. V. 10. № 1. P. 4846.
80. Jantz D., Berg J.M. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 20. P. 7589–7593.
81. Dovat S., Ronni T., Russell D., Ferrini R., Cobb B.S., Smale S.T. // Genes Dev. 2002. V. 16. № 23. P. 2985–2990.
82. Rizkallah R., Alexander K.E., Hurt M.M. // Cell Cycle. 2011. V. 10. № 19. P. 3327–3336.
83. Luo H., Yu Q., Liu Y., Tang M., Liang M., Zhang D., Xiao T.S., Wu L., Tan M., Ruan Y., et al. // Science Adv. 2020. V. 6. № 8. P. eaaw4651.
84. Caiafa P., Zlatanova J. // J. Cell. Physiol. 2009. V. 219. № 2. P. 265–270.
85. Farrar D., Rai S., Chernukhin I., Jagodic M., Ito Y., Yammine S., Ohlsson R., Murrell A., Klenova E. // Mol. Cell Biol. 2010. V. 30. № 5. P. 1199–1216.
86. Pavlaki I., Docquier F., Chernukhin I., Kita G., Gretton S., Clarkson C.T., Teif V.B., Klenova E. // Biochim. Biophys. Acta Gene Regul. Mech. 2018. V. 1861. № 8. P. 718–730.
87. Torrano V., Navascues J., Docquier F., Zhang R., Burke L.J., Chernukhin I., Farrar D., Leon J., Berciano M.T., Renkawitz R., et al. // J. Cell Sci. 2006. V. 119. Pt 9. P. 1746–1759.
88. Wang A.J., Han Y., Jia N., Chen P., Minden M.D. // Leukemia. 2020. V. 34. № 5. P. 1278–1290.
89. MacPherson M.J., Beatty L.G., Zhou W., Du M., Sadowski P.D. // Mol. Cell Biol. 2009. V. 29. № 3. P. 714–725.
90. Golovnin A., Volkov I., Georgiev P. // J. Cell Sci. 2012. V. 125. № Pt 8. P. 2064–2074.
91. Rosonina E., Akhter A., Dou Y., Babu J., Sri Theivakadadcham V.S. // Transcription. 2017. V. 8. № 4. P. 220–231.
92. Wallace J.A., Felsenfeld G. // Curr. Opin. Genet. Dev. 2007. V. 17. № 5. P. 400–407.
93. Barkess G., West A.G. // Epigenomics. 2012. V. 4. № 1. P. 67–80.
94. Ghirlando R., Felsenfeld G. // Genes Dev. 2016. V. 30. № 8. P. 881–891.
95. Farrell C.M., West A.G., Felsenfeld G. // Mol. Cell Biol. 2002. V. 22. № 11. P. 3820–3831.
96. West A.G., Huang S., Gaszner M., Litt M.D., Felsenfeld G. // Mol. Cell. 2004. V. 16. № 3. P. 453–463.
97. Dickson J., Gowher H., Strogantsev R., Gaszner M., Hair A., Felsenfeld G., West A.G. // PLoS Genet. 2010. V. 6. № 1. P. e1000804.
98. Gowher H., Brick K., Camerini-Otero R.D., Felsenfeld G. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 7. P. 2370–2375.
99. Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N., Mirny L.A. // Cell Reports. 2016. V. 15. № 9. P. 2038–2049.
100. Nishiyama T. // Curr. Opin. Cell Biol. 2019. V. 58. P. 8–14.
101. Morales C., Losada A. // Curr. Opin. Cell Biol. 2018. V. 52. P. 51–57.
102. Parelho V., Hadjur S., Spivakov M., Leleu M., Sauer S., Gregson H.C., Jarmuz A., Canzonetta C., Webster Z., Nesterova T., et al. // Cell. 2008. V. 132. № 3. P. 422–433.
103. Wendt K.S., Yoshida K., Itoh T., Bando M., Koch B., Schirghuber E., Tsutsumi S., Nagae G., Ishihara K., Mishiro T., et al. // Nature. 2008. V. 451. № 7180. P. 796–801.
104. Pugacheva E.M., Kubo N., Loukinov D., Tajmul M., Kang

S., Kovalchuk A.L., Strunnikov A.V., Zentner G.E., Ren B., Lobanenkov V.V. // Proc. Natl. Acad. Sci. USA. 2020. V. 117. № 4. P. 2020–2031.
105. Rao S.S.P., Huang S.C., Glenn St Hilaire B., Engreitz J.M., Perez E.M., Kieffer-Kwon K.R., Sanborn A.L., Johnstone S.E., Bascom G.D., Bochkov I.D., et al. // Cell. 2017. V. 171. № 2. P. 305–320 e324.
106. Wutz G., Varnai C., Nagasaka K., Cisneros D.A., Stocsits R.R., Tang W., Schoenfelder S., Jessberger G., Muhar M., Hossain M.J., et al. // EMBO. 2017. V. 36. № 24. P. 3573–3599.
107. Ladurner R., Bhaskara V., Huis in 't Veld P.J., Davidson I.F., Kreidl E., Petzold G., Peters J.M. // Curr. Biol. 2014. V. 24. № 19. P. 2228–2237.
108. Elbatsh A.M.O., Haarhuis J.H.I., Petela N., Chapard C., Fish A., Celie P.H., Stadnik M., Ristic D., Wyman C., Medema R.H., et al. // Mol. Cell. 2016. V. 61. № 4. P. 575–588.
109. Vian L., Pekowska A., Rao S.S.P., Kieffer-Kwon K.R., Jung S., Baranello L., Huang S.C., El Khattabi L., Dose M., Pruett N., et al. // Cell. 2018. V. 175. № 1. P. 292–294.
110. Vietri Rudan M., Barrington C., Henderson S., Ernst C., Odom D.T., Tanay A., Hadjur S. // Cell Rep. 2015. V. 10. № 8. P. 1297–1309.
111. de Wit E., Vos E.S., Holwerda S.J., Valdes-Quezada C., Verstegen M.J., Teunissen H., Splinter E., Wijchers P.J., Krijger P.H., de Laat W. // Mol. Cell. 2015. V. 60. № 4. P. 676–684.
112. Guo Y., Xu Q., Canzio D., Shou J., Li J., Gorkin D.U., Jung I., Wu H., Zhai Y., Tang Y., et al. // Cell. 2015. V. 162. № 4. P. 900–910.
113. Davidson I.F., Bauer B., Goetz D., Tang W., Wutz G., Peters J.M. // Science. 2019. V. 366. № 6471. P. 1338–1345.
114. Kim Y., Shi Z., Zhang H., Finkelstein I.J., Yu H. // Science. 2019. V. 366. № 6471. P. 1345–1349.
115. Stigler J., Camdere G.O., Koshland D.E., Greene E.C. // Cell Rep. 2016. V. 15. № 5. P. 988–998.
116. Ulianov S.V., Khrameeva E.E., Gavrilov A.A., Flyamer I.M., Kos P., Mikhaleva E.A., Penin A.A., Logacheva M.D., Imakaev M.V., Chertovich A., et al. // Genome Res. 2016. V. 26. № 1. P. 70–84.
117. Luzhin A.V., Flyamer I.M., Khrameeva E.E., Ulianov S.V., Razin S.V., Gavrilov A.A. // J. Cell. Biochem. 2019. V. 120. № 3. P. 4494–4503.
118. Holzmann J., Politi A.Z., Nagasaka K., Hantsche-Grininger M., Walther N., Koch B., Fuchs J., Durnberger G., Tang W., Ladurner R., et al. // eLife. 2019. V. 8. P. e46269.
119. Kentepozidou E., Aitken S.J., Feig C., Stefflova K., Ibarra-Soria X., Odom D.T., Roller M., Flicek P. // Genome Biol. 2020. V. 21. № 1. P. 5.
120. Haarhuis J.H.I., van der Weide R.H., Blomen V.A., Yanez-Cuna J.O., Amendola M., van Ruiten M.S., Krijger P.H.L., Teunissen H., Medema R.H., van Steensel B., et al. // Cell. 2017. V. 169. № 4. P. 693–707 e614.
121. Gassler J., Brandao H.B., Imakaev M., Flyamer I.M., Ladstatter S., Bickmore W.A., Peters J.M., Mirny L.A., Tachibana K. // EMBO J. 2017. V. 36. № 24. P. 3600–3618.
122. Zhang H., Emerson D.J., Gilgenast T.G., Titus K.R., Lan Y., Huang P., Zhang D., Wang H., Keller C.A., Giardine B., et al. // Nature. 2019. V. 576. № 7785. P. 158–162.
123. Owens N., Papadopoulou T., Festuccia N., Tachtsidi A., Gonzalez I., Dubois A., Vandormael-Pournin S., Nora E.P., Bruneau B.G., Cohen-Tannoudji M., et al. // eLife. 2019. V. 8. P. e47898.
124. Lambert S.A., Jolma A., Campitelli L.F., Das P.K., Yin Y., Albu M., Chen X., Taipale J., Hughes T.R., Weirauch M.T. // Cell. 2018. V. 172. № 4. P. 650–665.
125. Lambert S.A., Yang A.W.H., Sasse A., Cowley G., Albu M., Caddick M.X., Morris Q.D., Weirauch M.T., Hughes T.R. // Nat. Genet. 2019. V. 51. № 6. P. 981–989.
126. Imbeault M., Helleboid P.Y., Trono D. // Nature. 2017. V. 543. № 7646. P. 550–554.
127. Schmitges F.W., Radovani E., Najafabadi H.S., Barazandeh M., Campitelli L.F., Yin Y., Jolma A., Zhong G., Guo H., Kanagalingam T., et al. // Genome Res. 2016. V. 26. № 12. P. 1742–1752.
128. Barazandeh M., Lambert S.A., Albu M., Hughes T.R. // G3 (Bethesda). 2018. V. 8. № 1. P. 219–229.
129. Platt R.N., 2nd, Vandewege M.W., Ray D.A. // Chromosome Research. 2018. V. 26. № 1–2. P. 25–43.
130. Bruno M., Mahgoub M., Macfarlan T.S. // Annu. Rev. Genet. 2019. V. 53. P. 393–416.
131. Emerson R.O., Thomas J.H. // J. Virol. 2011. V. 85. № 22. P. 12043–12052.
132. Okumura K., Sakaguchi G., Naito K., Tamura T., Igarashi H. // Nucleic Acids Res. 1997. V. 25. № 24. P. 5025–5032.
133. Rohrmoser M., Kluge M., Yahia Y., Gruber-Eber A., Maqbool M.A., Forne I., Krebs S., Blum H., Greifenberg A.K., Geyer M., et al. // Nucleic Acids Res. 2019. V. 47. № 2. P. 700–715.
134. Diehl A.G., Ouyang N., Boyle A.P. // Nat. Commun. 2020. V. 11. № 1. P. 1796.
135. Herz H.M., Garruss A., Shilatifard A. // Trends in Biochemical Sciences. 2013. V. 38. № 12. P. 621–639.
136. Maeda T. // Int. J. Hematol. 2016. V. 104. № 3. P. 310–323.
137. Al Chiblak M., Steinbeck F., Thiesen H.J., Lorenz P. // BMC Molecular and Cell Biology. 2019. V. 20. № 1. P. 60.
138. Schumacher C., Wang H., Honer C., Ding W., Koehn J., Lawrence Q., Coulis C.M., Wang L.L., Ballinger D., Bowen B.R., et al. // J. Biol. Chem. 2000. V. 275. № 22. P. 17173–17179.
139. Yang P., Wang Y., Macfarlan T.S. // Trends Genet. 2017. V. 33. № 11. P. 871–881.
140. Francis M., Cheng H., Ma P., Grider A. // Biol. Trace Elem. Res. 2019. V. 192. № 2. P. 83–90.
141. Ogo O.A., Tyson J., Cockell S.J., Howard A., Valentine R.A., Ford D. // Mol. Cell Biol. 2015. V. 35. № 6. P. 977–987.
142. Kino T., Pavlatou M.G., Moraitis A.G., Nemery R.L., Raygada M., Stratakis C.A. // J. Clin. Endocrinol. Metab. 2012. V. 97. № 8. P. E1557–1566.
143. Fadda A., Syed N., Mackeh R., Papadopoulou A., Suzuki S., Jithesh P.V., Kino T. // Scientific Reports. 2017. V. 7. P. 41598.
144. Wagner S., Hess M.A., Ormonde-Hanson P., Malandro J., Hu H., Chen M., Kehrer R., Frodsham M., Schumacher C., Beluch M., et al. // J. Biol. Chem. 2000. V. 275. № 21. P. 15685–15690.
145. Frietze S., Lan X., Jin V.X., Farnham P.J. // J. Biol. Chem. 2010. V. 285. № 2. P. 1393–1403.
146. Brix D.M., Bundgaard Clemmensen K.K., Kallunki T. // Cells. 2020. V. 9. № 1. P. 223.
147. Galli G.G., Multhaupt H.A., Carrara M., de Lichtenberg K.H., Christensen I.B., Linnemann D., Santoni-Rugiu E., Calogero R.A., Lund A.H. // Oncogene. 2014. V. 33. № 25. P. 3342–3350.
148. Noll L., Peterson F.C., Hayes P.L., Volkman B.F., Sander T. // Leukemia Research. 2008. V. 32. № 10. P. 1582–1592.
149. Peterson F.C., Hayes P.L., Waltner J.K., Heisner A.K., Jensen D.R., Sander T.L., Volkman B.F. // J. Mol. Biol. 2006. V. 363. № 1. P. 137–147.
150. Helleboid P.Y., Heusel M., Duc J., Piot C., Thorball C.W., Coluccio A., Pontis J., Imbeault M., Turelli P., Aebersold R.,

et al. // EMBO. 2019. V. 38. № 18. P. e101220.
151. Duan Z., Person R.E., Lee H.H., Huang S., Donadieu J., Badolato R., Grimes H.L., Papayannopoulou T., Horwitz M.S. // Mol. Cell Biol. 2007. V. 27. № 19. P. 6889–6902.
152. Galli G.G., Carrara M., Francavilla C., de Lichtenberg K.H., Olsen J.V., Calogero R.A., Lund A.H. // Mol. Cell Biol. 2013. V. 33. № 22. P. 4504–4516.
153. Myslinski E., Krol A., Carbon P. // J. Biol. Chem. 1998. V. 273. № 34. P. 21998–22006.
154. Schuster C., Myslinski E., Krol A., Carbon P. // EMBO. 1995. V. 14. № 15. P. 3777–3787.
155. Bailey S.D., Zhang X., Desai K., Aid M., Corradin O., Cowper-Sal Lari R., Akhtar-Zaidi B., Scacheri P.C., Haibe-Kains B., Lupien M. // Nat. Commun. 2015. V. 2. P. 6186.
156. Heidari N., Phanstiel D.H., He C., Grubert F., Jahanbani F., Kasowski M., Zhang M.Q., Snyder M.P. // Genome Res. 2014. V. 24. № 12. P. 1905–1917.
157. Myslinski E., Gerard M.A., Krol A., Carbon P. // J. Biol. Chem. 2006. V. 281. № 52. P. 39953–39962.
158. Ngondo-Mbongo R.P., Myslinski E., Aster J.C., Carbon P. // Nucleic Acids Res. 2013. V. 41. № 7. P. 4000–4014.
159. Schaub M., Krol A., Carbon P. // Nucleic Acids Res. 2000. V. 28. № 10. P. 2114–2121.
160. Schaub M., Myslinski E., Krol A., Carbon P. // J. Biol. Chem. 1999. V. 274. № 35. P. 25042–25050.
161. Sathyan K.M., McKenna B.D., Anderson W.D., Duarte F.M., Core L., Guertin M.J. // Genes Dev. 2019. V. 33. № 19–20. P. 1441–1455.
162. Mourad R., Cuvier O. // Nucleic Acids Res. 2018. V. 46. № 5. P. e27.
163. Wen Z., Huang Z.T., Zhang R., Peng C. // Cell Biol. Toxicol. 2018. V. 34. № 6. P. 471–478.
164. Yang Y., Zhang R., Singh S., Ma J. // Bioinformatics. 2017. V. 33. № 14. P. i252–i260.
165. Raab J.R., Chiu J., Zhu J., Katzman S., Kurukuti S., Wade P.A., Haussler D., Kamakaka R.T. // EMBO. 2012. V. 31. № 2. P. 330–350.
166. van Bortle K., Phanstiel D.H., Snyder M.P. // Genome Biol. 2017. V. 18. № 1. P. 180.
167. Layat E., Probst A.V., Tourmente S. // Biochim. Biophys. Acta. 2013. V. 1829. № 3–4. P. 274–282.
168. Smith D.R., Jackson I.J., Brown D.D. // Cell. 1984. V. 37. № 2. P. 645–652.
169. Matthews N.E., White R. // BioEssays. 2019. P. e1900048.
170. Schwartz Y.B., Cavalli G. // Genetics. 2017. V. 205. № 1. P. 5–24.
171. Zolotarev N., Maksimenko O., Kyrchanova O., Sokolinskaya E., Osadchiy I., Girardot C., Bonchuk A., Ciglar L., Furlong E.E.M., Georgiev P. // Nucleic Acids Res. 2017. V. 45. № 21. P. 12285–12300.
172. Chung H.R., Schafer U., Jackle H., Bohm S. // EMBO Reports. 2002. V. 3. № 12. P. 1158–1162.
173. Chung H.R., Lohr U., Jackle H. // Mol. Biol. Evol. 2007. V. 24. № 9. P. 1934–1943.
174. Mackeh R., Marr A.K., Fadda A., Kino T. // Nuclear Receptor Signaling. 2018. V. 15. P. 1550762918801071.
175. Ecco G., Imbeault M., Trono D. // Development. 2017. V. 144. № 15. P. 2719–2729.
176. Jauch R., Bourenkov G.P., Chung H.R., Urlaub H., Reidt U., Jackle H., Wahl M.C. // Structure. 2003. V. 11. № 11. P. 1393–1402.
177. Zolotarev N., Fedotova A., Kyrchanova O., Bonchuk A., Penin A.A., Lando A.S., Eliseeva I.A., Kulakovskiy I.V., Maksimenko O., Georgiev P. // Nucleic Acids Res. 2016. V. 44. № 15. P. 7228–7241.
178. Gaszner M., Vazquez J., Schedl P. // Genes Dev. 1999. V. 13. № 16. P. 2098–2107.
179. Page A.R., Kovacs A., Deak P., Torok T., Kiss I., Dario P., Bastos C., Batista P., Gomes R., Ohkura H., et al. // EMBO. 2005. V. 24. № 24. P. 4304–4315.
180. Baxley R.M., Soshnev A.A., Koryakov D.E., Zhimulev I.F., Geyer P.K. // Dev. Biol. 2011. V. 356. № 2. P. 398–410.
181. Mohan M., Bartkuhn M., Herold M., Philippen A., Heinl N., Bardenhagen I., Leers J., White R.A., Renkawitz-Pohl R., Saumweber H., et al. // EMBO. 2007. V. 26. № 19. P. 4203–4214.
182. Gambetta M.C., Furlong E.E.M. // Genetics. 2018. V. 210. № 1. P. 129–136.
183. Baxley R.M., Bullard J.D., Klein M.W., Fell A.G., Morales-Rosado J.A., Duan T., Geyer P.K. // Nucleic Acids Res. 2017. V. 45. № 8. P. 4463–4478.
184. Maksimenko O., Kyrchanova O., Klimenko N., Zolotarev N., Elizarova A., Bonchuk A., Georgiev P. // Biochim. Biophys. Acta Gene Regul. Mech. 2020. V. 1863. № 1. P. 194446.
185. Li J., Gilmour D.S. // EMBO. 2013. V. 32. № 13. P. 1829–1841.
186. Schwartz Y.B., Linder-Basso D., Kharchenko P.V., Tolstorukov M.Y., Kim M., Li H.B., Gorchakov A.A., Minoda A., Shanower G., Alekseyenko A.A., et al. // Genome Res. 2012. V. 22. № 11. P. 2188–2198.
187. Soshnev A.A., He B., Baxley R.M., Jiang N., Hart C.M., Tan K., Geyer P.K. // Nucleic Acids Res. 2012. V. 40. № 12. P. 5415–5431.
188. Maksimenko O., Bartkuhn M., Stakhov V., Herold M., Zolotarev N., Jox T., Buxa M.K., Kirsch R., Bonchuk A., Fedotova A., et al. // Genome Res. 2015. V. 25. № 1. P. 89–99.
189. Negre N., Brown C.D., Shah P.K., Kheradpour P., Morrison C.A., Henikoff J.G., Feng X., Ahmad K., Russell S., White R.A., et al. // PLoS Genet. 2010. V. 6. № 1. P. e1000814.
190. Negre N., Brown C.D., Ma L., Bristow C.A., Miller S.W., Wagner U., Kheradpour P., Eaton M.L., Loriaux P., Sealfon R., et al. // Nature. 2011. V. 471. № 7339. P. 527–531.
191. Baumann D.G., Gilmour D.S. // Nucleic Acids Res. 2017. V. 45. № 18. P. 10481–10491.
192. Soshnev A.A., Baxley R.M., Manak J.R., Tan K., Geyer P.K. // Development. 2013. V. 140. № 17. P. 3613–3623.
193. Melnikova L., Elizar'ev P., Erokhin M., Molodina V., Chetverina D., Kostyuchenko M., Georgiev P., Golovnin A. // Sci. Rep. 2019. V. 9. № 1. P. 5314.
194. Kyrchanova O., Maksimenko O., Stakhov V., Ivlieva T., Parshikov A., Studitsky V.M., Georgiev P. // PLoS Genet. 2013. V. 9. № 7. P. e1003606.
195. Kyrchanova O., Chetverina D., Maksimenko O., Kullyev A., Georgiev P. // Nucleic Acids Res. 2008. V. 36. № 22. P. 7019–7028.
196. Maeda R.K., Karch F. // Chromosoma. 2015. V. 124. № 3. P. 293–307.
197. Kyrchanova O., Mogila V., Wolle D., Magbanua J.P., White R., Georgiev P., Schedl P. // Mech. Dev. 2015. V. 138. Pt 2. P. 122–132.
198. Gruzdeva N., Kyrchanova O., Parshikov A., Kullyev A., Georgiev P. // Mol. Cell Biol. 2005. V. 25. № 9. P. 3682–3689.
199. Barges S., Mihaly J., Galloni M., Hagstrom K., Muller M., Shanower G., Schedl P., Gyurkovics H., Karch F. // Development. 2000. V. 127. № 4. P. 779–790.
200. Iampietro C., Gummalla M., Mutero A., Karch F., Maeda R.K. // PLoS Genet. 2010. V. 6. № 12. P. e1001260.
201. Bender W., Lucas M. // Genetics. 2013. V. 193. № 4. P. 1135–1147.
202. Bowman S.K., Deaton A.M., Domingues H., Wang P.I.,

Sadreyev R.I., Kingston R.E., Bender W. // eLife. 2014. V. 3. P. e02833.
203. Kyrchanova O., Mogila V., Wolle D., Deshpande G., Parshikov A., Cleard F., Karch F., Schedl P., Georgiev P. // PLoS Genet. 2016. V. 12. № 7. P. e1006188.
204. Kyrchanova O., Sabirov M., Mogila V., Kurbidaeva A., Postika N., Maksimenko O., Schedl P., Georgiev P. // Proc. Natl. Acad. Sci. USA. 2019. V. 116. № 27. P. 13462–13467.
205. Wang Q., Sun Q., Czajkowsky D.M., Shao Z. // Nat. Commun. 2018. V. 9. № 1. P. 188.
206. Ramirez F., Bhardwaj V., Arrigoni L., Lam K.C., Gruning B.A., Villaveces J., Habermann B., Akhtar A., Manke T. // Nat. Commun. 2018. V. 9. № 1. P. 189.
207. Chathoth K.T., Zabet N.R. // Genome Res. 2019. V. 29. № 4. P. 613–625.
208. Dorsett D. // Trends Genet. 2019. V. 35. № 7. P. 542–551.
209. Perez-Rico Y.A., Barillot E., Shkumatava A. // iScience. 2020. V. 23. № 5. P. 101046.

Muscle-Specific Promoters for Gene Therapy

V. V. Skopenkova[1,2,3*], T. V. Egorova[1,2], M. V. Bardina[1,2,3]
[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia
[2]Marlin Biotech LLC, Moscow, 121205 Russia
[3]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia
*E-mail: v.silina.marlin@gmail.com
Received June 06, 2020; in final form, July 30, 2020
DOI: 10.32607/actanaturae.11063

ABSTRACT **Many genetic diseases that are responsible for muscular disorders have been described to date. Gene replacement therapy is a state-of-the-art strategy used to treat such diseases. In this approach, the functional copy of a gene is delivered to the affected tissues using viral vectors. There is an urgent need for the design of short, regulatory sequences that would drive a high and robust expression of a therapeutic transgene in skeletal muscles, the diaphragm, and the heart, while exhibiting limited activity in non-target tissues. This review focuses on the development and improvement of muscle-specific promoters based on skeletal muscle α-actin, muscle creatine kinase, and desmin genes, as well as other genes expressed in muscles. The current approaches used to engineer synthetic muscle-specific promoters are described. Other elements of the viral vectors that contribute to tissue-specific expression are also discussed. A special feature of this review is the presence of up-to-date information on the clinical and preclinical trials of gene therapy drug candidates that utilize muscle-specific promoters.**
KEYWORDS **Gene therapy, muscle-specific promoters, AAV, natural promoters, synthetic promoters.**
ABBREVIATIONS **AAV – adeno-associated virus; PCT – preclinical trials; CT – clinical trials; LGMD – limb-girdle muscular dystrophy; UTR – untranslated region; TF – transcription factor; CMV – cytomegalovirus; MVM – minute virus of mice; TSS – transcription start site; TFBS – transcription factor binding site.**

INTRODUCTION

Inherited muscle disorders are diagnosed in 4–5 people per 20,000 [1]. These diseases include the clinically and genetically heterogeneous group of muscular dystrophies, congenital myopathies, lysosomal storage disorders, channelopathies, and mitochondriopathies. Weakness of skeletal muscles limits locomotor activity, pharyngeal muscle dysfunction causes swallowing difficulties, while heart failure or respiratory insufficiency is a primary cause of early death. Unfortunately, effective treatment for such genetic muscle disorders does not exist [2].

Many inherited muscle diseases are caused by a protein deficiency resulting from mutations in the corresponding gene (*Table*). A promising strategy for treating such disorders is *gene replacement therapy*, which delivers a genetic construct with a functional copy of a gene (transgene) into muscle tissues. Adeno-associated virus (AAV) vectors are considered to be the most promising and safe for *in vivo* delivery of therapeutic genes [3]. Naturally occurring AAV serotypes such as AAV9, AAV8, AAVrh74, and AAV1 have an intrinsic tropism for muscles and allow for better targeting of affected tissues. Examples of AAV-based gene therapy candidates for inherited muscle disorders are listed in *Table*.

The therapeutic effect of gene therapy largely depends on the transgene expression levels in the targeting tissues. On one hand, muscles are a convenient target for gene therapy due to the long lifespan of muscle fibers, easy access for intramuscular injections, as well as high protein synthesis capacity [12]. On the other hand, muscles make up to 30–40% of body weight; so, high doses of the gene therapy drug are required [13]. Moreover, muscle tissue is structurally heterogeneous and is subdivided into cardiac, skeletal, and smooth muscles. This complicates the development of gene therapy drugs that would be equally effective in different types of muscle tissues [14].

A properly-selected promoter for transgene expression is the key to a successful gene therapy. This promoter should confer long-term sustained high expression in muscles affected by the disease, while exhibiting limited activity in other tissues. The popu-

Inherited muscle disorders and the potential gene therapy

Disorder	Mutated gene	Inheritance pattern	Protein	Gene therapy drugs* in clinical and preclinical studies
Duchenne muscular dystrophy Becker muscular dystrophy	*DMD*	XR	Dystrophin	CT:AAVrh74.MHCK7.miDMD NCT03769116 CT:AAV9.CK8e.miDMD, NCT03368742 CT: AAV9.tMCK.miDMD NCT04281485
Danon disease	*LAMP2*	XR	LAMP2	PCT: AAV9.CAG.LAMP2B [4] CT: NCT03882437
Barth syndrome	*TAZ*	XR	Tafazzin	PCT: AAV9.Des.TAZ [5]
Myotubular myopathy	*MTM1*	XR	Myotubularin	PCT: AAV8.DES.hMTM1 [6] CT: NCT03199469
Primary merosin deficiency	*LAMA2*	AR	Merosin	PCT: AAV9.CB.mini-agrin [7]
Pompe disease	*GAA*	AR	α-1,4-Glucosidase	PCT: AAV2/8.MHCK7.hGAA [8] CT: AAV2/8.LSP.hGAA NCT03533673 CT: rAAV9.DES.hGAA NCT02240407
Limb-girdle muscular dystrophy LGMD, 2A	*CAPN3*	AR	Calpain 3	PCT: AAV9.desmin.hCAPN3 [9]
LGMD, 2B	*DYSF*	AR	Dysferlin	CT: rAAVrh.74.MHCK7.DYSF NCT02710500
LGMD, 2D	*SGCA*	AR	α-Sarcoglycan	CT: rAAV1.tMCK.hαSG NCT00494195 CT: scAAVrh74.tMCK.hSGCA NCT01976091
LGMD, 2E	*SGCB*	AR	β-Sarcoglycan	CT:scAAVrh74.MHCK7.hSGCB NCT03652259
LGMD, 2I	*FKRP*	AR	Fukutin-related protein	PCT: AAV9.Des.mFkrp [10]
Oculopharyngeal muscular dystrophy	*PABPN1*	AD	PABPN1	PCT: AAV9.spc512.PABPN1 [11]

*Drug candidate name includes information about AAV serotype, promoter and transgene.
Note: AD – autosomal dominant; AR – autosomal recessive; XR – X-linked recessive; PCT – preclinical trials; CT – clinical trials.

larity of AAV as a gene therapy vector makes necessary a reduction of the promoter size because of the limited packaging capacity of the virus (4.7 kbp) [3]. Strong constitutive promoters, such as the promoters of the respiratory syncytial virus (RSV), cytomegalovirus (CMV), or elongation factor 1a (EF1a), are compact in size and achieve high expression levels in a variety of tissues. However, it has been demonstrated that expression in non-target tissues, especially in antigen-presenting cells, triggers an immune response to the transgene and induces cytotoxicity [15]. Furthermore, viral promoters are prone to transcriptional silencing in transduced cells due to methylation [16].

This review focuses on strategies for designing and improving natural muscle-specific promoters. It highlights current approaches to the engineering of synthetic promoters, and it discusses their application in gene therapy constructs.

THE STRUCTURE OF A EUKARYOTIC PROMOTER

Eukaryotic gene transcription is controlled by two classes of regulatory elements: promoters with core and proximal regions and distal regulatory elements (*Fig. 1*) [17].

The core (basal, or minimal) promoter is a specialized DNA sequence which directs transcription initiation and is located -50 to +50 bp from the transcription start site (TSS) [18]. There are several types of core promoters. The focused core promoter is a promoter with a single, well-defined TSS. The promoter with several closely positioned TSS within the 50–100 bp region is called dispersed [19]. The focused type is predominantly observed in promoters of tissue-specific genes, while dispersed core promoters are typical of universally expressed genes [18].

The core promoter supports the assembly of the preinitiation complex consisting of RNA polymerase II and basal transcription factors (TFs). Core promoters differ widely in terms of the conserved motifs that define their properties. Initiator (Inr) is the most common element of core promoters. The Inr sequence surrounds the TSS and is recognized by the multiprotein transcription factor II D (TFIID) [20]. Another well-characterized element of the core promoter is the TATA box. Approximately 28% of focused promoters in humans carry the TATA-like sequence [19]. This element is recognized by the TBP subunit of the transcription factor TFIID [21]. In promoters without the TATA box, Inr

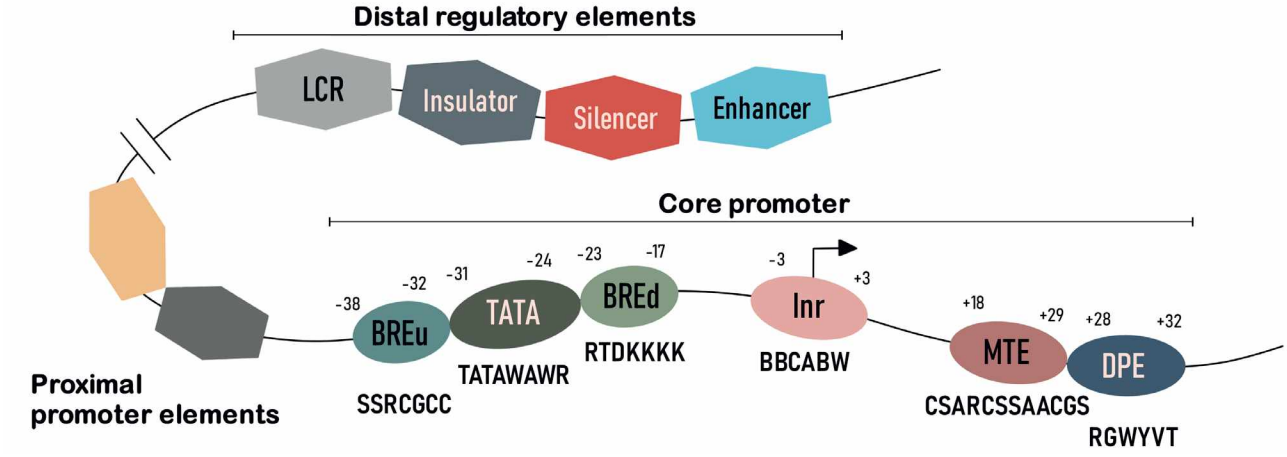


Fig. 1. The structure of a eukaryotic promoter. The eukaryotic promoter consists of the core promoter, the proximal promoter elements, and distal regulatory elements. In the core promoter conserved motifs are shown with consensus sequences and the position from the transcription start site

is often accompanied by the DPE motif (downstream promoter element), which is located downstream of the initiator and recognized by other subunits of TFIID [22]. The MTE (motif ten element) lies close to the DPE or overlaps with it [23]. Other typical elements of the core promoter include BREu (the upstream TFIIB recognition element) and BREd (the downstream TFIIB recognition element) [24].

The proximal promoter typically encompasses ~50–1,000 bp upstream of the TSS and contains many transcription factor binding sites (TFBSs) [25]. The unique combinations of these TFBSs in each promoter allow for tight regulation of the expression levels of ~25,000 human genes controlled by as few as ~1,600 TFs [26].

The distal regulatory elements of the eukaryotic gene include enhancers, silencers, insulators, and the locus-control regions (LCRs). Enhancer elements are of particular interest for promoter engineering. Enhancers are DNA sequences ~100- to 1,000-bp long that can increase the transcription of genes regardless of their orientation and distance to the target promoter [27]. These elements can be found in the 5' and 3' untranslated regions (UTRs) of the genes, within exons and introns, and even at a distance as large as 1 Mbp from the TSS [28]. Many enhancers are highly conserved sequences whose activity can be confined to a certain tissue or cell type, developmental stage, or certain physiological conditions [25].

NATURAL PROMOTERS

A straightforward way to design a muscle-specific promoter is to use the naturally occurring promoter of the gene with high expression levels in muscles. To reduce the size of the full-length natural promoter, only the core promoter and some proximal elements are left and supplemented with distal enhancers [29]. Poorly conserved sequences are typically excluded from the design; the importance for the expression of the remaining promoter elements was verified by mutation analysis [29]. A similar approach is creating hybrid/chimeric promoters by adding the enhancer elements of one gene to the promoter region of another gene [30]. The expression level and tissue specificity can be significantly improved by varying the copy number of the enhancers and individual TFBS and properly combining these sequences [31].

Human skeletal α-actin promoters

In their early attempts to create muscle-specific promoters, researchers focused on the promoter regions of proteins abundant in myocytes. Actin is the main protein that constitutes the sarcomere, the basic contractile unit of striated muscle. In higher vertebrates, six major isoforms of actin are distinguished, each encoded by a separate gene: skeletal and cardiac muscle α-actin, smooth muscle α-actin, smooth muscle γ-actin, and two isoforms with ubiquitous expression, cytoplasmic β-actin, and cytoplasmic γ-actin [32]. The human skeletal muscle α-actin gene (HSA) attracted the most interest from researchers, since this actin isoform prevails in adult muscles [33].

The first studies demonstrated that the region located 2,000 bp upstream of the *HSA* gene, as well as the first exon and the fragment of the first intron, is necessary and sufficient for muscle-specific expression in a cell culture (*Fig. 2A*) [34]. Three major promoter regions have been identified: the distal (from -1300 to -626 from the TSS), the proximal (-153...-87), and the basal

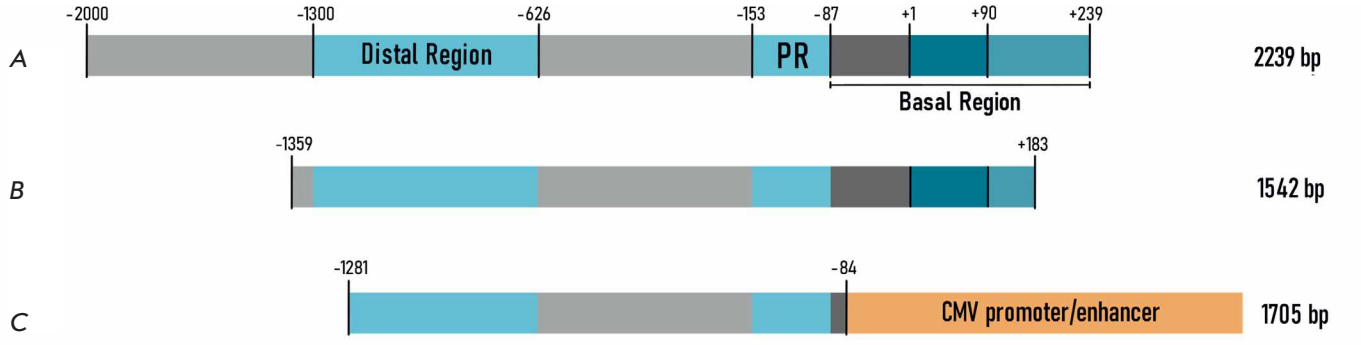


Fig. 2. Promoters based on the *ACTA1/HSA* gene. (*A*) – the full-length HSA promoter includes the distal region, the proximal region (PR), and the basal region, which consists of the noncoding exon (+1…+90) and the first intron fragment (+91…+239); (*B*) – shortened version of the HSA promoter; (*C*) – the chimeric HSA/CMV promoter consisting of a fragment of the HSA promoter and the CMV promoter

(-87…+239) regions. These promoter regions, lumped together or separately before the SV40 promoter, drive tissue-specific expression [34].

The fragment of *HSA* gene extending from c -2,000 bp to +7,500 bp (promoter region -2,000…+239 as in *Fig. 2A*) was used to generate a transgenic mouse line in [35]. It was demonstrated for the first time that the transgene expression was comparable to the expression level of endogenous mouse skeletal muscle α-actin in the striated muscles and the heart. The HSA promoter has become rather popular and has been used in a number of studies. For example, it was employed to produce transgenic mice carrying dystrophin gene deletions [36], mice with dysferlin overexpression [37], and a mouse model of spinal muscular atrophy [38], as well as to deliver microdystrophin into mouse muscles using lentiviral vectors [39].

Another truncated variant of the human HSA promoter was used in the AAV vector to treat Duchenne muscular dystrophy [40]. A high expression level in the muscles was achieved using a 1,542-bp fragment consisting of a distal region, a promoter, and a portion of the first intron (*Fig. 2B*). The transgene was actively expressed in skeletal muscles and the heart, but no transgene expression was detected in the liver.

The chimeric HSA promoter was used to produce coagulation factor IX in muscles and treat hemophilia B [41]. This promoter was a fragment of the HSA promoter (-1281…-84) ligated to the CMV promoter (*Fig. 2C*). In the myoblast cell culture, the transgene expression level ensured by this promoter was higher than the transgene expression levels induced by the CMV promoter and the full-length HSA promoter, while this promoter was as active as the CMV promoter in nonmuscle cell cultures. It appears that, although the addition of the universally expressed CMV promoter increased the activity of the chimeric promoter, it became tissue non-specific.

The regulatory regions of the homologous chicken [42], rat [43], and bovine [44] genes were modified to design muscle-specific promoters similar to the HSA one. The resulting constructs have been successfully used in *in vitro* and *in vivo* experiments, as well as to generate transgenic mice.

In general, the skeletal muscle α-actin promoters exhibited a high expression level and specificity in muscle cells; however, they are used in modern studies not so frequently, because of their large size.

Muscle creatine kinase promoters

The transcript of the muscle creatine kinase (*MCK/CKM,* creatine kinase, M-type) gene is the second-most abundant mRNA in skeletal muscles [45]. MCK catalyzes reversible phosphoryl transfer from ATP to creatine and from creatine phosphate to ADP, thus providing energy for muscle contractions. The *MCK* gene is also highly active in the cardiac muscle and is transcriptionally activated during the differentiation of myoblasts into myocytes [46].

The *MCK* promoter has been characterized well both *in vitro* and *in vivo*. One of the major regulatory regions of the mouse *Mck* gene is the muscle-specific 206-bp enhancer located within the -1256…-1050 region [47]. This enhancer exerts its function regardless of orientation and carries a number of binding sites for myogenic transcription factors (namely, E-boxes, CArG, and MEF2 sites). A mutation analysis of these motifs has confirmed their importance in muscle-specific expression [48].

The proximal promoter (358 bp) is the key regulatory element of the *MCK* gene [47]. As such, it alone ensured a high expression level of the transgene in limb muscles and abdominal skeletal muscles in mice but was inactive in cardiac and tongue muscles. Nevertheless, when the 206-bp enhancer and the 358-bp promoter were ligated together, expression was restored

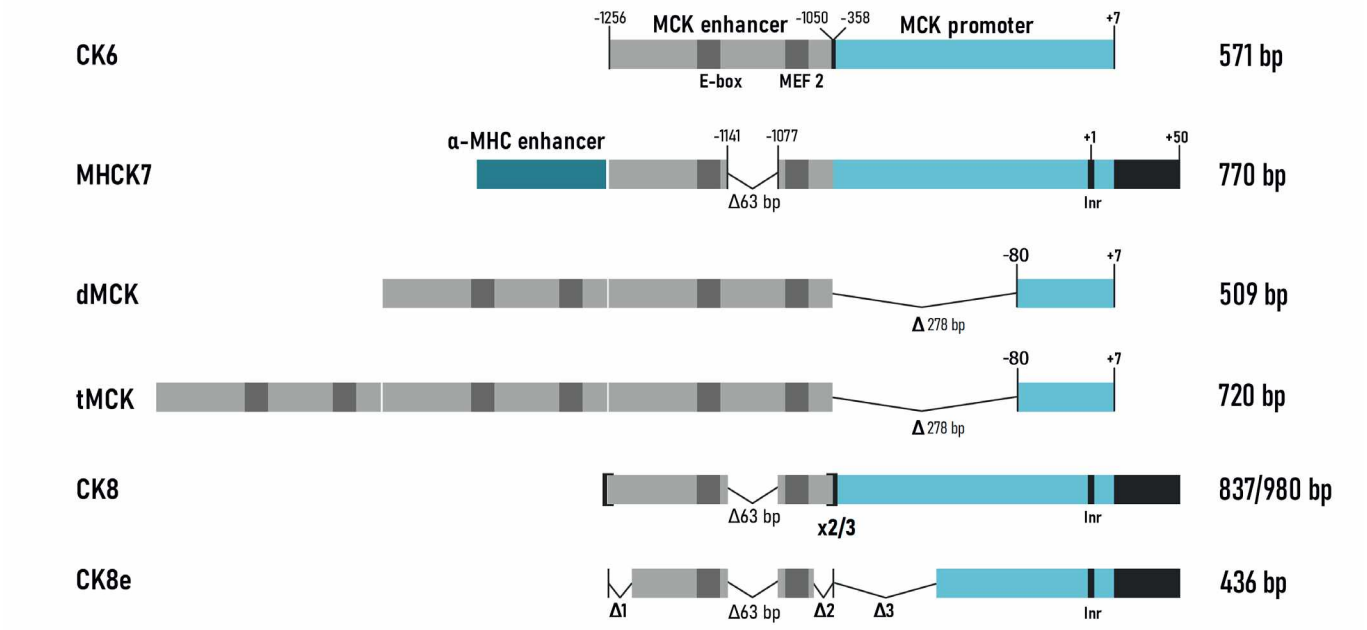


Fig. 3. *Mck*-based promoters. All the constructs contain the MCK enhancer and the MCK promoter, with different modifications

in all types of muscles [47]. The data obtained for transgenic mice have proved that the enhancer is required in order to induce expression in the heart [49].

Based on the discussed-above regulatory sequences, several series of small MCK expression cassettes for adenoviral vectors were developed and tested [50]. Thus, the **construct CK6,** consisting of an enhancer (206 bp) and a proximal promoter (358 bp) (*Fig. 3*), ensured high muscle specificity. However, the expression level in the muscles was ~12% compared to that attained using a similar construct with the CMV promoter; the expression level in the heart remained low [50].

The chimeric promoter **MHCK7** was developed to achieve a high expression level of the transgene in the cardiac muscle (*Fig. 3*) [29]. It included a 206-bp enhancer and a proximal promoter but contained four important modifications. Thus, the poorly conserved region between the right E-box and the MEF2 site was deleted in the 206-bp enhancer. The highly conserved 50-bp sequence from the first noncoding exon of *MCK* was added to the promoter. The TSS-containing sequence was replaced with the Inr consensus sequence. The most important modification was the following: a 188-bp enhancer from the mouse α-myosin heavy chain gene (α-*Mhc*), which ensures a high expression level in the heart, was added to the expression cassette described above [51]. The new MHCK7 promoter was tested in the context of AAV vectors. The promoter ensured a transgene expression level comparable to those for the CMV and RSV promoters in skeletal and cardiac muscles. Low expression levels were observed in the liver, lungs, and spleen after AAV6 had been intravenously injected to mice. Interestingly, the MHCK7 promoter was 400 and 50 times more active in the heart and the diaphragm, respectively, than promoter CK6.

The **dMCK** and **tMCK** promoters (*Fig. 3*) were developed in another laboratory, almost simultaneously with the MHCK7 promoter [52]. In these constructs, the proximal promoter (358 bp) was shortened to a 87-bp basal promoter (-80...+7) and two or three copies of the MCK enhancer (206 bp) were ligated to it. In the experiments where the transgene was delivered using AAV vectors, tMCK proved to be the most efficient promoter; the level of muscle-specific expression it ensured was higher than the expression levels ensured by the CMV, dMCK, and CK6 promoters. However, the dMCK promoter did not activate transgene expression in the heart or diaphragm.

However, the search for efficient muscle-specific promoters continued. The constructs named **CK8** (*Fig. 3*) were developed at the very same laboratory where the CK6 and MHCK7 promoters had been previously created. Thus, the CK8 promoter (MHCK7 with two copies of the MCK enhancer instead of the α-myosin heavy-chain enhancer) was used for intramuscular AAV8-mediated delivery of the growth hormone gene [53]. In mice treated with this construct, body length and weight were significantly higher compared to those in untreated mice. In another study, the CK8 promoter was similar to the construct described above but carried three copies of the MCK enhancer [31]. It was reported that using three instead of two

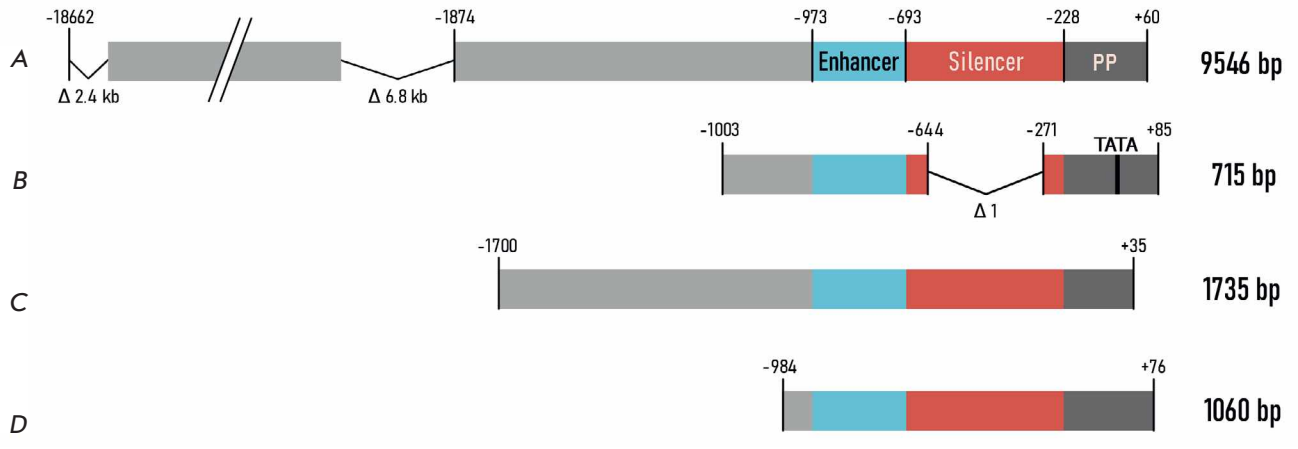


Fig. 4. Promoters based on the human *DES* gene. Promoter (*A*) includes the locus control region of the desmin gene (18.7 kbp) with introduced deletions, the enhancer, the silencer, and the proximal promoter (PP). Promoter (*B*) contains a deletion in the silencer and the TATA box added to the core promoter. Promoters (*C*) and (*D*) have deletions in the distal regions

enhancer copies increases the expression levels in skeletal myocytes and in the heart four- and threefold, respectively [31]. The 436-bp **CK8e** construct carried a number of deletions in the enhancer and proximal promoter of muscle-type creatine kinase (*Fig. 3*) and was more active than the CMV promoter in a differentiated human myoblast culture [54]. Deletion of the poorly conserved regions in the promoter reduced its length and simultaneously increased its activity [31].

Due to a high specificity and activity in muscle tissues, promoters based on the *MCK* gene are widely used in the gene therapy vectors that are currently undergoing preclinical and clinical testing (*Table*). The MHCK7 promoter is included in a vector undergoing preclinical studies for the treatment of the Pompe disease [8]. Clinical trials to treat LGMD type E (NCT03652259), where a functional copy of the β-sarcoglycan is delivered into patients under the control of the MHCK7 promoter, are currently underway [55]. The MHCK7 promoter is also being used as part of a construct to deliver the dysferlin gene (NCT02710500) [56]. In the ongoing clinical studies (NCT03769116) focused on the treatment of Duchenne muscular dystrophy the microdystrophin gene is delivered into patients under the control of the MHCK7 promoter [57]. The CK8e promoter was used in a clinical trial focusing on microdystrophin delivery using AAV9 NCT03368742) in [58]; another clinical study (NCT04281485) focused on the tMCK promoter and AAV9-mediated delivery of the microdystrophin gene [57]. A clinical trial (NCT01976091) evaluating the delivery of the α-sarcoglycan gene under the control of the tMCK promoter in the treatment of LGMD type 2D is also underway.

Desmin gene promoters

Desmin is a muscle-specific cytoskeletal protein belonging to the intermediate filament family [59]. It is encoded by the *DES* gene and is one of the earliest myogenic markers [59]. This protein is unique in that it is expressed in satellite cells and dividing myoblasts, while its abundance in differentiated muscle cells is several times higher [60].

A functional analysis of the 5'-flanking region of the human desmin gene revealed an enhancer (-973...-693) [60]. The 5'-region of the enhancer contains the MEF2-binding sites, the E-box, and the Mt element; it is needed for activating expression in muscle fibers. The 3'-half of the enhancer is responsible for desmin transcription in myoblasts because of binding to SP1 and KROX-20 [60]. The region -692...-228 is a silencer, which reduces expression in myoblasts and muscle fibers by up to 3- to 7-fold; the region -228...+75 was sufficient to initiate a low-level muscle-specific expression [61, 62].

The full-length dystrophin gene under the control of the human desmin promoter (9546 bp; region -18662...+60) in a plasmid vector was used for intra-arterial delivery in mice (*Fig. 4A*) [63]. This promoter ensured the same level of dystrophin expression as the CMV promoter did for at least 6 months.

A variant of the human desmin promoter (715 bp) with a deleted silencer was used in a comparative study of muscle-specific promoters for intravenous AAV9-mediated transgene delivery (*Fig. 4B*) [64]. In that study, the desmin promoter was superior to the CMV promoter and other muscle-specific promoters in terms of the transgene expression level attained in skeletal muscles and the diaphragm. In terms of the expression

level in the heart, it was less effective only than the CMV promoter, while it still ensured a high level of transgene expression in the brain [64].

Another variant of the human desmin promoter was used in a study focused on transgene delivery in mouse muscles using lentiviral vectors [65]. The region (-1700...+35) containing a promoter, a silencer, and an enhancer was used (*Fig. 4C*). The activity of this promoter was comparable to that of the CMV promoter in experiments *in vitro* and *in vivo*, being even higher than that of the human *MCK* promoter (-1061...+28).

Desmin promoters were used in a number of studies (*Table*) focused on the Pompe disease [66]. The construct rAAV9.DES.hGAA, intended to treat this disorder, is so far successfully undergoing clinical trials (NCT02240407). A promoter variant [64] was used in preclinical studies to develop gene therapy drugs for patients with the Barth syndrome [5]. Furthermore, the desmin promoter was used in preclinical studies as a vector to treat LGMD type 2A (calpain 3 deficiency) [9] and LGMD type 2I (fukutin-related protein deficiency) [10]. The human desmin promoter (-984...+76) (*Fig. 4D*) was evaluated in preclinical studies focused on the therapy of myotubular myopathy, a genetic disorder caused by mutations in the *MTM1* gene [6]; this medicinal product has almost completed clinical trials (NCT03199469).

Promoters based on other genes

The regulatory regions of many other genes that exhibit muscle-specific expression were also used to construct promoters. A search for a candidate promoter was simultaneously conducted in a number of laboratories, but only a few studies proved successful.

For instance, to treat the cardiac variant of the Fabry disease, lentiviral constructs with cardiac-specific transgene expression were developed [67]. Three different promoters were tested: the human α-myosin heavy chain gene (αMHC) promoter (region -1198...+1), the myosin light-chain promoter (**MLC2v**) (-250...+13), and the cardiac troponin T promoter (**cTnT**) (-300...+1). All three promoters were superior to the ubiquitous EF1α promoter in their expression levels of transgene in the heart. Besides the cardiac expression, the cTnT and MLC2v promoters also drove expression in the liver and spleen, while the transgene, under the control of αMHC, was active exclusively in the heart [67]. However, another study revealed that promoters based on these genes were inferior to the desmin promoter in terms of their expression level in skeletal muscles and the heart [64].

A chimeric promoter comprising the CMV-IE enhancer ligated to the 1.5-kbp fragment of the rat promoter **MLC** was used for AAV9-mediated delivery of microdystrophin in a mouse heart [68]. The cardiac activity of this promoter was four times as high as that of the CMV, and robust transgene expression in the heart was conferred for 10 months, but not in the skeletal muscles or the liver [69].

The ΔUSEx3 promoter was developed on the basis of the human troponin I (*TNNI1*) gene and consisted of three copies of the enhancer (-1036...-873) and the minimal promoter of the *TNNI1* gene with a portion of the first exon (-95...+56) [70]. The ΔUSEx3 promoter exhibited weak activity in nonmuscle cells and tissues in *in vivo* and *in vitro* experiments. Let us mention that the ΔUSEx3 promoter delivered by adenoviruses ensured a transgene expression level comparable to that induced by the synthetic SPcΔ5-12 promoter [71]; however, ΔUSEx3 delivered by lentiviruses was five times less active than the SPcΔ5-12 promoter, probably due to the effects related to its integration into the genome.

The unc45b gene encoding the muscle-specific myosin chaperone in fish was also used to develop a muscle-specific promoter [72]. Thus, the 195-bp promoter fragment (-505...-310) was able to induce expression in skeletal and cardiac muscles in fish and ensured reporter protein expression in mouse muscles when the plasmids were delivered by electroporation.

To summarize, the promoters discussed in this section proved capable of driving muscle-specific expression but were less potent than the promoters based on the actin, muscle creatine kinase, or desmin genes.

Synthetic promoters

A groundbreaking approach in promoter design is the creation of novel *synthetic promoters*. This strategy enables one to engineer promoters with defined properties, such as size and the expression profile of the transgene.

The development of synthetic promoters relies on computational algorithms, which are used to identify regulatory sequences and TFBSs within the genome, as well as to predict the promoter regions [73–75]. The binding sites for myogenic TFs are usually shorter than 10 bp [74], which allows one to create a library of constructs with different combinations of muscle-specific TFBSs. The key challenge in this approach is to analyze large libraries of novel synthetic constructs, which can be labor-intensive. Experiments are needed in order to determine the number of copies of the target motif and the distances between TFBS required for a successful binding of the transcription factor; not to mention identify the motifs having a synergistic function and the TFBS making the greatest contribution to expression enhancement. In order to overcome these obstacles, one can return to the analysis of natural promoters: extract the functioning combinations of muscle-specific TFBS

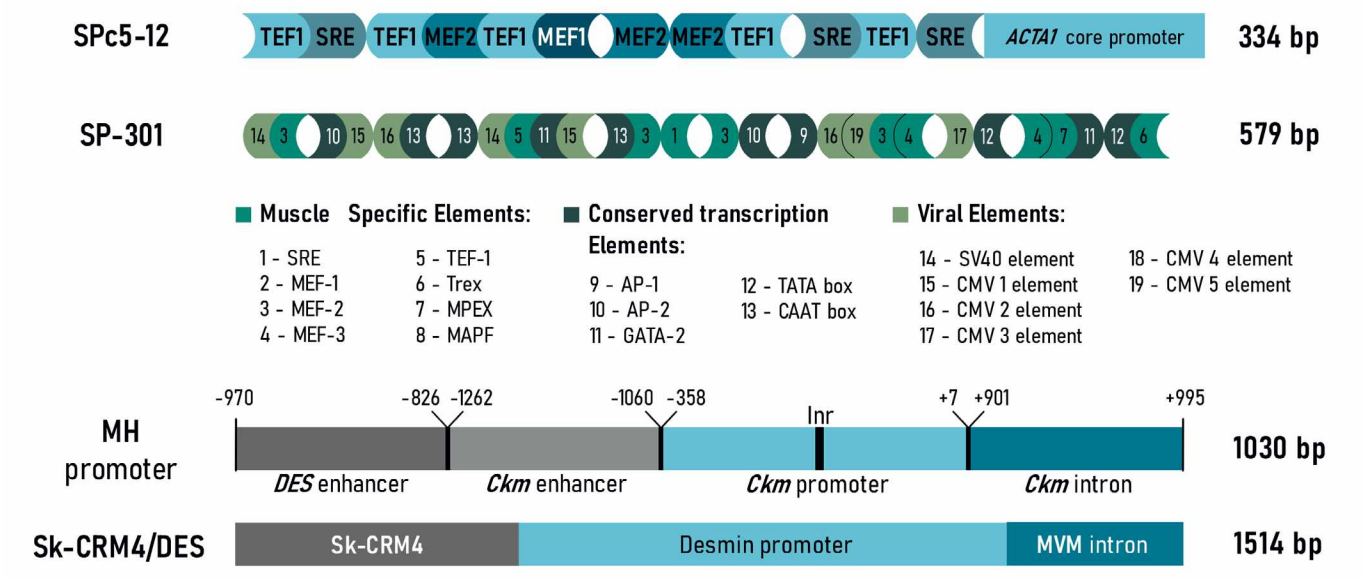


Fig. 5. Synthetic promoters. The SPc5-12 promoter consists of a combination of four muscle-specific TFBSs (TEF1, SRE, MEF1, and MEF2) and the core promoter (a fragment of the promoter of the chicken skeletal muscle α-actin gene). The SP-301 promoter is a combination of muscle-specific TFBSs, viral elements, and conserved *cis*-regulatory elements ligated in forward and reverse orientation. The MH promoter consists of the human desmin gene enhancer linked to the enhancer, the core promoter, and the first intron of the mouse *Ckm* gene. Sk-CRM4/Des is the regulatory module Sk-CRM4 ligated to the desmin promoter and the MVM intron

and construct promoters from similar clusters. An *in silico* analysis can substantially simplify the detection of regulatory regions.

In their pioneering study, Li *et al.* analyzed the sequences of strong muscle-specific promoters and identified the common binding sites of myogenic TFs (SRE, MEF-2, MEF-1, and TEF-1) within their structure [71]. These TFs were randomly ligated to each other in forward and reverse orientation, and the resulting fragments were inserted upstream of the minimal promoter of the chicken skeletal muscle α-actin gene (*Fig. 5*). As a result, a library consisting of more than 1,000 promoter variants was created. The synthetic promoter library was screened in primary myoblasts, and, based on the results, the **SPc5-12** promoter was selected (*Fig. 5*). SPc5-12 activity in muscle fibers was sixfold higher than that of the CMV promoter. The *in vivo* experiments confirmed that the SPc5-12 promoter is inactive in undifferentiated myoblasts and in various nonmuscle cell lines.

The SPc5-12 promoter was used to drive transgene expression in animal models of Duchenne muscular dystrophy [76], the Pompe disease [77], and dysferlinopathy [78], as well as to ensure growth hormone expression [79]. A gene therapy construct with the SPc5-12 promoter was utilized in preclinical studies for the treatment of oculopharyngeal muscular dystrophy [11].

Liu *et al.* used a similar strategy to construct synthetic promoters [80]. The promoters were designed from 19 elements, including eight muscle-specific TFBS, six viral elements (CMV and sv40 promoters), and five conserved *cis*-regulatory elements of eukaryotic promoters (TATA box, etc.). These motifs were randomly assembled to construct a library consisting of 1,200 primary clones, which were tested *in vitro* and *in vivo*. The strongest transcriptional activity was achieved with **the SP-301** promoter (*Fig. 5*); it was 6.6 times more active than the CMV promoter 2 days after intramuscular delivery of the construct in mice and remained active for at least a month. Many promoters achieved a higher *in vitro* activity compared to the CMV promoter, but they were less active *in vivo*. The tissue specificity of the SP-301 promoter was confirmed in transgenic mice. This study once again demonstrated the advantage of the strategy of designing synthetic promoters using a combination of TFBSs and also highlighted the the benefit of including viral motifs besides muscle-specific TFBSs for enhanced expression levels.

Another efficient approach, consisting in designing hybrid promoters, has already been partially discussed for the MHCK7 promoter [29]. In the study where this strategy was employed [30], Piekarowicz *et al.* conducted an *in silico* analysis of various tissue-specific genes and identified four clusters that con-

sisted of a combination of the binding sites of myogenic transcription factors. The first cluster was the previously discussed enhancer of the human desmin gene (-970...-826) [81]; the remaining three clusters were regions of the enhancer (-1256...-1051), the proximal promoter (-358...+7), and the first intron (+901...+995) of the mouse *Ckm* gene. The promoter that contained all four elements (the **MH promoter**) ensured the highest expression level in the muscle cell culture, being superior to the desmin and CMV promoters, as well as the remaining hybrid promoters. Interestingly, intron made the greatest contribution to the expression level, while deletion of one or the two enhancers or even the core promoter did not significantly alter the expression level. To test the *in vivo* activity of the hybrid promoter, AAV2/9 carrying the reporter gene was delivered intravenously in mice under the control of this promoter. The activity of the MH promoter in the cardiac and skeletal muscles was higher than that of the desmin and CMV promoters; however, the MH promoter did not induce transgene expression in the liver [30].

The strategy of using muscle-specific *cis*-regulatory modules (**Sk-CRM**) was employed in the next study [82]. TFBSs were mapped in the promoters of human genes highly expressed in skeletal muscles and analyzed for their tendency to form clusters. To identify conserved motifs, these clusters were subjected to multiple-sequence alignment across various animal species. It was assumed that the TFBS combinations conserved in evolution are more likely to retain potency and specificity following clinical translation. Open chromatin structure and the accessibility of candidate Sk-CRMs to TFs were also taken into consideration. Using this computational approach, seven novel evolutionarily conserved muscle-specific Sk-CRMs modules were identified and cloned upstream of the desmin promoter. Based on the results of a bioluminescence assay, the **Sk-CRM4** module was selected for further studies (*Fig. 5*). Six weeks after systemic delivery using AAV9, the Sk-CRM4 chimeric promoter enhanced the activity of the desmin promoter by 200–400 times in different skeletal muscles, the diaphragm, and the heart, while remaining inactive in non-target tissues. Moreover, the SkCRM4/Des promoter attained a 25–173 times higher expression in different muscles as compared to the CMV promoter and also outperformed the Sk-CRM4/SPc5-12 and SPc5-12 promoters. Therefore, the computationally designed Sk-CRM4/Des chimeric promoter demonstrated improved muscle-specific performance as compared to the other promoters commonly used for muscle gene therapy, with length (~1,500 bp) being its only drawback.

To conclude, implication of synthetic promoters in gene therapy holds great promise. Success in synthetic promoter engineering largely depends on the quality of the bioinformatic tools and efficient screening of the proposed variants. Expansion of the regulatory element databases, revealing new TFs, and improvement in the software for promoter identification will undoubtedly contribute to further development in this area of research [73–75].

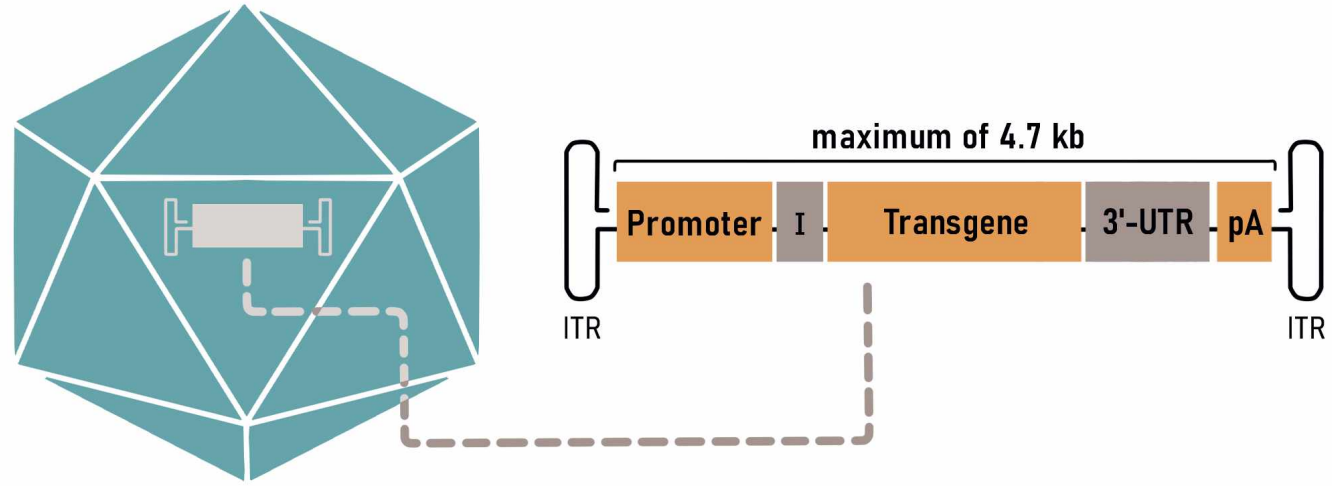


Fig. 6. Typical elements of the AVV expression cassette. Orange blocks (the promoter, the transgene, and polyadenylation signal (pA)) are the basic components of the cassette. Accessory *cis*-regulatory elements, such as intron (I), WPRE, and the microRNA binding sites (3′-UTR), can also be inserted to enhance expression efficiency. The cassette is flanked by inverted terminal repeats (ITR)

OTHER FACTORS DETERMINING MUSCLE-SPECIFIC EXPRESSION

When developing gene therapy drugs, one should comprehensively evaluate the expression of the genes of interest at the proper position and time, as this depends not only on promoter activity, but on other factors as well. Many factors affect the transgene expression at the post-transcriptional level.

Expression of the target gene can be enhanced due to the presence of an intron in the vector, which is usually positioned between the promoter and the coding region (*Fig. 6*). The presence of the intron increases RNA stability in the nucleus due to the incorporation of mRNA into the spliceosome [74] and promotes efficient export of spliced mRNA from the nucleus to the cytoplasm [83]. Introns can also contain regulatory sequences that affect tissue specificity and the expression level. A study focused on designing a chimeric promoter [30] showed that the presence of intron from the *Ckm* gene makes the greatest contribution to the transgene expression level. The MVM intron enhanced transgene expression during an AAV-mediated delivery of coagulation factor IX more than 80-fold compared to the construct without intron [84].

Along with the promoters, other *cis*-regulatory elements can also be added to the 3'-UTR of the expression cassette to enhance expression (*Fig. 6*). Thus, the 600-bp post-transcriptional regulatory element of the woodchuck hepatitis virus (WPRE) delivered using AAV led to a manifold enhancement of transgene expression in the liver, brain, and muscles [85]. WPRE promotes mRNA export from the nucleus and prevents post-translational gene silencing [86].

A different approach can also be used to achieve tissue specificity: not only inducing expression in the target tissues, but also suppressing it in non-target organs through RNA interference mechanisms [74]. For this purpose, the binding sites of microRNA that are present only in the non-target organs are added to the 3'-UTR of the expression cassette (*Fig. 6*) [87]. If transgenic mRNA is expressed in a non-target organ, microRNA binds to the complementary sites on the transgene and initiates its degradation [87].

A proper choice of viral vector also plays a significant role in the delivery of the transgene into the target organs and tissues. Along with the naturally occurring serotypes of AAVs (*Table*), capsids are also modified to design novel, genetically engineered vectors and improve the targeted delivery [88]. There is an ongoing search for other naturally occurring capsids with improved tropism for the heart and skeletal muscles [89]. Transgene expression patterns also differ depending on the route of administration (intravenous, intramuscular, etc.) [90]. An elaborate combination of the above-mentioned elements in the cassette, proper choice of the viral vector, and an optimal delivery route for the genetically engineered drug can significantly enhance the expression of the gene of interest, while maintaining tissue-specific expression.

CONCLUSIONS

The efforts to develop optimal muscle-specific promoters started more than 30 years ago and are still underway. Early versions of natural muscle-specific promoters consisted of the regulatory regions of the actin, desmin, and muscle creatine kinase genes and exceeded 1 kbp in length. The latest generations of synthetic promoters contain combinations of TFBS from common muscle-specific genes, are much shorter, but the expression efficiency achieved by these promoters is comparable to or higher than that of natural promoters [30, 71].

It has been demonstrated in many studies that transcription factors and their binding sites in vertebrates are appreciably conserved [42, 72]. Thanks to this property of promoters, various animal models can be used in preclinical studies to prove the effectiveness of gene therapy drugs. However, when conducting *in vitro* studies, one should keep in mind that promoter activity in this case does not necessarily coincide with *in vivo* activity [80].

Since the group of genetic muscular disorders is heterogeneous, there is no universal promoter that could be used to develop vectors intended for the treatment of all diseases. This can be largely attributed to the features of the pathogenesis and the different functions of the proteins whose deficiency or dysfunction causes a given disorder (*Table*). Different muscle groups and types of muscle fibers are affected in patients with different disorders [14]. The gained experience in developing muscle-specific synthetic promoters provides hope that researchers will eventually design ideal constructs that mimic the unique expression profile of muscle-specific proteins and fully restore their lost functions. ●

REFERENCES

1. Theadom A., Rodrigues M., Poke G., O'Grady G., Love D., Hammond-Tooke G., Parmar P., Baker R., Feigin V., Jones K., et al. // Neuroepidemiology. 2019. V. 52. № 3–4. P. 128–135.
2. Shieh P.B. // Neurol. Clin. 2013. V. 31. № 4. P. 1009–1029.
3. Anguela X.M., High K.A. // Annu. Rev. Med. 2019. V. 70. P. 273–288.
4. Manso A., Hashem S., Nelson B., Gault E., Soto-Hermida A., Villarruel E., Brambatti M., Bogomolovas J., Bushway P., Chen C., et al. // Sci. Transl. Med. 2020. V. 12. P. eaax1744.
5. Suzuki-Hatano S., Saha M., Soustek M.S., Kang P.B.,

Byrne B.J., Cade W.T., Pacak C.A. // Mol. Ther. Methods Clin. Dev. 2019. V. 13. P. 167–179.
6. Childers M.K., Joubert R., Poulard K., Moal C., Grange R.W., Doering J.A., Lawlor M.W., Rider B.E., Jamet T., Danièle N., et al. // Sci. Transl. Med. 2014. V. 6. № 220. P. 220ra10.
7. Qiao C., Dai Y., Nikolova V.D., Jin Q., Li J., Xiao B., Li J., Moy S.S., Xiao X. // Mol. Ther. Methods Clin. Dev. 2018. V. 9. P. 47–56.
8. Han S., Li S., Bird A., Koeberl D. // Hum. Gene Ther. 2015. V. 26. № 11. P. 743–750.
9. Lostal W., Roudaut C., Faivre M., Karine C., Suel L., Bourg N., Best H., Smith J., Gohlke J., Corre G., et al. // Sci. Transl. Med. 2019. V. 11. P. eaat6072.
10. Gicquel E., Maizonnier N., Foltz S.J., Martin W.J., Bourg N., Svinartchouk F., Charton K., Beedle A.M., Richard I. // Hum. Mol. Genet. 2017. V. 26. № 10. P. 1952–1965.
11. Malerba A., Klein P., Bachtarzi H., Jarmin S.A., Cordova G., Ferry A., Strings V., Espinoza M.P., Mamchaoui K., Blumen S.C., et al. // Nat. Commun. 2017. V. 8. P. 14848.
12. Marshall D.J., Leiden J.M. // Curr. Opin. Genet. Dev. 1998. V. 8. № 3. P. 360–365.
13. Janssen I., Heymsfield S.B., Wang Z.M., Ross R. // J. Appl. Physiol. Bethesda Md 1985. 2000. V. 89. № 1. P. 81–88.
14. Gaina G., Budisteanu M., Manole E., Ionica E. // Muscular Dystrophies. 2019.
15. Shirley J.L., de Jong Y.P., Terhorst C., Herzog R.W. // Mol. Ther. J. Am. Soc. Gene Ther. 2020. V. 28. № 3. P. 709–722.
16. Duan B., Cheng L., Gao Y., Yin F.X., Su G.H., Shen Q.Y., Liu K., Hu X., Liu X., Li G.P. // Theriogenology. 2012. V. 78. № 4. P. 793–802.
17. Razin S.V., Gavrilov A.A., Ulianov S.V. // Molecular Biology (rus). 2015. V. 49. № 2. P. 212.
18. Haberle V., Stark A. // Nat. Rev. Mol. Cell Biol. 2018. V. 19. № 10. P. 621–637.
19. Vo Ngoc L., Wang Y.-L., Kassavetis G.A., Kadonaga J.T. // Genes Dev. 2017. V. 31. № 13. P. 1289–1301.
20. Vo Ngoc L., Cassidy C.J., Huang C.Y., Duttke S.H.C., Kadonaga J.T. // Genes Dev. 2017. V. 31. № 1. P. 6–11.
21. Sainsbury S., Bernecky C., Cramer P. // Nat. Rev. Mol. Cell Biol. 2015. V. 16. № 3. P. 129–143.
22. Burke T.W., Kadonaga J.T. // Genes Dev. 1997. V. 11. № 22. P. 3020–3031.
23. Lim C.Y., Santoso B., Boulay T., Dong E., Ohler U., Kadonaga J.T. // Genes Dev. 2004. V. 18. № 13. P. 1606–1617.
24. Deng W., Roberts S.G.E. // Genes Dev. 2005. V. 19. № 20. P. 2418–2423.
25. Maston G.A., Evans S.K., Green M.R. // Annu. Rev. Genomics Hum. Genet. 2006. V. 7. P. 29–59.
26. Lambert S.A., Jolma A., Campitelli L.F., Das P.K., Yin Y., Albu M., Chen X., Taipale J., Hughes T.R., Weirauch M.T. // Cell. 2018. V. 172. № 4. P. 650–665.
27. Long H.K., Prescott S.L., Wysocka J. // Cell. 2016. V. 167. № 5. P. 1170–1187.
28. Riethoven J.-J.M. // Methods Mol. Biol. Clifton NJ. 2010. V. 674. P. 33–42.
29. Salva M.Z., Himeda C.L., Tai P.W., Nishiuchi E., Gregorevic P., Allen J.M., Finn E.E., Nguyen Q.G., Blankinship M.J., Meuse L., et al. // Mol. Ther. J. Am. Soc. Gene Ther. 2007. V. 15. № 2. P. 320–329.
30. Piekarowicz K., Bertrand A.T., Azibani F., Beuvin M., Julien L., Machowska M., Bonne G., Rzepecki R. // Mol. Ther. Methods Clin. Dev. 2019. V. 15. P. 157–169.
31. Himeda C.L., Chen X., Hauschka S.D. // Methods Mol. Biol. Clifton NJ. 2011. V. 709. P. 3–19.
32. Somlyo A.V., Siegman M.J. In: Muscle. // Elsevier. 2012 (Accessed April 23, 2020). 1117–1132.
33. Gunning P., Ponte P., Blau H., Kedes L. // Mol. Cell. Biol. 1983. V. 3. № 11. P. 1985–1995.
34. Muscat G.E., Kedes L. // Mol. Cell. Biol. 1987. V. 7. № 11. P. 4089–4099.
35. Brennan K.J., Hardeman E.C. // J. Biol. Chem. 1993. V. 268. № 1. P. 719–725.
36. Crawford G.E., Faulkner J.A., Crosbie R.H., Campbell K.P., Froehner S.C., Chamberlain J.S. // J. Cell Biol. 2000. V. 150. № 6. P. 1399–1410.
37. Glover L.E., Newton K., Krishnan G., Bronson R., Boyle A., Krivickas L.S., Brown R.H. // Ann. Neurol. 2010. V. 67. № 3. P. 384–393.
38. Miniou P., Tiziano D., Frugier T., Roblot N., Le Meur M., Melki J. // Nucleic Acids Res. 1999. V. 27. № 19. P. e27.
39. Kimura E., Li S., Gregorevic P., Fall B.M., Chamberlain J.S. // Mol. Ther. J. Am. Soc. Gene Ther. 2010. V. 18. № 1. P. 206–213.
40. Strimpakos G., Corbi N., Pisani C., Di Certo M.G., Onori A., Luvisetto S., Severini C., Gabanella F., Monaco L., Mattei E., et al. // J. Cell. Physiol. 2014. V. 229. № 9. P. 1283–1291.
41. Hagstrom J.N., Couto L.B., Scallan C., Burton M., McCleland M.L., Fields P.A., Arruda V.R., Herzog R.W., High K.A. // Blood. 2000. V. 95. № 8. P. 2536–2542.
42. Petropoulos C.J., Rosenberg M.P., Jenkins N.A., Copeland N.G., Hughes S.H. // Mol. Cell. Biol. 1989. V. 9. № 9. P. 3785–3792.
43. Asante E.A., Boswell J.M., Burt D.W., Bulfield G. // Transgenic Res. 1994. V. 3. № 1. P. 59–66.
44. Hu Q., Tong H., Zhao D., Cao Y., Zhang W., Chang S., Yang Y., Yan Y. // Cell. Mol. Biol. Lett. 2015. V. 20. № 1. P. 160–176.
45. Welle S., Bhatt K., Thornton C.A. // Genome Res. 1999. V. 9. № 5. P. 506–513.
46. Jaynes J.B., Chamberlain J.S., Buskin J.N., Johnson J.E., Hauschka S.D. // Mol. Cell. Biol. 1986. V. 6. № 8. P. 2855–2864.
47. Shield M.A., Haugen H.S., Clegg C.H., Hauschka S.D. // Mol. Cell. Biol. 1996. V. 16. № 9. P. 5058–5068.
48. Amacher S.L., Buskin J.N., Hauschka S.D. // Mol. Cell. Biol. 1993. V. 13. № 5. P. 2753–2764.
49. Donoviel D.B., Shield M.A., Buskin J.N., Haugen H.S., Clegg C.H., Hauschka S.D. // Mol. Cell. Biol. 1996. V. 16. № 4. P. 1649–1658.
50. Hauser M.A., Robinson A., Hartigan-O'Connor D., Williams-Gregory D.A., Buskin J.N., Apone S., Kirk C.J., Hardy S., Hauschka S.D., Chamberlain J.S. // Mol. Ther. J. Am. Soc. Gene Ther. 2000. V. 2. № 1. P. 16–25.
51. Molkentin J.D., Jobe S.M., Markham B.E. // J. Mol. Cell. Cardiol. 1996. V. 28. № 6. P. 1211–1225.
52. Wang B., Li J., Fu F.H., Chen C., Zhu X., Zhou L., Jiang X., Xiao X. // Gene Ther. 2008. V. 15. № 22. P. 1489–1499.
53. Martari M., Sagazio A., Mohamadi A., Nguyen Q., Hauschka S.D., Kim E., Salvatori R. // Hum. Gene Ther. 2009. V. 20. № 7. P. 759–766.
54. Gonçalves M.A., Janssen J.M., Nguyen Q.G., Athanasopoulos T., Hauschka S.D., Dickson G., de Vries A.A. // Mol. Ther. 2011. V. 19. № 7. P. 1331–1341.
55. Pozsgai E.R., Griffin D.A., Heller K.N., Mendell J.R., Rodino-Klapac L.R. // Mol. Ther. J. Am. Soc. Gene Ther. 2017. V. 25. № 4. P. 855–869.
56. Potter R.A., Griffin D.A., Sondergaard P.C., Johnson R.W., Pozsgai E.R., Heller K.N., Peterson E.L., Lehtimäki K.K., Windish H.P., Mittal P.J., et al. // Hum. Gene Ther. 2018. V. 29. № 7. P. 749–762.

57. Duan D. // Mol. Ther. J. Am. Soc. Gene Ther. 2018. V. 26. № 10. P. 2337–2356.
58. Hakim C.H., Wasala N.B., Pan X., Kodippili K., Yue Y., Zhang K., Yao G., Haffner B., Duan S.X., Ramos J., et al. // Mol. Ther. Methods Clin. Dev. 2017. V. 6. P. 216–230.
59. Buckingham M. // Trends Genet. TIG. 1992. V. 8. № 4. P. 144–148.
60. Raguz S., Hobbs C., Yagüe E., Ioannou P.A., Walsh F.S., Antoniou M. // Dev. Biol. 1998. V. 201. № 1. P. 26–42.
61. Li Z.L., Paulin D. // J. Biol. Chem. 1991. V. 266. № 10. P. 6562–6570.
62. Li Z., Colucci E., Babinet C., Paulin D. // Neuromuscul. Disord. NMD. 1993. V. 3. № 5–6. P. 423–427.
63. Zhang G., Ludtke J.J., Thioudellet C., Kleinpeter P., Antoniou M., Herweijer H., Braun S., Wolff J.A. // Hum. Gene Ther. 2004. V. 15. № 8. P. 770–782.
64. Pacak C.A., Sakai Y., Thattaliyath B.D., Mah C.S., Byrne B.J. // Genet. Vaccines Ther. 2008. V. 6. P. 13.
65. Talbot G.E., Waddington S.N., Bales O., Tchen R.C., Antoniou M.N. // Mol. Ther. J. Am. Soc. Gene Ther. 2010. V. 18. № 3. P. 601–608.
66. Salabarria S.M., Nair J., Clement N., Smith B.K., Raben N., Fuller D.D., Byrne B.J., Corti M. // J. Neuromuscul. Dis. V. 7. № 1. P. 15–31.
67. Lee C.-J., Fan X., Guo X., Medin J.A. // J. Cardiol. 2011. V. 57. № 1. P. 115–122.
68. Müller O.J., Leuchs B., Pleger S.T., Grimm D., Franz W.-M., Katus H.A., Kleinschmidt J.A. // Cardiovasc. Res. 2006. V. 70. № 1. P. 70–78.
69. Schinkel S., Bauer R., Bekeredjian R., Stucka R., Rutschow D., Lochmüller H., Kleinschmidt J.A., Katus H.A., Müller O.J. // Hum. Gene Ther. 2012. V. 23. № 6. P. 566–575.
70. Robert M.-A., Lin Y., Bendjelloul M., Zeng Y., Dessolin S., Broussau S., Larochelle N., Nalbantoglu J., Massie B., Gilbert R. // J. Gene Med. 2012. V. 14. .
71. Li X., Eastman E.M., Schwartz R.J., Draghia-Akli R. // Nat. Biotechnol. 1999. V. 17. № 3. P. 241–245.
72. Rudeck S., Etard C., Khan M.M., Rottbauer W., Rudolf R., Strähle U., Just S. // Genes. N. Y. N 2000. 2016. V. 54. № 8. P. 431–438.
73. Aysha J., Noman M., Wang F., Liu W., Zhou Y., Li H., Li X. // Mol. Biotechnol. 2018. V. 60. № 8. P. 608–620.
74. Domenger C., Grimm D. // Hum. Mol. Genet. 2019. V. 28. № R1. P. R3–R14.
75. Boeva V. // Front. Genet. 2016. V. 7. P. 24.
76. Le Guiner C., Servais L., Montus M., Larcher T., Fraysse B., Moullec S., Allais M., François V., Dutilleul M., Malerba A., et al. // Nat. Commun. 2017. V. 8. P. 16105.
77. Colella P., Sellier P., Costa Verdera H., Puzzo F., van Wittenberghe L., Guerchet N., Daniele N., Gjata B., Marmier S., Charles S., et al. // Mol. Ther. Methods Clin. Dev. 2019. V. 12. P. 85–101.
78. Escobar H., Schöwel V., Spuler S., Marg A., Izsvák Z. // Mol. Ther. Nucleic Acids. 2016. V. 5. P. e277.
79. Khan A.S., Draghia-Akli R., Shypailo R.J., Ellis K.I., Mersmann H., Fiorotto M.L. // Mol. Ther. J. Am. Soc. Gene Ther. 2010. V. 18. № 2. P. 327–333.
80. Liu Y., He Y., Wang Y., Liu M., Jiang M., Gao R., Wang G. // Plasmid. 2019. V. 106. P. 102441.
81. Li H., Capetanaki Y. // EMBO J. 1994. V. 13. № 15. P. 3580–3589.
82. Sarcar S., Tulalamba W., Rincon M.Y., Tipanee J., Pham H.Q., Evens H., Boon D., Samara-Kuko E., Keyaerts M., Loperfido M., et al. // Nat. Commun. 2019. V. 10.
83. Crane M.M., Sands B., Battaglia C., Johnson B., Yun S., Kaeberlein M., Brent R., Mendenhall A. // Sci. Rep. 2019. V. 9.
84. Wu Z., Sun J., Zhang T., Yin C., Yin F., Van Dyke T., Samulski R.J., Monahan P.E. // Mol. Ther. J. Am. Soc. Gene Ther. 2008. V. 16. № 2. P. 280–289.
85. Wang L., Wang Z., Zhang F., Zhu R., Bi J., Wu J., Zhang H., Wu H., Kong W., Yu B., et al. // Int. J. Med. Sci. 2016. V. 13. № 4. P. 286–291.
86. Powell S.K., Rivera-Soto R., Gray S.J. // Discov. Med. 2015. V. 19. № 102. P. 49–57.
87. Geisler A., Fechner H. // World J. Exp. Med. 2016. V. 6. № 2. P. 37–54.
88. Büning H., Srivastava A. // Mol. Ther. Methods Clin. Dev. 2019. V. 12. P. 248–265.
89. Bello A., Chand A., Aviles J., Soule G., Auricchio A., Kobinger G.P. // Sci. Rep. 2014. V. 4. № 1. P. 1–11.
90. Fumoto S., Kawakami S., Hashida M., Nishida K. // Nov. Gene Ther. Approaches. 2013.

Diversity and Functions of Type II Topoisomerases

D. A. Sutormin[1,2*], A. K. Galivondzhyan[3,4], A. V. Polkhovskiy[1,2], S. O. Kamalyan[1,2], K. V. Severinov[2,5,6], S. A. Dubiley[1,2]

[1]Institute of Gene Biology RAS, Moscow, 119334 Russia
[2]Centre for Life Sciences, Skolkovo Institute of Science and Technology, Moscow, 121205 Russia
[3]Lomonosov Moscow State University, Moscow, 119991 Russia
[4]Institute of Molecular Genetics RAS, Moscow, 123182 Russia
[5]Centre for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology RAS, Moscow, 119334 Russia
[6]Waksman Institute for Microbiology, Piscataway, New Jersey, 08854 USA
*E-mail: sutormin94@gmail.com

Received July 01, 2020; in final form, October 09, 2020
DOI: 10.32607/actanaturae.11058

ABSTRACT **The DNA double helix provides a simple and elegant way to store and copy genetic information. However, the processes requiring the DNA helix strands separation, such as transcription and replication, induce a topological side-effect – supercoiling of the molecule. Topoisomerases comprise a specific group of enzymes that disentangle the topological challenges associated with DNA supercoiling. They relax DNA supercoils and resolve catenanes and knots. Here, we review the catalytic cycles, evolution, diversity, and functional roles of type II topoisomerases in organisms from all domains of life, as well as viruses and other mobile genetic elements.**

KEYWORDS **topoisomerases, supercoiling, decatenation, transcription, replication, DNA segregation, spatial chromosome organization.**

ABBREVIATIONS **LUCA – last universal common ancestor; CTD – C-terminal domain; TAD – topologically associating domain; kb – kilobase.**

DNA TOPOLOGY

The topological state of DNA and the level of its supercoiling are described using the linking number concept (Lk) [1]. If one of the strands of a covalently closed circular DNA molecule is thought to be the edge of an imaginary surface, then the linking number of DNA strands is the number of intersections of this surface with the second DNA strand, with allowance for the sign of this intersection (*Fig. 1A*). Lk does not depend on molecule deformations and can only be altered through cleavage, passage, and religation of DNA strands (*Fig. 1A*) [2]. For a relaxed DNA molecule, the theoretical linking number (Lk^0) can be calculated as a ratio between the DNA length in base pairs (N) and the period of DNA (h = 10.5 bp/turn for the canonical B-form of DNA) (1). Lk of DNA molecules isolated from living organisms differs from Lk^0: it can either exceed Lk^0 ($\Delta Lk > 0$, a positively supercoiled molecule) or be less than Lk^0 ($\Delta Lk < 0$, a negatively supercoiled molecules) (2). Lk is the sum of two geometrical parameters of the double helix, called the twist (Tw) and the writhe (Wr) (3). The twist is defined as the number of times DNA chains turn around each other along the double helix axis, while the writhe is a measure of the supercoiling of the DNA axis [3]. When Lk is different from Lk^0, supercoiling is partitioned between the twist and writhe (4), which can interconvert to each other. For example, according to the electron microscopy of plasmids, the writhe and twist account for 75% and 25% of DNA supercoiling, respectively [3]. In nature, supercoiled DNA in the form of writhe stably exists in two forms: plectoneme (a higher order double helix) and a solenoid (a higher order single helix, which is typical of DNA wrapped around a protein) (*Fig. 1B*) [3]. A more detailed and comprehensive discussion of DNA topology may be found, for example, in the book *DNA Topology* by Bates & Maxwell, 2005 [3].

STRUCTURE, EVOLUTION, AND CATALYTIC MECHANISM OF TYPE II TOPOISOMERASES

Special enzymes, topoisomerases, regulate the level of DNA supercoiling and resolve knots and catenanes [4, 5]. According to their structure, homology, and catalytic mechanism, topoisomerases are usually divided into type I and type II [4]. Type I topoisomerases introduce a single-strand DNA break (nick) and alter the supercoil-

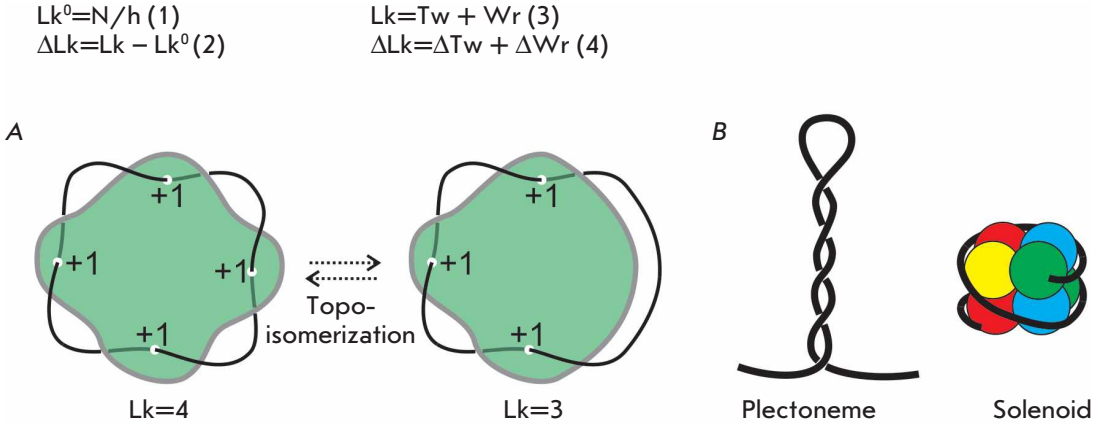


Fig. 1. DNA topology. (*A*) Linking number of a circular DNA molecule and changes in the linking number resulting from strand cleavage and transfer. (*B*) Spatial structures, plectoneme and solenoid, arising from DNA supercoiling

ing state of a molecule either by rotating the DNA duplex around the intact second strand (class IB, change Lk of the molecule by an arbitrary integer number per catalytic event) or by passing the intact strand through the nick (class IA, change Lk by ±1 per catalytic event). Type II topoisomerases cleave both strands in a DNA fragment, termed the G-segment, and pass the second duplex, the T-segment, through this break, hydrolyzing two ATP molecules (*Fig. 3*) [6–8]. This process is topologically equivalent to a change in Lk by ±2 [9]. DNA supercoiling is altered if G- and T-segments belong to the same molecule, but if they come from different molecules, action of the toposiomerase results in catenation or decatenation of DNAs (*Fig. 3C*). Below, we will analyze the diversity, mechanisms, and physiological role of type II enzymes.

Type II topoisomerases are found in organisms of all domains of life and are encoded in most, except for a few extremely reduced ones, sequenced genomes of cellular organisms [10, 11]. In all studied cases, type II topoisomerases have been shown to be necessary for transcription, replication, and segregation of chromosomes during cell division.

On the basis of their structure and catalytic cycle features, type II topoisomerases are subdivided into two classes: IIA and IIB (*Fig. 2, 3*) [4]. Topoisomerases can be either heterotetramers consisting of two B and two A subunits or homodimers in which the B and A subunits are combined into a single polypeptide [10]. The topoisomerase subunits have dimerization interfaces, referred to as gates. The conserved ATP-hydrolysis GHKL (**G**yrase, **H**sp90, **H**istidine Kinase, **M**utL) domain [12] forms the **N-gate**, and the Toprim and WHD (**Topo**isomerase/Primase and **W**inged-**h**elix **d**omain) domains form the **DNA-gate** [13]. The G-segment of DNA binds to the DNA-gate region of the enzyme and is cleaved by active site tyrosyl residues of the WHD domains [14]. The third dimerization interface (**C-gate**), formed by the **c**oiled-**c**oil (CC) domain, is present only in type IIA enzymes (*Fig. 2*) [15]. The **C**-**t**erminal **d**omains (CTD) are located either at the C-termini of A-subunits or at the end of fused polypeptides. CTD determines the specificity of topoisomerases IIA to DNA structures (supercoils or crossovers), interacts with other proteins, and, in eukaryotes, is subject to post-translational modifications regulating the activity of the enzyme [16–18].

At the first stage of the catalytic cycle, topoisomerase IIA is believed to bind the G-segment of DNA in the DNA-gate region [19]. The binding causes DNA bending, which is probably the basis of the topological scanning of DNA by the enzyme: topoisomerase preferentially binds to supercoiled regions of the molecules that are either already bent or can be easily bent due to energy of supercoiling [20–22]. Next, the T-segment of DNA is trapped between the GHKL domains and the DNA-gate. Binding of two ATP molecules to ATPase centers leads to dimerization of the GHKL domains, closure of the N-gate, and secure capture of the T-segment [23]. Hydrolysis of the first ATP molecule to ADP triggers cleavage of the DNA G-segment by the catalytic site tyrosyl residues of the WHDs and opens the DNA-gate, which results in the T-segment passage through the break to the protein cavity at the C-gate [7, 13, 24, 25]. To stabilize the double-stranded break, the hydroxyl groups of the tyrosyl residues remain linked to the DNA 5'-ends by phosphodiester bonds. Opening of the C-gate, which releases the T-segment from the enzymatic complex, follows closure of the DNA-gate and ligation of the G-segment due to hydrolysis of the second ATP molecule [26]. The release of ADP molecules, which have low affinity for active centers, leads to the opening of the N-gate and transition of the enzyme to its original state *(Fig. 3A)* [23].

Binding of ATP molecules is believed to be necessary for the unidirectional passage of the T-segment, since this segment is incapable of leaving the enzyme through the N-gate until both ATP molecules are hydrolyzed [24]. It should be noted that the role of ATP hydrolysis in segment passage has not been fully

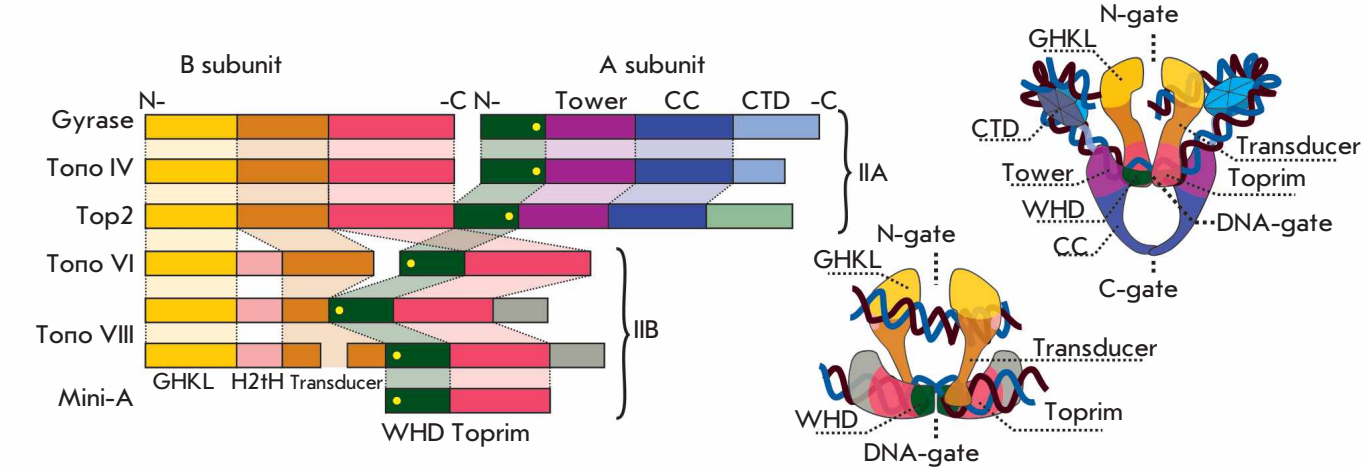


Fig. 2. Type II topoisomerase structure. *Left* – variants of the enzyme domain architecture. Homologous domains are shown in the same colors. In the WHD, the catalytic tyrosine residue responsible for DNA cleavage is depicted by a yellow circle. *Right* – domain organization of type IIA (DNA gyrase) and IIB (Topo VI) topoisomerases

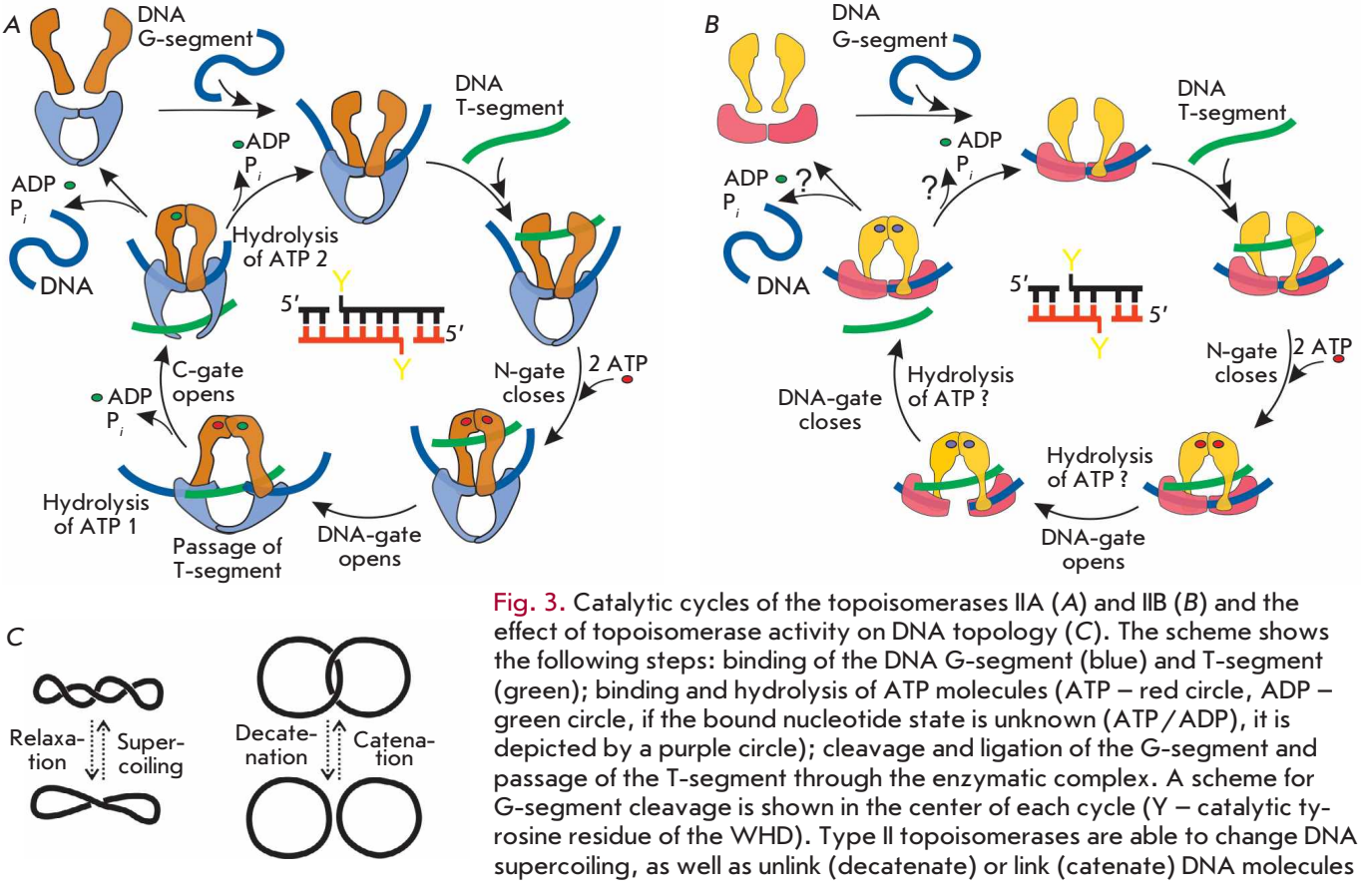


Fig. 3. Catalytic cycles of the topoisomerases IIA (*A*) and IIB (*B*) and the effect of topoisomerase activity on DNA topology (*C*). The scheme shows the following steps: binding of the DNA G-segment (blue) and T-segment (green); binding and hydrolysis of ATP molecules (ATP – red circle, ADP – green circle, if the bound nucleotide state is unknown (ATP/ADP), it is depicted by a purple circle); cleavage and ligation of the G-segment and passage of the T-segment through the enzymatic complex. A scheme for G-segment cleavage is shown in the center of each cycle (Y – catalytic tyrosine residue of the WHD). Type II topoisomerases are able to change DNA supercoiling, as well as unlink (decatenate) or link (catenate) DNA molecules

elucidated. According to one of the existing models, sequential hydrolysis of two ATP molecules promotes the T-segment passage by induced conformational rearrangements [27, 28]. According to another model, the hydrolysis is required only for "restarting" the enzyme and trapping a new T-segment [29]. For example, in the presence of ADPNP, a non-hydrolyzable ATP analogue, topoisomerase is able to perform one act of

T-segment passage, and then the enzyme remains in an inactive state with a closed N-gate [30]. According to recent single-molecule studies of DNA and DNA gyrase using magnetic tweezers, ATP hydrolysis is important both for accelerating T-segment passage and for "restarting" the enzyme [7]. An alternative explanation considers ATP binding and GHKL domain dimerization as a safeguard that is necessary to stabilize the two halves of the enzymatic complex and to prevent the formation of double-strand breaks during T-segment transfer due to accidental dissociation of the two enzyme halves [8].

The catalytic mechanism of type IIB topoisomerases is considered to be similar to that of type IIA topoisomerases (*Fig. 3B*) [31–33]. However, due to the absence of a C-gate, the T-segment immediately leaves the enzymatic complex after passing through the DNA-gate and the break in the G-segment [31]. In type IIB topoisomerases, the tyrosyl residues of WHDs are located on different secondary structure elements compared to the homologous domains of type IIA enzymes. When cleaving the G-segment of DNA, they generate two-nucleotide 5'-overhanging ends instead of the four-nucleotide overhangs characteristic of type IIA topoisomerases [34, 35]. G-segment cleavage was shown to depend on ATP binding for IIB enzymes. This is considered necessary for the stabilization of the complex and that of the temporary double-stranded break [8, 32].

The evolutionary relationships within type IIA and IIB topoisomerase groups and between these groups remain the subject of debate. Only a few evolutionary events can be reliably traced; for example, the duplication of a type IIA topoisomerase gene in the ancestor of bacteria, which led to the emergence of two enzymes with specific functions: DNA gyrase and Topo IV. Similarly, a duplication in the ancestor of vertebrates resulted in the emergence of Top2α and Top2β. Horizontal transfer of gyrase genes from different bacterial groups to Euryarchaeota and reverse transfer of Topo VI genes have also been described. Bacterial gyrase found in Archaeplastida is likely to be inherited from chloroplasts during establishing of primary endosymbiosis [10]. For more ancient events of topoisomerase evolution, there is no consensus.

BACTERIAL TOPOISOMERASES

Free-living fast-growing bacteria, such as *Escherichia coli*, *Caulobacter crescentus*, and *Bacillus subtilis*, usually possess a wide spectrum of topoisomerases. This includes type I topoisomerases I and III, as well as type II, class IIA DNA gyrase and topoisomerase IV [4, 36–38]. Slow-growing bacteria (e.g., *Mycobacterium tuberculosis*) or symbiotic/parasitic bacteria with reduced genomes (e.g., *Helicobacter pylori*), in contrast, often have the minimum essential set of one type I (topoisomerase I) and one type II (DNA gyrase) enzymes [39, 40]. The genomes of several endosymbiotic bacteria, for example *Hodgkinia cicadicola* and *Tremblaya princeps*, lack topoisomerase II genes or, like *Carsonella rudii*, encode only one subunit [41–43]. These organisms have extremely reduced (139–160 kb) genomes.

DNA gyrase and topoisomerase IV are the targets for many antibiotics that, according to their mechanism of action, may be divided into two groups: poisons and catalytic inhibitors. Poisons stabilize an intermediate covalent complex of topoisomerase with the DNA G-segment. Accidental dissociation of enzyme subunits from such a complex (for example induced by the collision with the replisome or RNA polymerase) causes double-stranded DNA breaks and ultimately leads to cell death. Catalytic inhibitors do not cause DNA breaks, but they inhibit enzymatic activity, for example, by binding to the ATPase center of the GHKL domain and competing with ATP [44, 45].

Quinolone and fluoroquinolone drugs (ciprofloxacin, levofloxacin, etc.), which are often used in clinical practice, are topoisomerase poisons [44, 46]. Structural studies have shown that movement of divalent metal ions (most often magnesium) in the topoisomerase catalytic center is necessary for DNA cleavage and ligation. Gyrase poisons stabilize a metal ion in the position that promotes DNA cleavage, but not the sealing of the break [47, 48]. The latter fact explains the effects of the most prevalent gyrase mutations leading to antibiotic resistance. The conserved serine and glutamine residues of the WHD were found to coordinate water molecules and magnesium ions, which are necessary for the binding of fluoroquinolones [47]. Replacing at least one of these residues with a non-polar moiety leads to poison resistance [49].

Classical catalytic inhibitors are aminocoumarin compounds (novobiocin and coumermycin A1) that compete with ATP for the interaction with the ATPase center [44, 50]. Inhibition of gyrase activity leads to inhibition of replication and transcription and cell division arrest. Due to the low solubility of aminocoumarins and their toxicity to humans, aminocoumarin drugs are not currently used in clinical practice, but they found application in veterinary medicine [45].

The spread of antibiotic resistance necessitates a search for new antibacterial drugs; several new classes of topoisomerase inhibitors are currently in clinical trials [45, 51, 52].

DNA gyrase

Bacterial DNA gyrases are conserved enzymes (*Fig. 4A*) sharing a unique ability to induce negative

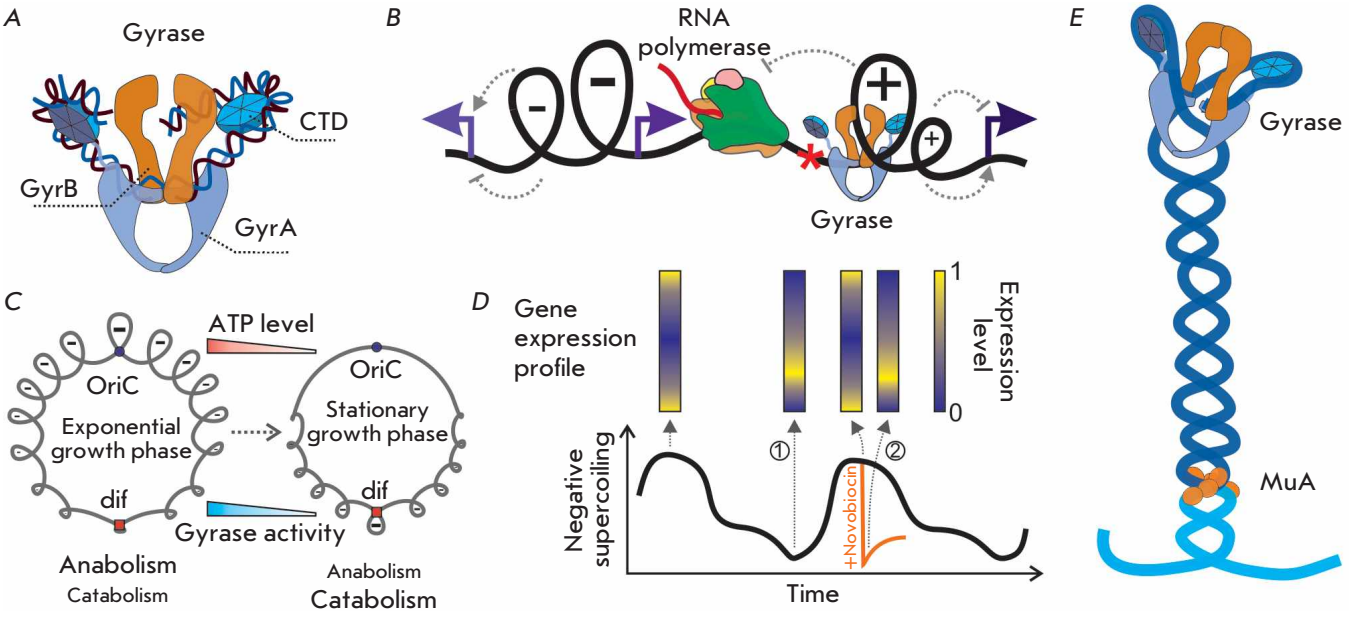


Fig. 4. DNA gyrase and its function. (*A*) Structure of a DNA gyrase complex with DNA. (*B*) a twin-domain model illustrating positive supercoiling upstream of the elongating RNA-polymerase and negative supercoiling downstream [62]. Co-transcriptional positive, and negative, supercoiling moves along the DNA molecule and influences the initiation of transcription from adjacent promoters (indicated by arrows). Depending on the promoter, the effect can be either activating or inhibiting. DNA gyrase promotes transcription elongation through relaxation of positive supercoiling ahead of RNA polymerase. (*C*) Changes in genome supercoiling during *E. coli* culture transition from the exponential to stationary growth phase promote switching of the cell from a mainly anabolic to catabolic physiological state [63]. OriC – origin of replication, dif – site recognized by XerC/XerD recombinases. (*D*) Circadian oscillations of the *S. elongatus* genome supercoiling level (at the bottom) correlate with changes in the gene transcriptional profile (at the top). A sharp decrease in the genome supercoiling level (indicated by the orange arrow) in the presence of the DNA gyrase inhibitor novobiocin causes rapid change in the transcriptional profile (2), making it similar to the profile of bacteria in the physiologically relaxed genome state (1) [64]. (*E*) DNA gyrase is essential for the spatial organization of the Mu prophage and its transposition. The prophage DNA is shown in dark blue, and bacterial genome DNA is in blue

supercoiling using the energy of ATP hydrolysis, which was demonstrated in *in vitro* experiments for enzymes from *E. coli*, *B. subtilis*, *C. crescentus*, *M. tuberculosis*, and many other bacteria. In addition, DNA gyrases effectively relax positive supercoils and are capable of decatenating circular DNA molecules [39, 53–56]. The *gyrA* and *gyrB* genes encoding the enzyme subunits are essential, and inhibitors that reduce gyrase activity significantly decrease cell viability [57–60]. Gyrase inhibition induces a similar phenotype in different bacteria: elongated cells incapable of dividing [60, 61].

Gyrase maintains negative supercoiling of the genome, facilitating the initiation of transcription and replication. It also relaxes positive supercoils in front of elongating polymerases. Early ChIP-chip (immunoprecipitation of protein-bound DNA and its subsequent analysis on a chip to determine protein binding sites) experiments with *E. coli* revealed a positive correlation between gyrase binding and a gene's transcription level [65]. Later, using the Topo-Seq method that enables highly accurate mapping of topoisomerase activity sites, catalytically active DNA gyrase from *E. coli* was directly shown to be located at the ends of active genes and in the regions downstream of transcription terminators [66]. Similarly, the results of ChIP-Seq (immunoprecipitation of protein-bound DNA and its subsequent sequencing to determine protein binding sites) experiments with *M. tuberculosis* gyrase indicate preferential binding of the enzyme to transcriptionally active regions [67]. In *C. crescentus*, suppression of the *gapR* gene expression inhibits initiation and elongation of replication and increases the sensitivity of cells to gyrase inhibitors. *In vitro* experiments have shown that the GapR protein preferentially binds to positively supercoiled DNA and interacts with the gyrase, increasing its ability to relax positive supercoils. Probably, GapR recruits the gyrase to the positive supercoils formed in front of the moving replication complex, fa-

cilitating their relaxation and thus stimulating replication [55]. Single-molecule experiments have shown that in the absence of gyrase, transcription on topologically constrained DNA molecules quickly slows down and eventually stops due to the accumulation of positive supercoiling (*Fig. 4B*). The binding of gyrase to such molecules results in rapid restoration of the normal rate of transcription (transcriptional burst) [68].

In addition to its ability to relax positive supercoiling in front of elongating RNA polymerase, by introduction of negative supercoiling the gyrase can both activate and suppress transcription initiation [69]. Up to half of *E. coli* genes were found to respond to genome relaxation by changing their transcription level [70, 71]. Ontological analysis of *E. coli* genes sensitive to supercoiling revealed that the products of genes responding to relaxation of negative supercoils by increasing their transcription level are preferentially involved in catabolic reactions (for example, Krebs cycle enzymes). These genes are located closer to the terminus of replication. In contrast, genes that require negative genome supercoiling for initiation of their transcription are predominantly associated with anabolic processes (synthesis of amino acids and nucleotides) and are located closer to the region of replication origin [71, 72]. According to one model, during active growth of a *E. coli* culture, activity of DNA gyrase generates a negative supercoiling gradient in the genome, with the maximum and minimum levels being in the replication origin and the terminus regions, respectively. This leads to a predominant expression of the genes involved in the anabolic process, promoting cell growth and division. Depletion of nutrients in the stationary phase decreases the ATP concentration, which reduces DNA gyrase activity. This decreases the genome supercoiling level and, in combination with other factors, inverts the gradient of chromosome supercoiling, resulting in a predominant expression of the genes involved in catabolic processes [63]. It was hypothesized that *E. coli* uses supercoiling to globally modulate gene transcription upon starvation [72–74] (*Fig. 4C*).

Promoters of the *E. coli gyrA*, *gyrB*, and *topA* genes that encode gyrase and topoisomerase I subunits are highly sensitive to supercoiling. They contain supercoiling sensors: the *gyrA* and *gyrB* transcription is activated upon genome relaxation, while *topA* is better transcribed upon enhancement of negative supercoiling [75, 76]. This enables the mutually regulated synthesis of two topoisomerases with opposite activities, which provides a homeostat for the genome-wide supercoiling level [77, 78]. Similar mechanisms are operational in *S. coelicolor* and *C. crescentus* [58, 79].

The supercoiling level in *Salmonella typhimurium* is believed to regulate the transition from anaerobic metabolism to aerobic respiration [80]. In *H. pylori*, negative supercoiling is an important regulator of flagellar synthesis [81]. Circadian oscillations of DNA supercoiling in the cyanobacterium *Synechococcus elongatus* correlate with specific changes in gene transcription and relaxation of negative supercoiling by the addition of the DNA gyrase inhibitor novobiocin, leading to a rapid change in the gene transcription pattern, mimicking the changes observed during the circadian cycle (*Fig. 4D*) [64]. Overall, these data allow one to consider supercoiling as a global transcription factor and show that the structure of regulatory regions has evolved to allow specific responses to this factor [63, 69, 72].

A number of studies have indicated that gyrase and gyrase-induced negative supercoiling are involved in the spatial organization of bacterial genomes. For example, *in vivo* fluoroquinolone induces cleavage of *E. coli* genomic DNA by the gyrase into 50- to 100-kb fragments, which roughly corresponds to the length of supercoiled chromosome domains [82–84]. Activity of DNA gyrase at a high-affinity site located at the center of the bacteriophage Mu prophage was shown to cause a local increase in negative supercoiling, leading to plectonemic compaction of the chromosome region with the prophage. This brings prophage termini into proximity with each other and promotes their recombination by the MuA transposase [85, 86] (*Fig. 4E*). Similarly, excessive negative supercoiling accumulated in *E. coli* cells with a mutation in topoisomerase I is believed to lead to chromosome compaction [87]. As shown by Hi-C experiments (a method for determining the chromosome conformation) in *C. crescentus*, gyrase inhibition by novobiocin, on the contrary, makes the spatial structure of the chromosome more diffuse [88]. It should be noted that for the *E. coli* genome no significant associations between gyrase active sites and either the boundaries or locations of topologically associating domains (TADs) determined by Hi-C were found [66]. Further research is needed to elucidate the role of supercoiling in the regulation of the spatial organization of prokaryotic genomes.

Topoisomerase IV

In vitro experiments have demonstrated that despite their structural similarity topoisomerases IV (Topo IV) and gyrases have different spectra of activity. Topo IV is able to effectively relax positive supercoils. Negative supercoils are relaxed at a much slower rate. Unlike the gyrase, Topo IV cannot introduce excessive negative supercoiling [55, 56, 89]. At the same time, Topo IV is an efficient decatenase that separates interlinked circular DNA molecules much better than gyrase [90–94]. Accordingly, Topo IV, but not gyrase, is capable of resolving knotted DNA molecules *in vivo*

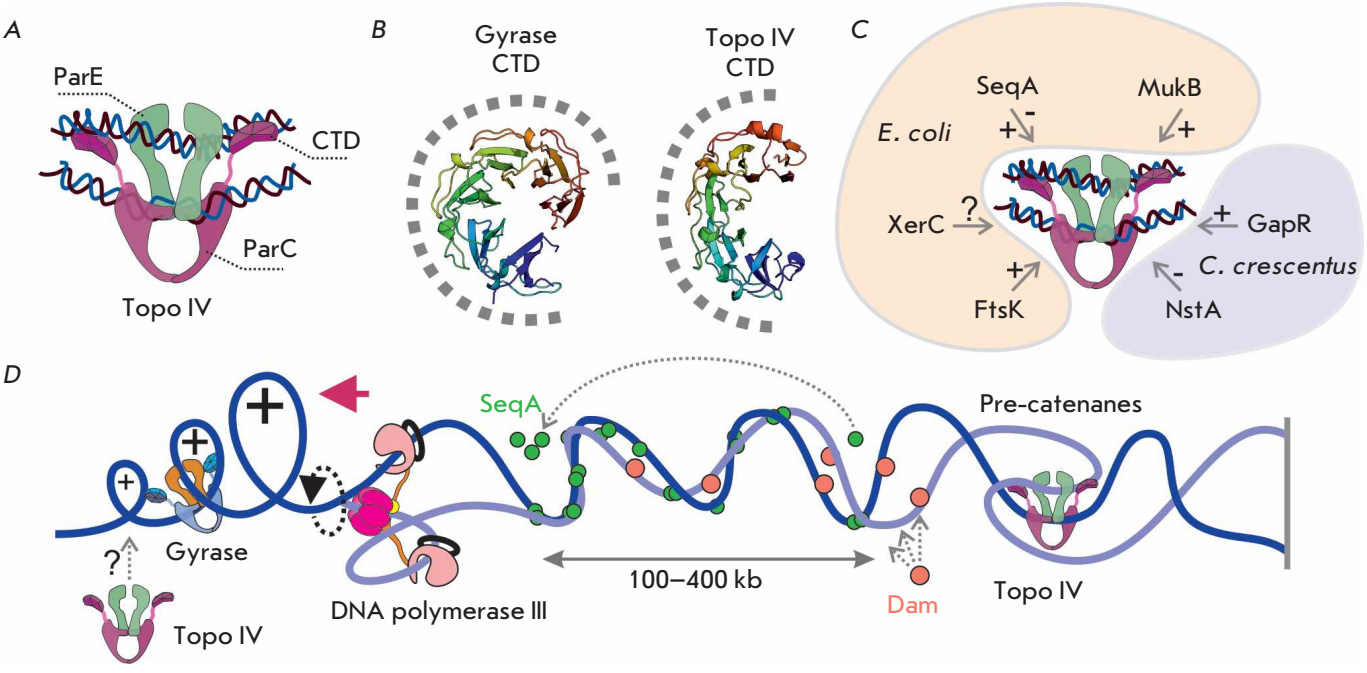


Fig. 5. Topoisomerase IV and its function. (*A*) Structure of the Topo IV complex with DNA. (*B*) Comparison of the GyrA CTD (PDB ID: **1zi0**) and Topo IV ParC CTD (PDB ID: **1zvt**) structures. A putative position of DNA is shown as a dashed line. (*C*) Proteins interacting with Topo IV. The effect of each protein on Topo IV activity is depicted as "+" (activation), "–" (inhibition), or "?" (interaction is not confirmed). (*D*) Topological effects associated with DNA replication. Positive supercoils formed in front of the moving replisome are relaxed by DNA gyrase and, presumably, Topo IV. Accumulation of DNA supercoiling leads to replisome rotation, thereby producing DNA pre-catenanes. In *E. coli*, the SeqA protein binds to the hemimethylated GATC sites of newly replicated DNA molecules. Dam methylates GATC sites and displaces SeqA; so, the SeqA concentration gradient extends 100–400 kb over the replisome and moves together with it. Topo IV cannot interact with SeqA-bound DNA regions, which explains the temporary cohesion of daughter chromosomes during replication in *E. coli*; however, when all GATC sites are methylated and SeqA is no longer associated with DNA, topoisomerase removes pre-catenanes, enabling daughter chromosome separation [110]

[95]. It is hypothesized that these differences are related to the structures of CTD domains in the GyrA subunit of gyrase and in the homologous ParC subunit of Topo IV (*Fig. 5B*). The gyrase CTD enables wrapping of DNA around the enzyme, such that DNA located *in cis* and close to the G-segment of DNA serves as a T-segment, which allows for the introduction of negative supercoils in one DNA molecule [7, 96]. The Topo IV CTD does not bend the G-segment; instead it traps as a T-segment remote DNA sites or *in trans* DNA molecules. Since the T-segment must be perpendicular to the enzyme-bound G-segment, catenanes are effectively recognized and resolved [89, 93, 97] (*Fig. 4A*, *Fig. 5A*).

Like gyrase, Topo IV is necessary for bacterial division. Mutations in the *parC* and *parE* topoisomerase subunit genes or inhibition of the enzyme activity by drugs causes the development of the so-called *par* phenotype in different bacteria. The *par* phenotype is characterized by elongated cells that are not capable of division and contain an increased amount of unsegregated DNA [36, 98–101]. However, the lack of Topo IV activity does not interfere with *E. coli* chromosome replication and its termination [99, 100]. The biochemical properties of the enzyme suggest that the main function of Topo IV in the cell is to resolve pre-catenanes during replication (intersections between sister DNA molecules arising from replisome rotation) and to separate catenanes of circular molecules upon the completion of replication [100, 102] 2] (*Fig. 5D*). According to this hypothesis, Topo IV is not essential for *Streptomyces* with linear chromosomes but is important for the maintenance of circular plasmids [38]. Yet, *E. coli* cells with artificial linear chromosomes exhibit *par* phenotype upon Topo IV inactivation. This may be an indication of the importance of the early removal of pre-catenanes and knots along the entire length of the replicating chromosome [103]. An increase in the Topo IV expression level leads to accelerated DNA segregation during the division of *E. coli* cells [100].

Decatenation in bacteria lacking Topo IV is supposed to be performed by DNA gyrase and type I topoisomerases. For example, *M. tuberculosis* gyrase is an efficient decatenase. The ChIP-Seq experiment demonstrated that the *M. tuberculosis* gyrase is significantly enriched in the chromosomal replication terminus region, which suggests that it acts as Topo IV [39, 67, 104]. However, no such enrichment was observed for the *E. coli* gyrase [66]. The involvement of *H. pylori* gyrase in chromosome segregation is indirectly confirmed by the fact that bacteria with deletion of the *xerH* gene, which encodes the recombinase involved in the resolution of chromosome dimers and, possibly, decatenation, are more sensitive to the gyrase inhibitor ciprofloxacin [99, 105].

The ability of Topo IV to relax positive supercoils [56, 89] suggests that it may cooperate with the DNA gyrase in the removal of positive supercoils formed during transcription and replication [55, 106] (*Fig. 5D*). For example, treatment of *E. coli* cells with the RNA polymerase inhibitor rifampicin was found to reduce both the gyrase and Topo IV activities, at least in some regions of the genome [83, 107]. Interestingly, an increase in the copy number of the *parC* and *parE* genes is a common suppressor mutation associated with deletion of the topoisomerase I gene in *E. coli* and *B. subtilis*. In this case, Topo IV is believed to compensate for the loss of topoisomerase I and perform its function by removing negative supercoiling [37, 98, 108].

Topo IV interacts with a number of proteins that have completely different functions and structures, but are involved in the organization and separation of replicated chromosomes. In *E. coli*, these are the SeqA protein that binds to the hemimethylated GATC sites behind the moving replisome [109, 110], the MukBEF cohesin [111, 112], the DNA translocase FtsK [113], and, probably, the XerC recombinase [107, 114] (*Fig. 5C*). *C. crescentus* Topo IV interacts with GapR and NstA. These proteins have opposite effects on Topo IV – GapR stimulates enzyme activity, while NstA suppresses it [55, 115]. *In vivo*, Topo IV and the *E. coli* cohesin complex MukBEF form clusters consisting of ~15 topoisomerase molecules and ~10 cohesin molecules [116, 117]. These clusters colocalize with replication origins, determine their position in the cell, and are necessary for segregation of the origins of daughter chromosomes during division [116, 118, 119]. *C. crescentus* Topo IV is also required for the correct movement of one of the origins to the opposite cell pole [101].

Topoisomerase NM

A unique type II topoisomerase, called TopoNM, was discovered in *M. smegmatis* [120]. It consists of two subunits (TopoN and TopoM), homologous to the ParE/GyrB and ParC/GyrA subunits of topoisomerase IV and gyrase, respectively. According to a phylogenetic analysis of amino acid sequences, TopoNM is distant from all known type IIA topoisomerases, which indicates early divergence of enzyme genes [120]. The significant divergence from other topoisomerases II and the absence of TopoNM in other, even related, bacteria may indicate the viral origin of the enzyme. TopoNM has reduced sensitivity to fluoroquinolones and coumarins. The enzyme relaxes positive and negative supercoils and decatenates circular DNA molecules, which is typical of type II topoisomerases. A unique property of TopoNM is the ability to introduce positive supercoils into relaxed plasmids [120]. Besides TopoNM, only reverse gyrase – a type I topoisomerase – is capable of introducing positive supercoils using the energy of ATP hydrolysis [121]. Neither the mechanism of positive supercoiling by TopoNM nor the functions of this enzyme are known.

An unusual system for protection against mobile genetic elements was found in *M. smegmatis*. It consists of genes encoding a cohesin-like complex that prevents effective transformation of bacteria with plasmids [122, 123]. TopoNM may be part of this defense system, in the way some bacterial topoisomerases interact with cohesins [111, 112, 124].

ARCHAEAL TOPOISOMERASES

Members of the Archaea domain usually harbour type IIB topoisomerases (Topo VI). Some archaea from the Euryarchaeota phylum have lost their Topo IV genes but independently acquired, through horizontal gene transfer, DNA gyrase genes from different bacterial groups [11]. Hyperthermophilic archaea encode reverse gyrases as an adaptation to high temperatures, since this enzyme is believed to be essential for maintaining DNA duplex stability at high temperatures and is involved in DNA repair [125–127].

Topoisomerase VI

Topoisomerase VI (Topo VI) was first found in the hyperthermophilic archaeon *Sulfolobus shibatae* [128] and, later, in most other archaea, except for some members of the Thermoplasmatales group in which it is replaced by the DNA gyrase [11]. *In vitro*, Topo VI can relax both positive and negative supercoils and exhibits decatenation activity [32, 129]. Similarity between the amino acid sequences of IIA and IIB topoisomerases is rather low. Additionally, the catalytic tyrosine residues of WHDs are located on non-homologous secondary structure elements in the two groups [32, 33, 130] (*Fig. 2*). Despite these, the catalytic mechanism of Topo VI is supposed to be similar to that of type

IIA topoisomerases, a conclusion based on biochemical and structural analyses (*Fig. 3B*).

The physiological role of Topo VI has not been established. The activity of the enzyme demonstrated *in vitro* and the fact that Topo VI can be replaced with DNA gyrase indicate that the topoisomerase may be involved in the decatenation of replicated chromosomes and in the relaxation of supercoils formed during transcription and replication [129]. The expression level of Topo VI in *S. islandicus* was found to increase 7 h after one elevates the cultivation temperature above its optimal level. Probably, Topo VI compensates for an increase in reverse gyrase activity under these conditions [131].

DNA gyrase

Gyrase genes have been found in members of several Euryarchaeota groups [11]. Like bacterial gyrase, the archaeal enzyme is sensitive to coumarins and quinolones [132–134]. *In vitro* experiments have shown that *Thermoplasma acidophilum* gyrase has a typical spectrum of activities: it relaxes positive supercoils, introduces negative supercoils, and decatenates circular DNA molecules [134]. Inhibition of gyrase activity by the addition of novobiocin to *Halobacterium halobium* cells leads to the inhibition of DNA replication and a significant decrease in the levels of transcription and translation [132]. Thus, the archaeal gyrases are believed to perform functions typical of bacterial homologues: relaxation of positive supercoils formed during transcription and replication, as well as decatenation of linked DNA molecules during cell division.

EUKARYOTIC TOPOISOMERASES

Homodimeric topoisomerase IIA (Top2) is common to all known eukaryotes. It is encoded by one *Top2* gene in most species; vertebrates, however, have two paralogous genes, *Top2α* and *Top2β* [10]. Archaeplastida and eukaryotes related to them via secondary endosymbiosis of plastids (Apicomplexa, etc.) contain DNA gyrase genes. The enzyme is of bacterial origin and is encoded by nuclear genes that had been transferred from the chloroplast genome after the establishment of endosymbiosis [11, 135]. The ubiquitous eukaryotic proteins involved in a complex required for generating DNA breaks during meiotic recombination are homologous to Topo VI from Archaea: Spo11 and Rec102/Rec6/MEI-P22 are homologues of the A and B subunits respectively [128, 136, 137]. These proteins are not considered topoisomerases, and we will not discuss them in detail. However, a full-length heterotetrameric Topo VI possessing typical enzymatic activities is found in plants, making it another distinctive feature of Archaeplastida [138].

Most agents used in cancer chemotherapy are topoisomerase poisons, with etoposide being the most common [139–145]. They induce double-strand breaks (DSBs), thus causing apoptosis [146–151]. The selectivity of these drugs is determined by the neoplastic features of tumor cells: they actively proliferate and have an increased topoisomerase expression level [152]. Severe side effects caused by DNA damage in normal cells, especially actively proliferating, remain a crucial issue in chemotherapy [153, 154]. Top2-mediated DSBs can lead to chromosomal translocations and induce secondary malignancies [155]. For example, etoposide therapy often leads to secondary leukemia [156–158]. The oncogenic effects often arise due to the inhibition of Top2β that is actively expressed in most tissues and is associated with promoter regions [159–163]. A possible solution to this problem may be searching for and using inhibitors targeting Top2α selectively.

Catalytic inhibitors of Top2 (merbarone, suramin, bis-dioxypiperazine derivative ICRF-187) have not been used broadly in clinical practice as antineoplastic drugs [164]. However, some of them are used as cardioprotectors, simultaneously with oncotherapy involving Top2 poisons [165, 166]. According to one of the existing hypotheses, the protective properties of inhibitors are associated with a decrease in the number of DNA-Top2 covalent complexes and, accordingly, DNA breaks due to inhibition of Top2 activity [167, 168].

Top2

Eukaryotic Top2 is a classic type IIA topoisomerase. It relaxes positive and negative supercoils and decatenates DNA molecules [169–172]. Top2 inactivation impairs chromatin condensation, leads to changes in chromosome morphology, chromosomal rearrangements, and abnormalities of embryogenesis and nervous system development in vertebrates [170, 173–180].

Eukaryotic Top2 primarily has a nuclear localization but is also present in the mitochondria of mammalian cells [181]. An increased expression level of the *Top2* gene (*Top2α* in vertebrates) is common to actively proliferating tissues, since the enzyme is essential to chromosome condensation and separation during mitosis [182, 183]. The level of *Top2β* gene expression is less dependent on the tissue type [184].

The Top2 CTD is the least conserved Top2 region. The CTD undergoes post-translational modifications, most prominently, phosphorylation, which changes in a cell cycle-dependent manner. The divergence between Top2α and Top2β CTDs determines the functional differences between the paralogs and their regulation [185]. By studying the properties of chimeric enzymes (Top2α with the CTD of Top2β and *vice versa*) it was demonstrated that Top2α CTD (CTDα) attracts topoi-

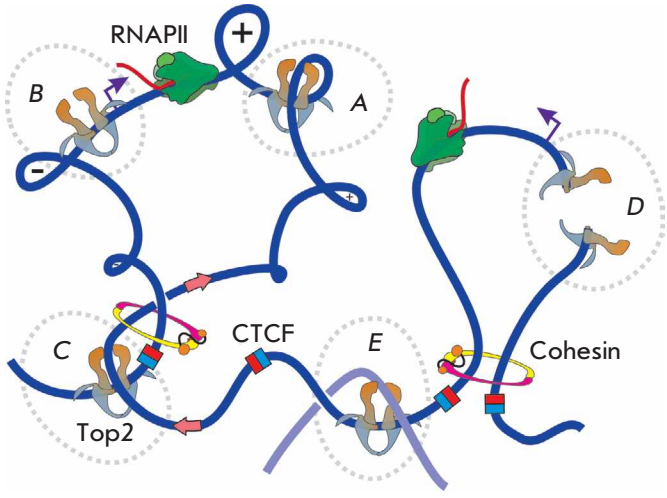


Fig. 6. Function of eukaryotic Top2. (*A*) Relaxation of supercoils during transcription. Promoters are depicted by purple arrows. (*B*) Top2 is involved in transcription initiation. (*C*) Top2, CTCF, and cohesin are colocalized at the TAD boundaries. Pink arrows display the direction of loop extrusion, mediated by CTCF and cohesin. Red-blue squares depict CTCF binding sites. Top2 facilitates cohesin-mediated DNA translocation through the relaxation of topological stress. (*D*) Top2-mediated introduction of DNA double-strand breaks in the promoter region induces transcription. (*E*) Decatenation of daughter chromosomes

somerase to chromosomes during mitosis and that a topoisomerase with CTDα is required for cell proliferation [186]. In contrast, CTDβ was shown to decrease the affinity between topoisomerase it is attached to for DNA and reduce the efficiency of catalysis [187, 188].

Top2 is required during the transcription of highly active and, especially, long genes. It relaxes positive supercoils in front of the elongating RNA polymerase (*Fig. 6A*) [189–193]. Moreover, Top2 recruits RNA polymerase II to gene promoters [194, 195] and plays an important role in the transcription initiation of some inducible yeast genes (*Fig. 6B*) [196]. Induction of genes regulated by nuclear receptors (androgens, estrogens, glucocorticoids) is associated with the promoter-mediated assembly of a complex comprising chromatin-remodeling factors (BRG1), components of the DSB repair system (PARP1, Ku70), and Top2β [197–200]. In response to hormones, Top2β, which is part of this complex, introduces a double-strand break in DNA, efficiently relaxing supercoils during transcription (*Fig. 6D*). Similar data on the activating effect of Top2β-induced breaks were obtained for several genes in NMDA-stimulated neurons [201].

Recent studies have shown that, with rare exceptions, eukaryotic genomes are organized into topologically associating domains (TADs) [202–204]. Architectural proteins, particularly CTCF and cohesin, are associated with TAD boundaries [205, 206]. Colocalization of these proteins and Top2β at the boundaries of TADs was established using the ChIP-Seq and ChIP-exo approaches (the later method has enhanced precision because of exonuclease treatment of DNA–protein complexes) (*Fig. 6C*) [207]. In addition, mapping of Top2–DNA cleavage sites stabilized by etoposide has demonstrated that they are predominantly located near the CTCF binding sites [208–211]. Presumably, TADs are composed of loops formed by extrusion due to the activity of cohesin and CTCF [212–214]. Top2 is supposed to play an important role in the functioning of chromatin loops and is necessary in order to relieve the topological stress at TAD boundaries (*Fig. 6C*). TAD compactization, according to some models, may be maintained due to the negative DNA supercoiling that can be considered a universal factor that spatially organizes both prokaryotic and eukaryotic genomes [215].

Top2 was also found to interact with the ATP-dependent chromatin-remodeling complexes [171, 216–218] that perform nucleosome assembly and movement along the DNA, and to replace canonical histones with histone variants, thus maintaining a tissue-specific chromatin structure [219–221]. The interaction between Top2 and remodeling complexes affect the catalytic properties of topoisomerase and its ability to bind DNA [171, 222], which is probably required for structural rearrangements within chromatin. The chromatin remodelers might be responsible for recruiting topoisomerases and CTCF to the TAD boundaries [223]. To date, the interplay between the chromatin architecture and Top2 activity has remained insufficiently explored and requires further investigation.

DNA gyrase

Eukaryotic gyrase, similarly to a bacterial enzyme, is capable of introducing negative supercoils *in vitro* and is sensitive to coumarins and quinolones [224–226]. Plant gyrase is able to complement a mutated enzyme in *E. coli* [225, 227, 228].

The nuclear genome of *Arabidopsis thaliana* contains one gene encoding the GYRA subunit and three paralogous genes encoding the GYRB subunit of gyrase [225]. It was shown that AtGYRA interacts with AtGYRB1 and AtGYRB2, forming complexes capable of introducing negative supercoils. In contrast, AtGYRA does not interact with AtGYRB3 [228, 229]. B-subunits contain signal peptides that are responsible for the localization of AtGYRB1 and AtGYRB2 in chloroplasts and

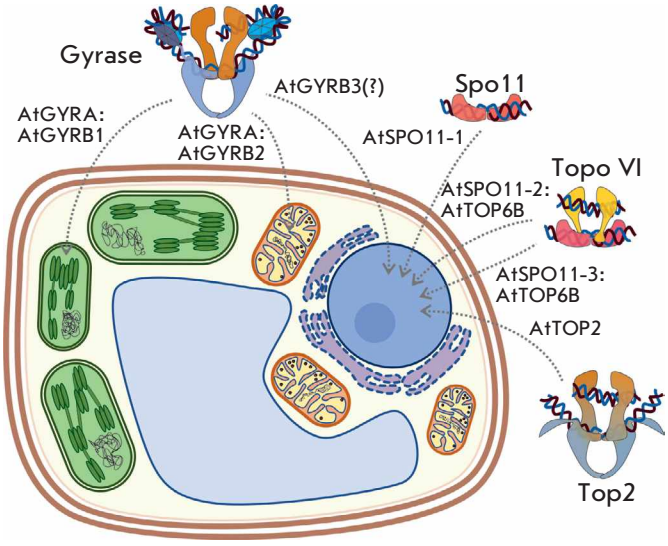


Fig. 7. Cellular localization of type II topoisomerases and homologues proteins in *A. thaliana*

mitochondria, respectively. Therefore, it is believed that the AtGYRA:AtGYRB1 complex functions in chloroplasts, and that the AtGYRA:AtGYRB2 complex functions in mitochondria. The AtGYRB3 subunit lacks a canonical signal peptide, but it is believed to localize in the nucleus [225, 228] (*Fig.* 7). *N. benthamiana* has one *GYRA* gene and two *GYRB* genes, with the GYRA and GYRB1 subunits being localized in chloroplasts and mitochondria [227]. Similar results were obtained for the GYRA subunit of *Pisum sativum* [230].

In plants, gyrase inhibition primarily affects chloroplasts and mitochondria. For example, treatment of *Chlamydomonas reinhardtii* algae with enzyme inhibitors (nalidixic acid, novobiocin) leads to transcription alteration in chloroplasts [224]. The addition of the nalidixic acid to *Nicotiana tabacum* cell cultures suppresses DNA synthesis in plastids [231]. Gyrase inhibitors reduce the number of chloroplasts and mitochondria, change the structure of chloroplasts, and, probably, disrupt their division [229]. Cultivation of plants on media supplemented with gyrase inhibitors or treatment of *A. thaliana* plants with these compounds retards their growth and induces etiolation, which ultimately leads to plant death [225, 229]. Similar results were obtained by suppression of gyrase expression in *N. benthamiana* plants by virus-induced gene silencing (VIGS), and in *A. thaliana* by RNA interference [227, 229]. Such data suggest that the enzyme in plants probably retains its characteristic role in double-membrane organelles and is necessary for the segregation of DNA, division, and transcription.

The role of the AtGYRB3 subunit remains unknown. This polypeptide lacks some amino acid motifs in its ATPase domain, which are conserved in type II topoisomerases. At the same time, it contains a histone-binding SANT domain not found in other topoisomerases [228, 232]. Analysis of *AtGYRB3* gene expression in *A. thaliana* revealed no correlation with the expression of other gyrase genes: the highest expression level of *AtGYRB3* was found in the stamens and pollen, while expression of the other subunits was most active in the seeds and shoot apical meristem. We hypothesize that the AtGYRB3 protein could be involved in meiosis, where it assists Topo VI or SPO11.

Bimolecular fluorescence complementation (BiFC) and co-immunoprecipitation experiments revealed some interaction between RNase H1 (AtRNH1C), which removes RNA from RNA–DNA heteroduplexes (R-loops) formed during transcription, and the AtGYRA subunit in *A. thaliana* chloroplasts [233]. The interaction between enzymes was thought to promote replication fork progression through R-loops that often form in actively transcribed regions of the chloroplast genome; e.g., in rRNA genes.

Topoisomerase VI

Among eukaryotes, the full-length heterotetrameric Topo VI is found only in *Archaeplastida* [138]. Similar to the case of gyrase, plants contain several paralogous genes encoding Topo VI subunits. *A. thaliana* has one B subunit gene (*AtTOP6B*) and three A subunit genes *AtSPO11-1,2,3*, while *Oryza sativa* has five paralogous genes of A-subunits and one gene of B-subunits [138, 234, 235]. Plant Topo VI subunits are localized in the nucleus, which had been predicted bioinformatically and was confirmed by microscopy [234, 236, 237].

Two-hybrid screening and co-immunoprecipitation experiments revealed that not all A-subunits form a complex with the B-subunit: *in A. thaliana*, AtTOP6B interacts with AtSPO11-2 and AtSPO11-3, but not with AtSPO11-1; in *O. sativa*, OsTOP6B interacts with OsSPO11-2, OsSPO11-3, and OsSPO11-4, but not with OsSPO11-1 and OsSPO11-5 [138, 234, 235]. The A subunits, which do not interact with the B subunit, likely function as Spo11 proteins in other eukaryotes. For example, AtSPO11-1 and OsSPO11-1 are required for meiotic recombination [238, 239]. Although OsSPO11-4 interacts with OsTOP6B, it is also required for meiosis in pollen grains; therefore, participation in this process may be one of the functions of plant Topo VI [235].

Mutations in or suppressed expression of the genes of Topo VI subunits that form full-length topoisomerase and are not involved in meiotic recombination cause a dwarf phenotype in plants and a decrease in cell size. These plants lack trichomes and root hairs [237, 240,

241]. Mutants were shown to have impaired endoreduplication – somatic cell polyploidization that normally occurs in plant cells [237, 241]. For efficient functioning, Topo VI forms a complex with the MID, RHL1, and BIN4 proteins (interestingly, RHL1 and BIN4 are distantly similar to the Top2α CTD of vertebrates) [237, 242, 243]. This complex is believed to participate in the regulation of the endoreduplication cycles and, probably, decatenation of chromosomes in cells with high ploidy [236, 237, 242, 243].

Overexpression of the Topo VI components of several plants in *A. thaliana* increases cell ploidy and significantly stimulates the resistance of organisms to stress conditions, such as increased salt content or drought, and reduces the sensitivity of plants to stress hormone abscisic acid [234, 241]. Overexpression of topoisomerase genes changes the levels of many transcripts. For example, it leads to the activation of stress-response cascades [234]. Topo VI was found to be also involved in plant response to oxidative stress through binding to the promoters of some genes [244]. The mechanism by which Topo VI affects transcription – relaxation of supercoils, introduction of breaks in DNA (like Top2β), or chromatin remodeling – remains unknown. In addition, it is not clear how endoreduplication and response to stress, both processes that involve Topo VI, are related.

TOPOISOMERASES OF VIRUSES AND MOBILE GENETIC ELEMENTS

Top2-like topoisomerases

Viruses with large double-stranded DNA genomes (e.g., T4-like viruses and nucleo-cytoplasmic large DNA viruses (NCLDV)) encode their own Top2-like enzymes [11]. NCLDV topoisomerases (eukaryotic viruses) are a sister group of Top2 of their hosts. The phylogenetic position of bacteriophage T4 topoisomerases is less certain; their amino acid sequences are equally distant from those of bacterial and eukaryotic type IIA enzymes [11]. However, the structure and activity of these virus topoisomerases are conserved: the bacteriophage T4 enzyme, which is encoded by three genes, relaxes supercoils, decatenates circular DNA, and is sensitive to some Top2 inhibitors [245, 246]. Topoisomerases are believed to be necessary for the removal of positive supercoils that arise during the replication of the viral genome [247, 248].

DNA gyrase

DNA gyrase genes have been predicted in the genome of the giant bacteriophage AR9 and several related viruses from the Myoviridae group [249]. The functions and role of this enzyme are unknown.

Topoisomerase VIII

Genes of topoisomerases with predicted domains similar to the Topo VI domains are found in some archaeal and bacterial plasmids, as well as in integrated mobile genetic elements. The topoisomerases encoded by these genes are allocated into a separate group of type IIB topoisomerases and are referred to as "Topo VIII" [250, 251]. Several Topo VIII were shown to relax supercoiled plasmids and decatenate circular DNA molecules *in vitro* [250]. Recently, a new group of proteins homologous to the A-subunit of Topo VIII was identified; they are called Mini-A because of their relatively small size (*Fig. 2*) [251]. The function of these topoisomerases is unknown. Probably, they help to maintain plasmids and promote their propagation in host cells.

CONCLUSION

Topoisomerases resolve topological problems that arise from DNA helicity. These enzymes are rather abundant and are required for fundamental cellular processes. According to one hypothesis, topoisomerases arose and spent the early stages of their evolution in viruses where they formed all known groups at or before the time when the last universal common ancestor (LUCA) existed. During the division of cellular organisms into modern domains, viruses spread, transferred, and mixed topoisomerase genes [11, 250]. It is likely that the variety of topoisomerases is only the tip of the iceberg, and that further exploration of "viral dark matter" could lead to the discovery of new types and classes of enzymes with unusual properties. ●

Preparation of this review was supported by the Agreement with the Ministry of Science and Higher Education of Russian Federation No. 075-15-2019-1661 and the Russian Foundation for Basic Research (project No. 20-34-90069).

REFERENCES

1. Crick F.H.C. // Proc. Natl. Acad. Sci. USA. 1976. V. 73. № 8. P. 2639–2643.
2. Mirkin S. Encyclopedia of Life Sciences. Chichester: John Wiley & Sons Ltd., 2001.
3. Bates A., Maxwell A. DNA topology. Second Ed. Oxford: Oxford University Press, 2005.
4. Maxwell A., Bush N.G., Evans-Roberts K. // EcoSal Plus. 2015. V. 6. № 2. P. 1–34.
5. Seol Y., Neuman K.C. // Biophys. Rev. 2016. V. 8. P. 101–111.
6. Bizard A.H., Hickson I.D. // J. Biol. Chem. 2020. V. 295. P. 7138–7153.
7. Basu A., Hobson M., Lebel P., Fernandes L.E., Tretter E.M., Berger J.M., Bryant Z. // Nat. Chem. Biol. 2018. V. 14. № 6.

P. 565–574.
8. Bates A.D., Berger J.M., Maxwell A. // Nucl. Acids Res. 2011. V. 39. № 15. P. 6327–6339.
9. Brown P.O., Cozzarelli N.R. // Science. 1979. V. 206. № 4422. P. 1081–1083.
10. Forterre P., Simonetta G., Daniele G., Serre M.-C. // Biochimie. 2007. V. 89. P. 427–446.
11. Forterre P., Gadelle D. // Nucl. Acids Res. 2009. V. 37. № 3. P. 679–692.
12. Dutta R., Inouye M. // TIBS. 2000. V. 25. P. 24–28.
13. Broeck A. V., Lotz C., Ortiz J., Lamour V. // Nat. Commun. 2019. V. 10. № 1. P. 4935.
14. Schmidt B.H., Burgin A.B., Deweese J.E., Osheroff N., Berger J.M. // Nature. 2010. V. 465. № 7298. P. 641–644.
15. Berger J.M., Gambling S., Harrison S., Wang J.C. // Nature. 1996. V. 379. № 18. P. 225–232.
16. Clarke D.J., Azuma Y. // Int. J. Mol. Sci. 2017. V. 18. № 2438. P. 1–14.
17. Vos S.M., Lee I., Berger J.M. // J. Mol. Biol. 2014. V. 425. № 17. P. 1–26.
18. Mcclendon A.K., Gentry A.C., Dickey J.S., Brinch M., Bendsen S., Andersen A.H., Osheroff N. // Biochemistry. 2009. V. 47. № 50. P. 13169–13178.
19. Morais Cabral J.H., Jackson A.P., Smith C.V., Shikotra N., Maxwell A., Liddington R.C. // Nature. 1997. V. 388. № 6645. P. 903–906.
20. Vologodskii A.V., Zhang W., Rybenkov V.V., Podtelezhnikov A.A., Subramanian D., Griffith J.D., Cozzarelli N.R. // Proc. Natl. Acad. Sci. USA. 2001. V. 98. № 6. P. 3045–3049.
21. Dong K.C., Berger J.M. // Nature. 2007. V. 450. № 7173. P. 1201–1205.
22. Vologodskii A. // Nucl. Acids Res. 2009. V. 37. № 10. P. 3125–3133.
23. Gubaev A., Klostermeier D. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 34. P. 14085–14090.
24. Roca J. // J. Biol. Chem. 2004. V. 279. № 24. P. 25783–25788.
25. Gubaev A., Klostermeier D. // DNA Repair (Amst.). 2014. V. 16. P. 130–141.
26. Rudolph M.G., Klostermeier D. // J. Mol. Biol. 2013. V. 425. № 15. P. 2632–2640.
27. Basu A., Parente A.C., Bryant Z. // J. Mol. Biol. 2016. V. 428. № 9. P. 1833–1845.
28. Baird C.L., Harkins T.T., Morris S.K., Lindsley J.E. // Proc. Natl. Acad. Sci. USA. 1999. V. 96. № 24. P. 13685–13690.
29. Sugino A., Higgins N.P., Brown P.O., Peebles C.L., Cozzarelli N.R. // Proc. Natl. Acad. Sci. USA. 1978. V. 75. № 10. P. 4838–4842.
30. Bates A.D., O'Dea M.H., Gellert M. // Biochemistry. 1996. V. 35. № 5. P. 1408–1416.
31. Corbett K.D., Berger J.M. // EMBO J. 2003. V. 22. № 1. P. 151–163.
32. Corbett K.D., Benedetti P., Berger J.M. // Nat. Struct. Mol. Biol. 2007. V. 14. № 7. P. 611–619.
33. Graille M., Cladie L., Durand D., Lecointe F., Gadelle D., Quevillon-Cheruel S., Vachette P., Forterre P., van Tilbeurgh H. // Structure. 2008. V. 16. № 3. P. 360–370.
34. Buhler C., Lebbink J.H.G., Bocs C., Ladenstein R., Forterre P. // J. Biol. Chem. 2001. V. 276. № 40. P. 37215–37222.
35. Morrison A., Cozzarelli N.R. // Cell. 1979. V. 17. № 1. P. 175–184.
36. Ward D., Newton A. // Mol. Microbiol. 1997. V. 26. № 5. P. 897–910.
37. Reuß D.R., Faßhauer P., Mroch P.J., Ul-Haq I., Koo B.M., Pöhlein A., Gross C.A., Daniel R., Brantl S., Stülke J. // Nucl. Acids Res. 2019. V. 47. № 10. P. 5231–5242.
38. Huang T.W., Hsu C.C., Yang H.Y., Chen C.W. // Nucl. Acids Res. 2013. V. 41. № 22. P. 10403–10413.
39. Aubry A., Mark Fisher L., Jarlier V., Cambau E. // Biochem. Biophys. Res. Commun. 2006. V. 348. № 1. P. 158–165.
40. Ambur O.H., Davidsen T., Frye S.A., Balasingham S.V., Lagesen K., Rognes T., Tønjum T. // FEMS Microbiol. Rev. 2009. V. 33. № 3. P. 453–470.
41. McCutcheon J.P., McDonald B.R., Moran N.A. // PLoS Genet. 2009. V. 5. № 7. P. 1–11.
42. López-Madrigal S., Latorre A., Porcar M., Moya A., Gil R. // J. Bacteriol. 2011. V. 193. № 19. P. 5587–5588.
43. Tamames J., Gil R., Latorre A., Peretó J., Silva F.J., Moya A. // BMC Evol. Biol. 2007. V. 7. № 181. P. 1–7.
44. Collin F., Karkare S., Maxwell A. // Appl. Microbiol. Biotechnol. 2011. V. 92. № 3. P. 479–497.
45. Dighe S.N., Collet T.A. // Eur. J. Med. Chem. 2020. V. 199. № 8. P. 1–100.
46. Hooper D.C., Jacoby G.A. // Cold Spring Harb. Perspect. Med. 2016. V. 6. № 9. P. 1–22.
47. Wohlkonig A., Chan P.F., Fosberry A.P., Homes P., Huang J., Kranz M., Leydon V.R., Miles T.J., Pearson N.D., Perera R.L., et al. // Nat. Struct. Mol. Biol. 2010. V. 17. № 9. P. 1152–1153.
48. Bax B.D., Chan P.F., Eggleston D.S., Fosberry A., Gentry D.R., Gorrec F., Giordano I., Hann M.M., Hennessy A., Hibbs M., et al. // Nature. 2010. V. 466. № 7309. P. 935–940.
49. Aldred K.J., McPherson S.A., Turnbough C.L., Kerns R.J., Osheroff N. // Nucl. Acids Res. 2013. V. 41. № 8. P. 4628–4639.
50. Alt S., Mitchenall L.A., Maxwell A., Heide L. // J. Antimicrob. Chemother. 2011. V. 66. № 9. P. 2061–2069.
51. Butler M.S., Paterson D.L. // J. Antibiot. (Tokyo). 2020. V. 2019. № 10. P. 329–364.
52. Lewis K. // Cell. 2020. V. 181. № 1. P. 29–45.
53. Weidlich D., Klostermeier D. // J. Biol. Chem. 2020. V. 295. № 8. P. 2299–2312.
54. Rovinskiy N.S., Agbleke A.A., Chesnokova O.N., Patrick Higgins N. // Microorganisms. 2019. V. 7. № 3. P. 1–22.
55. Guo M.S., Haakonsen D.L., Zeng W., Schumacher M.A., Laub M.T. // Cell. 2018. V. 175. № 2. P. 583–597.
56. Ashley R.E., Dittmore A., McPherson S.A., Turnbough C.L., Neuman K.C., Osheroff N. // Nucl. Acids Res. 2017. V. 45. № 16. P. 9611–9624.
57. Grompone G., Ehrlich S.D., Michel B. // Mol. Microbiol. 2003. V. 48. № 3. P. 845–854.
58. Rizzo M.F., Shapiro L., Gober J. // J. Bacteriol. 1993. V. 175. № 21. P. 6970–6981.
59. Ogasawara N., Seiki M., Yoshikawa H. // Nature. 1979. V. 281. № 5733. P. 702–704.
60. Guha S., Udupa S., Ahmed W., Nagaraja V. // J. Mol. Biol. 2018. V. 430. № 24. P. 4986–5001.
61. Nakanishi A., Oshida T., Matsushita T., Imajoh-Ohmi S., Ohnuki T. // J. Biol. Chem. 1998. V. 273. № 4. P. 1933–1938.
62. Wu H.Y., Shyy S., Wang J.C., Liu L.F. // Cell. 1988. V. 53. № 3. P. 433–440.
63. Travers A., Muskhelishvili G. // Nat. Rev. Microbiol. 2005. V. 3. № 2. P. 157–169.
64. Vijayan V., Zuzow R., O'Shea E.K. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 52. P. 22564–22568.
65. Jeong K.S., Ahn J., Khodursky A.B. // Genome Biol. 2004. V. 5. № 11. P. 1–10.
66. Sutormin D., Rubanova N., Logacheva M., Ghilarov D., Severinov K. // Nucl. Acids Res. 2019. V. 47. № 3. P. 1–16.
67. Ahmed W., Sala C., Hegde S.R., Jha R.K., Cole S.T., Nagaraja V. // PLoS Genet. 2017. V. 13. № 5. P. 1–20.

68. Chong S., Chen C., Ge H., Xie X.S. // Cell. 2015. V. 158. № 2. P. 314–326.
69. Dorman C.J. // BMC Mol. Cell Biol. 2019. V. 20. № 1. P. 1–9.
70. Peter B.J., Arsuaga J., Breier A.M., Khodursky A.B., Brown P.O., Cozzarelli N.R. // Genome Biol. 2004. V. 5. № 11. P. 1–16.
71. Blot N., Mavathur R., Geertz M., Travers A., Muskhelishvili G. // EMBO Rep. 2006. V. 7. № 7. P. 710–715.
72. Sobetzko P., Travers A., Muskhelishvili G. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 2. P. E42–E50.
73. Martis B.S., Forquet R., Reverchon S., Nasser W., Meyer S. // Comput. Struct. Biotechnol. J. 2019. V. 17. P. 1047–1055.
74. Muskhelishvili G., Forquet R., Reverchon S., Nasser W., Meyer S. // Microorganisms. 2019. V. 7. № 12. P. 10–15.
75. Pruss G.J., Drlica K. // Cell. 1989. V. 56. № 2. P. 521–523.
76. Ahmed W., Menon S., Karthik P.V.D.N.B., Nagaraja V. // Nucl. Acids Res. 2015. V. 44. № 4. P. 1541–1552.
77. Menzel R., Gellert M. // Cell. 1983. V. 34. № 8. P. 105–113.
78. Drlica K. // Mol. Microbiol. 1992. V. 6. № 4. P. 425–433.
79. Szafran M.J., Gongerowska M., Gutkowski P., Zakrzewska-Czerwin J., Jakimowicz D. // J. Bacteriol. 2016. V. 198. № 21. P. 3016–3028.
80. Yamamoto N., Droffner M.L. // Proc. Natl. Acad. Sci. USA. 1985. V. 82. № 7. P. 2077–2081.
81. Ye F., Brauer T., Niehus E., Drlica K., Josenhans C., Suerbaum S. // Int. J. Med. Microbiol. 2007. V. 297. № 2. P. 65–81.
82. Hsu Y.H., Chung M.W., Li T.K. // Nucl. Acids Res. 2006. V. 34. № 10. P. 3128–3138.
83. Condemine G., Smith C.L. // Nucl. Acids Res. 1990. V. 18. № 24. P. 7389–7396.
84. Postow L., Hardy C.D., Arsuaga J., Cozzarelli N.R. // Genes Dev. 2004. V. 18. № 7. P. 1766–1779.
85. Pato M.L., Karlok M., Wall C., Higgins N.P. // J. Bacteriol. 1995. V. 177. № 20. P. 5937–5942.
86. Saha R.P., Lou Z., Meng L., Harshey R.M. // PLoS Genet. 2013. V. 9. № 11. P. 1–17.
87. Sawitzke J.A., Austin S. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. № 4. P. 1671–1676.
88. Le T.B., Imakaev M.V., Mirny L.A., Laub M.T. // Science. 2014. V. 342. № 6159. P. 731–734.
89. Crisona N.J., Strick T.R., Bensimon D., Croquette V., Cozzarelli N.R. // Genes Dev. 2000. V. 14. № 22. P. 2881–2892.
90. Peng H., Marians K.J. // Proc. Natl. Acad. Sci. USA. 1993. V. 90. № 18. P. 8571–8575.
91. Ullsperger C., Cozzarelli N.R. // J. Biol. Chem. 1996. V. 271. № 49. P. 31549–31555.
92. Hiasa H., DiGate R.J., Marians K.J. // J. Biol. Chem. 1994. V. 269. № 3. P. 2093–2099.
93. Seol Y., Hardin A.H., Strub M.P., Charvin G., Neuman K.C. // Nucl. Acids Res. 2013. V. 41. № 8. P. 4640–4649.
94. Blanche F., Cameron B., Bernard F.X., Maton L., Manse B., Ferrero L., Ratet N., Lecoq C., Goniot A., Bisch D., et al. // Antimicrob. Agents Chemother. 1996. V. 40. № 12. P. 2714–2720.
95. Deibler R., Rahmati S., Zechiedrich E.L. // Genes Dev. 2001. V. 15. P. 748–761.
96. Kampranis S.C., Maxwell A. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. № 25. P. 14416–14421.
97. Hsieh T.J., Farh L., Huang W.M., Chan N.L. // J. Biol. Chem. 2004. V. 279. № 53. P. 55587–55593.
98. Kato J., Nishimura Y., Imamura R., Niki H., Hiraga S., Suzuki H., Grainge I., Bregu M., Vazquez M., Sivanathan V., et al. // EMBO J. 1990. V. 26. № 2. P. 4228–4238.
99. Grainge I., Bregu M., Vazquez M., Sivanathan V., Ip S.C.Y., Sherratt D.J. // EMBO J. 2007. V. 26. № 19. P. 4228–4238.
100. Wang X., Reyes-Lamothe R., Sherratt D.J. // Genes Dev. 2008. V. 22. № 17. P. 2426–2433.
101. Wang S.C., Shapiro L. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 25. P. 9251–9256.
102. Zechiedrich E.L., Khodursky A.B., Cozzarelli N.R. // Genes Dev. 1997. V. 11. № 19. P. 2580–2592.
103. Cui T., Moro-oka N., Ohsumi K., Kodama K., Ohshima T., Ogasawara N., Mori H., Wanner B., Niki H., Horiuchi T. // EMBO Rep. 2007. V. 8. № 2. P. 181–187.
104. Manjunatha U.H., Dalal M., Chatterji M., Radha D.R., Visweswariah S.S., Nagaraja V. // Nucl. Acids Res. 2002. V. 30. № 10. P. 2144–2153.
105. Debowski A.W., Carnoy C., Verbrugghe P., Nilsson H.O., Gauntlett J.C., Fulurija A., Camilleri T., Berg D.E., Marshall B.J., Benghezal M. // PLoS One. 2012. V. 7. № 4. P. 1–15.
106. Khodursky A.B., Peter B.J., Schmid M.B., DeRisi J., Botstein D., Brown P.O., Cozzarelli N.R. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. № 17. P. 9419–9424.
107. El Sayyed H., Le Chat L., Lebailly E., Vickridge E., Pages C., Cornet F., Cosentino Lagomarsino M., Espéli O. // PLoS Genet. 2016. V. 12. № 5. P. 1–22.
108. Brochu J., Vlachos-Breton E., Sutherland S., Martel M., Drolet M. // PLoS Genet. 2018. № 9. P. 1–25.
109. Kang S., Han J.S., Park J.H., Skarstad K., Hwang D.S. // J. Biol. Chem. 2003. V. 278. № 49. P. 48779–48785.
110. Joshi M.C., Magnan D., Montminy T.P., Lies M., Stepankiw N., Bates D. // PLoS Genet. 2013. V. 9. № 8. P. 1–13.
111. Li Y., Stewart N.K., Berger A.J., Vos S., Schoeffler A.J., Berger J.M., Chait B.T., Oakley M.G. // Proc. Natl. Acad. Sci. USA. 2010. V. 107. № 44. P. 18832–18837.
112. Hayama R., Marians K.J. // Proc. Natl. Acad. Sci. USA. 2010. V. 107. № 44. P. 18826–18831.
113. Espeli O., Lee C., Marians K.J. // J. Biol. Chem. 2003. V. 278. № 45. P. 44639–44644.
114. Hojgaard A., Szerlong H., Tabor C., Kuempel P. // Mol. Microbiol. 1999. V. 33. № 5. P. 1027–1036.
115. Narayanan S., Janakiraman B., Kumar L., Radhakrishnan S.K. // Genes Dev. 2015. V. 29. № 11. P. 1175–1187.
116. Zawadzki P., Stracy M., Ginda K., Zawadzka K., Lesterlin C., Kapanidis A.N., Sherratt D.J. // Cell Rep. 2015. V. 13. № 11. P. 2587–2596.
117. Badrinarayanan A., Reyes-Lamothe R., Uphoff S., Leake M.C., Sherratt D.J. // Science. 2012. V. 338. № 6106. P. 528–531.
118. Badrinarayanan A., Lesterlin C., Reyes-Lamothe R., Sherratt D. // J. Bacteriol. 2012. V. 194. № 17. P. 4669–4676.
119. Nicolas E., Upton A.L., Uphoff S., Henry O., Badrinarayanan A., Sherratt D. // mBio. 2014. V. 5. № 1. P. 1–10.
120. Jain P., Nagaraja V. // Mol. Microbiol. 2005. V. 58. № 5. P. 1392–1405.
121. Rudolph M.G., Del Toro Duany Y., Jungblut S.P., Ganguly A., Klostermeier D. // Nucl. Acids Res. 2013. V. 41. № 2. P. 1058–1070.
122. Panas M.W., Jain P., Yang H., Mitra S., Biswas D., Wattam A.R., Letvin N.L., Jacobs W.R. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. № 37. P. 13264–13271.
123. Doron S., Melamed S., Ofir G., Leavitt A., Lopatina A., Keren M., Amitai G., Sorek R. // Science. 2018. V. 359. № 6379. P. 1–12.
124. Tadesse S., Mascarenhas J., Kösters B., Hasilik A., Graumann P.L. // Microbiology. 2005. V. 151. № 11. P. 3729–3737.
125. Lipscomb G.L., Hahn E.M., Crowley A.T., Adams M.W.W. // Extremophiles. 2017. V. 21. № 3. P. 603–608.
126. Couturier M., Gadelle D., Forterre P., Nadal M., Garnier

F. // Mol. Microbiol. 2020. V. 113. № 2. P. 356–368.
127. Han W., Feng X., She Q. // Int. J. Mol. Sci. 2017. V. 18. № 7. P. 1–13.
128. Bergerat A., De Massy B., Gadelle D., Varoutas P.C., Nicolas A., Forterre P. // Nature. 1997. V. 386. № 6623. P. 414–417.
129. Wendorff T.J., Berger J.M. // Elife. 2018. V. 7. P. 1–35.
130. Visone V., Vettone A., Serpe M., Valenti A., Perugino G., Rossi M., Ciaramella M. // Int. J. Mol. Sci. 2014. V. 15. № 9. P. 17162–17187.
131. Lopez-Garcia P., Forterre P. // Mol. Microbiol. 1999. V. 33. № 4. P. 766–777.
132. Sioud M., Possot O., Elie C., Sibold L., Forterre P. // J. Bacteriol. 1988. V. 170. № 2. P. 946–953.
133. Holmes M.L., Dyall-Smith M.L. // J. Bacteriol. 1991. V. 173. № 2. P. 642–648.
134. Yamashiro K., Yamagishi A. // J. Bacteriol. 2005. V. 187. № 24. P. 8531–8536.
135. Nagano S., Lin T.Y., Edula J.R., Heddle J.G. // BMC Bioinformatics. 2014. V. 15. № 1. P. 1–15.
136. Vrielynck N., Chambon A., Vezon D., Pereira L., Chelysheva L., De Muyt A., Mézard C., Mayer C., Grelon M. // Science. 2016. V. 351. № 6276. P. 939–944.
137. Robert T., Nore A., Brun C., Maffre C., Crimi B., Bourbon H.-M., de Massy B. // Science. 2016. V. 351. № 6276. P. 943–949.
138. Hartung F., Puchta H. // Gene. 2001. V. 271. № 1. P. 81–86.
139. Pommier Y., Leo E., Zhang H., Marchand C. // Chem. Biol. 2010. V. 17. № 5. P. 421–433.
140. Pommier Y., Sun Y., Huang S.Y.N., Nitiss J.L. // Nat. Rev. Mol. Cell Biol. 2016. V. 17. № 11. P. 703–721.
141. Pommier Y. // ACS Chem. Biol. 2013. V. 8. № 1. P. 82–95.
142. Deweese J.E., Osheroff N. // Nucl. Acids Res. 2009. V. 37. № 3. P. 738–748.
143. Atkin N.D., Raimer H.M., Wang Y.H. // Genes (Basel). 2019. V. 10. № 10. P. 791–813.
144. Walker J.V., Nitiss J.L. // Cancer Invest. 2002. V. 20. № 4. P. 570–589.
145. Bailly C. // Chem. Rev. 2012. V. 112. № 7. P. 3611–3640.
146. Holden J.A. // Curr. Med. Chem. Anticancer. Agents. 2001. V. 1. № 1. P. 1–25.
147. Kaufmann S.H. // Biochim. Biophys. Acta–Gene Struct. Expr. 1998. V. 1400. № 1–3. P. 195–211.
148. Baguley B.C., Ferguson L.R. // Biochim. Biophys. Acta–Gene Struct. Expr. 1998. V. 1400. № 1–3. P. 213–222.
149. Wilstermann A., Osheroff N. // Curr. Top. Med. Chem. 2005. V. 3. № 3. P. 321–338.
150. Ketron A.C., Osheroff N. // Phytochem. Rev. 2014. V. 13. № 1. P. 19–35.
151. Baldwin E.L., Osheroff N. // Curr. Med. Chem.–Anti-Cancer Agents. 2005. V. 5. № 4. P. 363–372.
152. Nitiss J.L. // Nat. Rev. Cancer. 2009. V. 9. № 5. P. 338–350.
153. Milyavsky M., Gan O.I., Trottier M., Komosa M., Tabach O., Notta F., Lechman E., Hermans K.G., Eppert K., Konovalova Z., et al. // Cell Stem Cell. 2010. V. 7. № 2. P. 186–197.
154. Bracker T.U., Giebel B., Spanholtz J., Sorg U.R., Klein-Hitpass L., Moritz T., Thomale J. // Stem Cells. 2006. V. 24. № 3. P. 722–730.
155. Ashour M.E., Atteya R., El-Khamisy S.F. // Nat. Rev. Cancer. 2015. V. 15. № 3. P. 137–151.
156. Álvarez-Quilón A., Terrón-Bautista J., Delgado-Sainz I., Serrano-Benítez A., Romero-Granados R., Martínez-García P.M., Jimeno-González S., Bernal-Lozano C., Quintero C., García-Quintanilla L., et al. // Nat. Commun. 2020. V. 11. № 1. P. 1–14.
157. Felix C.A. // Biochim. Biophys. Acta – Gene Struct. Expr. 1998. V. 1400. № 1–3. P. 233–255.
158. Morton L.M., Dores G.M., Tucker M.A., Kim C.J., Onel K., Gilbert E.S., Fraumeni J.F., Curtis R.E. // Blood. 2013. V. 121. № 15. P. 2996–3004.
159. Azarova A.M., Lyu Y.L., Lin C.P., Tsai Y.C., Lau J.Y.N., Wang J.C., Liu L.F. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. № 26. P. 11014–11019.
160. Cowell I.G., Sondka Z., Smith K., Lee K.C., Manville C.M., Sidorczuk-Lesthuruge M., Rance H.A., Padget K., Jackson G.H., Adachi N., et al. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 23. P. 8989–8994.
161. Zhang S., Liu X., Bawa-Khalfe T., Lu L.S., Lyu Y.L., Liu L.F., Yeh E.T.H. // Nat. Med. 2012. V. 18. № 11. P. 1639–1642.
162. Willmore E., Frank A.J., Padget K., Tilby M.J., Austin C.A. // Mol. Pharmacol. 1998. V. 54. № 1. P. 78–85.
163. Willmore E., Errington F., Tilby M.J., Austin C.A. // Biochem. Pharmacol. 2002. V. 63. № 10. P. 1807–1815.
164. Larsen A.K., Escargueil A.E., Skladanowski A. // Pharmacol. Ther. 2003. V. 99. № 2. P. 167–181.
165. Henriksen P.A. // Heart. 2018. V. 104. № 12. P. 971–977.
166. Marinello J., Delcuratolo M., Capranico G. // Int. J. Mol. Sci. 2018. V. 19. № 11. P. 3480–3498.
167. Deng S., Yan T., Jendrny C., Nemecek A., Vincetic M., Gödtel-Armbrust U., Wojnowski L. // BMC Cancer. 2014. V. 14. № 1. P. 1–11.
168. Vavrova A., Jansova H., Mackova E., Machacek M., Haskova P., Tichotova L., Sterba M., Simunek T. // PLoS One. 2013. V. 8. № 10. P. 1–13.
169. McClendon A.K., Rodriguez A.C., Osheroff N. // J. Biol. Chem. 2005. V. 280. № 47. P. 39337–39345.
170. Uemura T., Ohkura H., Adachi Y., Morino K., Shiozaki K., Yanagida M. // Cell. 1987. V. 50. № 6. P. 917–925.
171. Swanston A., Zabrady K., Ferreira H.C. // Nucl. Acids Res. 2019. V. 47. № 12. P. 6172–6183.
172. Kawano S., Kato Y., Okada N., Sano K., Tsutsui K., Tsutsui K.M., Ikeda S. // J. Biochem. 2015. V. 159. № 3. P. 363–369.
173. Ladouceur A.M., Ranjan R., Smith L., Fadero T., Heppert J., Goldstein B., Maddox A.S., Maddox P.S. // J. Cell Biol. 2017. V. 216. № 9. P. 2645–2655.
174. Mengoli V., Bucciarelli E., Lattao R., Piergentili R., Gatti M., Bonaccorsi S. // PLoS Genet. 2014. V. 10. № 10. P. 15–17.
175. Samejima K., Samejima I., Vagnarelli P., Ogawa H., Vargiu G., Kelly D.A., Alves F. de L., Kerr A., Green L.C., Hudson D.F., et al. // J. Cell Biol. 2012. V. 199. № 5. P. 755–770.
176. Gonzalez R.E., Lim C.U., Cole K., Bianchini C.H., Schools G.P., Davis B.E., Wada I., Roninson I.B., Broude E.V. // Cell Cycle. 2011. V. 10. № 20. P. 3505–3514.
177. Dovey M., Patton E.E., Bowman T., North T., Goessling W., Zhou Y., Zon L.I. // Mol. Cell. Biol. 2009. V. 29. № 13. P. 3746–3753.
178. Yang X., Li W., Prescott E.D., Burden S.J., Wang J.C. // Science. 2000. V. 287. № 5450. P. 131–134.
179. Lyu Y.L., Lin C.-P., Azarova A.M., Cai L., Wang J.C., Liu L.F. // Mol. Cell. Biol. 2006. V. 26. № 21. P. 7929–7941.
180. Petrova B., Dehler S., Kruitwagen T., Heriche J.-K., Miura K., Haering C.H. // Mol. Cell. Biol. 2013. V. 33. № 5. P. 984–998.
181. Zhang H., Zhang Y.W., Yasukawa T., Dalla Rosa I., Khiati S., Pommier Y. // Nucl. Acids Res. 2014. V. 42. № 11. P. 7259–7267.
182. Woessner R.D., Mattern M.R., Mirabelli C.K., Johnson R.K., Drake F.H. // Cell Growth Differ. 1991. V. 2. № 4.

P. 209–214.
183. Singh B.N., Mudgil Y., Sopory S.K., Reddy M.K. // Plant Mol. Biol. 2003. V. 52. № 5. P. 1063–1076.
184. Turley H., Comley M., Houlbrook S., Nozaki N., Kikuchi A., Hickson I.D., Gatter K., Harris A.L. // Br. J. Cancer. 1997. V. 75. № 9. P. 1340–1346.
185. Austin C.A., Marsh K.L. // BioEssays. 1998. V. 20. № 3. P. 215–226.
186. Linka R.M., Porter A.C.G., Volkov A., Mielke C., Boege F., Christensen M.O. // Nucl. Acids Res. 2007. V. 35. № 11. P. 3810–3822.
187. Meczes E.L., Gilroy K.L., West K.L., Austin C.A. // PLoS One. 2008. V. 3. № 3. P. 1–7.
188. Gilroy K.L., Austin C.A. // PLoS One. 2011. V. 6. № 2. P. 1–7.
189. Kouzine F., Gupta A., Baranello L., Wojtowicz D., Ben-Aissa K., Liu J., Przytycka T.M., Levens D. // Nat. Struct. Mol. Biol. 2013. V. 20. № 3. P. 396–403.
190. Joshi R.S., Piña B., Roca J. // Nucl. Acids Res. 2012. V. 40. № 16. P. 7907–7915.
191. Canela A., Maman Y., Huang S. yin N., Wutz G., Tang W., Zagnoli-Vieira G., Callen E., Wong N., Day A., Peters J.M., et al. // Mol. Cell. 2019. V. 75. № 2. P. 252–266.e8.
192. King I.F., Yandava C.N., Mabb A.M., Hsiao J.S., Huang H.S., Pearson B.L., Calabrese J.M., Starmer J., Parker J.S., Magnuson T., et al. // Nature. 2013. V. 501. № 7465. P. 58–62.
193. Yu X., Davenport J.W., Urtishak K.A., Carillo M.L., Gosai S.J., Kolaris C.P., Byl J.A.W., Rappaport E.F., Osheroff N., Gregory B.D., et al. // Genome Res. 2017. V. 27. № 7. P. 1238–1249.
194. Durand-Dubief M., Persson J., Norman U., Hartsuiker E., Ekwall K. // EMBO J. 2010. V. 29. № 13. P. 2126–2134.
195. Sperling A.S., Jeong K.S., Kitada T., Grunstein M. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 31. P. 12693–12698.
196. Pedersen J.M., Fredsoe J., Roedgaard M., Andreasen L., Mundbjerg K., Kruhøffer M., Brinch M., Schierup M.H., Bjergbaek L., Andersen A.H. // PLoS Genet. 2012. V. 8. № 12. P. 1–15.
197. Ju B.G., Lunyak V.V., Perissi V., Garcia-Bassets I., Rose D.W., Glass C.K., Rosenfeld M.G. // Science. 2006. V. 312. № 5781. P. 1798–1802.
198. Williamson L.M., Lees-Miller S.P. // Carcinogenesis. 2011. V. 32. № 3. P. 279–285.
199. Haffner M.C., Aryee M.J., Toubaji A., Esopi D.M., Albadine R., Gurel B., Isaacs W.B., Bova G.S., Liu W., Xu J., et al. // Nat. Genet. 2010. V. 42. № 8. P. 668–675.
200. Trotter K.W., King H.A., Archer T.K. // Mol. Cell. Biol. 2015. V. 35. № 16. P. 2799–2817.
201. Madabhushi R., Gao F., Pfenning A.R., Pan L., Yamakawa S., Seo J., Rueda R., Phan T.X., Yamakawa H., Pao P.C., et al. // Cell. 2015. V. 161. № 7. P. 1592–1605.
202. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Liu J.S., Ren B. // Nature. 2012. V. 485. № 7398. P. 376–380.
203. Sati S., Cavalli G. // Chromosoma. 2017. V. 126. № 1. P. 33–44.
204. Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., et al. // Science. 2009. V. 326. № 5950. P. 289–293.
205. Rao S.S.P., Huntley M.H., Durand N.C., Stamenova E.K., Bochkov I.D., Robinson J.T., Sanborn A.L., Machol I., Omer A.D., Lander E.S., et al. // Cell. 2014. V. 159. № 7. P. 1665–1680.
206. Merkenschlager M., Nora E.P. // Annu. Rev. Genomics Hum. Genet. 2016. V. 17. № 1. P. 17–43.
207. Uusküla-Reimand L., Hou H., Samavarchi-Tehrani P., Rudan M.V., Liang M., Medina-Rivera A., Mohammed H., Schmidt D., Schwalie P., Young E.J., et al. // Genome Biol. 2016. V. 17. № 1. P. 1–22.
208. Canela A., Sridharan S., Sciascia N., Tubbs A., Meltzer P., Sleckman B.P., Nussenzweig A. // Mol. Cell. 2016. V. 63. № 5. P. 898–911.
209. Gothe H.J., Bouwman B.A.M., Gusmao E.G., Piccinno R., Petrosino G., Sayols S., Drechsel O., Minneker V., Josipovic N., Mizi A., et al. // Mol. Cell. 2019. V. 75. № 2. P. 267–283.
210. Canela A., Maman Y., Jung S., Wong N., Callen E., Day A., Kieffer-Kwon K.R., Pekowska A., Zhang H., Rao S.S.P., et al. // Cell. 2017. V. 170. № 3. P. 507–521.e18.
211. Gittens W.H., Johnson D.J., Allison R.M., Cooper T.J., Thomas H., Neale M.J. // Nat. Commun. 2019. V. 10. № 1. P. 1–16.
212. Sanborn A.L., Rao S.S.P., Huang S.-C., Durand N.C., Huntley M.H., Jewett A.I., Bochkov I.D., Chinnappan D., Cutkosky A., Li J., et al. // Proc. Natl. Acad. Sci. USA. 2015. V. 112. № 47. P. E6456–E6465.
213. Racko D., Benedetti F., Goundaroulis D., Stasiak A. // Polymers (Basel). 2018. V. 10. № 10. P. 1–11.
214. Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N., Mirny L.A. // Cell Rep. 2016. V. 15. № 9. P. 2038–2049.
215. Benedetti F., Dorier J., Stasiak A. // Nucl. Acids Res. 2014. V. 42. № 16. P. 10425–10432.
216. Havugimana P.C., Hart G.T., Nepusz T., Yang H., Turinsky A.L., Li Z., Wang P.I., Boutz D.R., Fong V., Phanse S., et al. // Cell. 2012. V. 150. № 5. P. 1068–1081.
217. Varga-Weisz P.D., Wilm M., Bonte E., Dumas K., Mann M., Becker P.B. // Nature. 1997. V. 388. № 6642. P. 598–602.
218. LeRoy G., Loyola A., Lane W.S., Reinberg D. // J. Biol. Chem. 2000. V. 275. № 20. P. 14787–14790.
219. Ho L., Crabtree G.R. // Nature. 2010. V. 463. № 7280. P. 474–484.
220. Wang W., Xue Y., Zhou S., Kuo A., Cairns B.R., Crabtree G.R. // Genes Dev. 1996. V. 10. № 17. P. 2117–2130.
221. Hota S.K., Bruneau B.G. // Dev. 2016. V. 143. № 16. P. 2882–2897.
222. Dykhuizen E.C., Hargreaves D.C., Miller E.L., Cui K., Korshunov A., Kool M., Pfister S., Cho Y.J., Zhao K., Crabtree G.R. // Nature. 2013. V. 497. № 7451. P. 624–627.
223. Rasim Barutcu A., Lian J.B., Stein J.L., Stein G.S., Imbalzano A.N. // Nucleus. 2017. V. 8. № 2. P. 150–155.
224. Thompson R.J., Mosig G. // Nucl. Acids Res. 1985. V. 13. № 3. P. 873–891.
225. Wall M.K., Mitchenall L.A., Maxwell A. // Proc. Natl. Acad. Sci. USA. 2004. V. 101. № 20. P. 7821–7826.
226. Tang Girdwood S.C., Nenortasa E., Shapiro T.A. // Physiol. Behav. 2017. V. 176. № 12. P. 139–148.
227. Cho H.S., Lee S.S., Kim K.D., Hwang I., Lim J.-S., Park Y.-I., Pai H.-S. // Plant Cell. 2004. V. 16. № 10. P. 2665–2682.
228. Evans-Roberts K.M., Breuer C., Wall M.K., Sugimoto-Shirasu K., Maxwell A. // PLoS One. 2010. V. 5. № 3. P. 1–7.
229. Evans-Roberts K.M., Mitchenall L.A., Wall M.K., Leroux J., Mylne J.S., Maxwell A. // J. Biol. Chem. 2015. V. 291. № 7. P. jbc.M115.689554.
230. Reddy M.K., Achary V.M.M., Singh B.N., Manna M., Sheri V., Panditi V., James D., Fartyal D., Ram B., Kaul T. // J. Plant Biochem. Biotechnol. 2019. V. 28. № 3. P. 291–300.
231. Heinhorst S., Cannon G., Weissbach A. // Arch. Biochem. Biophys. 1985. V. 239. № 2. P. 475–479.

232. Boyer L.A., Latek R.R., Peterson C.L. // Nat. Rev. Mol. Cell Biol. 2004. V. 5. № 2. P. 158–163.
233. Yang Z., Hou Q., Cheng L., Xu W., Hong Y., Li S., Sun Q. // Plant Cell. 2017. V. 29. № 10. P. 2478–2497.
234. Jain M., Tyagi A.K., Khurana J.P. // FEBS J. 2006. V. 273. № 23. P. 5245–5260.
235. An X.J., Deng Z.Y., Wang T. // PLoS One. 2011. V. 6. № 5. P. 1–14.
236. Sugimoto-Shirasu K., Roberts G.R., Stacey N.J., McCann M.C., Maxwell A., Roberts K. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. № 51. P. 18736–18741.
237. Sugimoto-Shirasu K., Stacey N.J., Corsar J., Roberts K., McCann M.C. // Curr. Biol. 2002. V. 12. № 20. P. 1782–1786.
238. Grelon M., Vezon D., Gendrot G., Pelletier G. // EMBO J. 2001. V. 20. № 3. P. 589–600.
239. Yu H., Wang M., Tang D., Wang K., Chen F., Gong Z., Gu M., Cheng Z. // Chromosoma. 2010. V. 119. № 6. P. 625–636.
240. Mittal A., Balasubramanian R., Cao J., Singh P., Subramanian S., Hicks G., Nothnagel E.A., Abidi N., Janda J., Galbraith D.W., et al. // J. Exp. Bot. 2014. V. 65. № 15. P. 4217–4239.
241. Tian Y., Gu H., Fan Z., Shi G., Yuan J., Wei F., Yang Y., Tian B., Cao G., Huang J. // Planta. 2019. V. 249. № 4. P. 1119–1132.
242. Kirik V., Schrader A., Uhrig J.F., Hulskamp M. // Plant Cell. 2007. V. 19. № 10. P. 3100–3110.
243. Breuer C., Stacey N.J., West C.E., Zhao Y., Chory J., Tsukaya H., Azumi Y., Maxwell A., Roberts K., Sugimoto-Shirasu K. // Plant Cell. 2007. V. 19. № 11. P. 3655–3668.
244. Šimková K., Moreau F., Pawlak P., Vriet C., Baruah A., Alexandre C., Hennig L., Apel K., Laloi C. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. № 40. P. 16360–16365.
245. Kreuzer K.N., Jongeneel C.V. // Methods Enzymol. 1983. V. 100. № 1980. P. 557–580.
246. Kreuzer K.N. // Biochim. Biophys. Acta. 1998. V. 1400. № 1–3. P. 339–347.
247. Nossal N.G., Dudas K.C., Kreuzer K.N. // Mol. Cell. 2001. V. 7. № 1. P. 31–41.
248. Hong G., Kreuzer K.N. // Mol. Cell. Biol. 2000. V. 20. № 2. P. 594–603.
249. Lavysh D., Sokolova M., Minakhin L., Yakunina M., Artamonova T., Kozyavkin S., Makarova K.S., Koonin E.V., Severinov K. // Virology. 2016. V. 495. P. 185–196.
250. Gadelle D., Krupovic M., Raymann K., Mayer C., Forterre P. // Nucl. Acids Res. 2014. V. 42. № 13. P. 8578–8591.
251. Takahashi T.S., Da Cunha V., Krupovic M., Mayer C., Forterre P., Gadelle D. // Nucl. Acids Res. 2020. V. 2. № 1. P. 1–13.

Snail-Family Proteins: Role in Carcinogenesis and Prospects for Antitumor Therapy

M. A. Yastrebova[1*], A. I. Khamidullina[1], V. V. Tatarskiy[1,2], A. M. Scherbakov[2]
[1]Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia
[2]Blokhin National Medical Research Center of Oncology, Moscow, 115478 Russia
[*]E-mail: ritayastrebova2009@gmail.com
Received June 29, 2020; in final form, September 04, 2020
DOI: 10.32607/actanaturae.11062

ABSTRACT **The review analyzes Snail family proteins, which are transcription factors involved in the regulation of the epithelial-mesenchymal transition (EMT) of tumor cells. We describe the structure of these proteins, their post-translational modification, and the mechanisms of Snail-dependent regulation of genes. The role of Snail proteins in carcinogenesis, invasion, and metastasis is analyzed. Furthermore, we focus on EMT signaling mechanisms involving Snail proteins. Next, we dissect Snail signaling in hypoxia, a condition that complicates anticancer treatment. Finally, we offer classes of chemical compounds capable of down-regulating the transcriptional activity of Snails. Given the important role of Snail proteins in cancer biology and the potential for pharmacological inhibition, Snail family proteins may be considered promising as therapeutic targets.**
KEYWORDS **transcription factors, Snail, Slug, epithelial-mesenchymal transition, mesenchymal-epithelial transition, breast cancer.**
ABBREVIATIONS **ECM – extracellular matrix; MMP – matrix metalloproteinase; MET – mesenchymal-epithelial transition; TSCs – tumor stem cells; BC – breast cancer; EMT – epithelial-mesenchymal transition.**

INTRODUCTION

During metastasis, tumor cells acquire a locomotor phenotype, enter the bloodstream, and form premetastatic niches in target organs. The colonization of metastatic niches by tumor cells leads to the formation of secondary tumors [1, 2]. The process by which highly differentiated polarized epithelial cells acquire a locomotor phenotype of mesenchymal cells is called the epithelial-mesenchymal transition (EMT) [3]. The key role in the regulation of this process is played by Snail family proteins, which are transcription factors that control the expression of the genes whose products determine the EMT phenotype(s) and, ultimately, the progression of neoplasms [4]. Over the past 15 years, new-generation antineoplastic agents have been developed. Antitumor therapy has become targeted and has focused on the individual mechanisms that regulate the vital activity of tumor cells. Clinical practice has been expanded by the introduction of protein kinase inhibitors, modulators of the death/survival balance, proteasome inhibitors, etc., which yield significant therapeutic results in certain groups of patients [5–8]. Along with classic chemotherapy regimens, personalized approaches based on the biological characteristics of a particular neoplasm have been tested. These approaches are especially important in the development of optimal treatment regimens for patients with metastasis.

Despite the progress achieved in understanding the mechanisms of metastasis, there are still no effective antimetastatic drugs; therefore, the investigation of molecules that reduce the metastatic potential of a tumor remains topical.

The review discusses the signaling pathways of Snail family proteins, their role in maintaining an aggressive behavior of a tumor cell, and prospects for the pharmacological regulation of EMT in clinical practice.

METASTASIS AND EMT

Since the first description of the EMT phenomenon [3], more light has been shed on its key mechanisms. The main EMT criteria include changes in the expression of the marker genes of epithelial and mesenchymal cells, as well as the changes taking place in the morphology of cells and the increase in their migration ability. Cytokines, growth factors, and extracellular matrix (ECM) molecules activate the signaling pathways that trigger the EMT program. These pathways

are mediated by a number of transcription factors (Slug, Snail, ZEB1/2, Twist1/2, etc.) that bind to the regulatory regions of target genes. Regulation of EMT by the products of these genes leads to the inhibition of epithelial markers (E-cadherin, claudins, occludin, etc.) and activation of mesenchymal markers (vimentin, fibronectin, N-cadherin, etc.). Mesenchymal cells exhibit enhanced motility, invasiveness, resistance to apoptosis, and production of ECM components [9, 10].

After acquiring a mesenchymal phenotype, tumor cells are able to migrate from the epithelial layer, via the bloodstream, and, after reaching a metastatic niche, return to their initial phenotype through the mesenchymal-epithelial transition (MET), which leads to the formation of metastases. There are studies that have explored the mechanisms of MET regulation, including the dynamic regulation of the factors that induce MET during the metastatic cascade. A gradual decrease in Snail expression in tumor cells during colonization, which is due to inhibition by microRNAs, causes MET induction: in particular, miR-34 and miR-200 inhibit Snail and ZEB1/2 transcription factors [11–13]. However, it is not entirely clear whether MET is an actively regulated process triggered by certain signaling molecules, or whether it occurs passively in the absence of factors that stimulate and maintain EMT in the metastatic site, as compared to the primary tumor.

EMT occurs in many processes in embryonic (mesoderm formation, migration of neural crest cells, left-right asymmetry determination, and parietal endoderm formation) and postnatal development [14, 15]. In disease, EMT is associated with malignant transformation, tumor progression, and fibrosis development. There are studies of Snail and Slug proteins as EMT regulators during tumor progression where they are involved in the regulation of cell survival and proliferation, invasion, and metastasis [16–18], as well as regulate energy metabolism and maintain resistance to therapy [19].

The new EMT classification includes four stages: epithelial, early hybrid, late hybrid, and mesenchymal. Snail activity was shown to increase starting from the early hybrid stage, while changes in the shape of cells, from round to elongated, occur only at the late hybrid stage. These changes are accompanied by a gradual loss of intercellular adhesion [20].

STRUCTURE OF Snail FAMILY PROTEINS

Snail family proteins, Snail/*SNAI1* and Slug/*SNAI2*, are transcriptional repressors [21]. These proteins contain a highly conserved C-terminal region that includes four (Snail) and five (Slug) zinc fingers and is involved in the binding of the proteins to the target gene promoters containing the E-box sequence. The N-terminal regions contain the evolutionarily conserved SNAG domain required for transcriptional repression and capable of binding methyltransferases and histone deacetylases [4]. Despite the similarity of the N- and C-terminal regions of Snail and Slug, the central proline-rich regions, which mediate ubiquitination and the proteolytic degradation of these proteins, are different. Snail contains a protein destruction box (DB) domain and a nuclear export signal (NES) domain, while Slug comprises a specific SLUG domain. The SNAG and SLUG domains of the Slug protein are required for the repression of the E-cadherin gene promoter. The SLUG domain interacts with the CtBP1 corepressor, while the SNAG domain interacts with the NCoR corepressor [22]. Interestingly, the SNAG domain is required for EMT induction, while the SLUG domain probably negatively regulates the Slug-mediated EMT [23] (*Fig. 1*).

The functional activity of the proteins is determined by their structure, configuration, and post-translational modifications [24].

POST-TRANSLATION MODIFICATIONS OF Snail FAMILY PROTEINS

Snail is a labile protein whose half-life is less than 4 h [25]. Like many proteins, Snail undergoes various post-translational modifications that affect its stability, intracellular localization, and transcriptional activity. There are two Snail phosphorylation sites: one controls the proteolysis of the protein in the proteasome, and the other determines its intracellular localization. Glycogen synthase kinase-3β (GSK-3β) binds to Snail and phosphorylates it, causing export of the protein from the nucleus to the cytoplasm. Subsequent GSK-3β-mediated phosphorylation in the cytoplasm promotes the binding of Snail to E3-ubiquitin ligase β-TrCP and degradation of Snail in the proteasome [26]. Both phosphorylated and non-phosphorylated Snail forms can bind to ubiquitin ligase FBXL14, which also leads to proteasomal degradation of Snail. DUB3 deubiquitinase was shown to be able to prevent the degradation of Snail in the proteasome, thereby stabilizing it [27]. Stabilization of Snail in the nucleus also involves protein kinase PAK1 that enables Snail phosphorylation at the serine residue in position 246. In turn, Snail phosphorylation by protein kinase A (PKA) at serines 11 and 92 enhances Snail transactivation [28].

Stability of the Slug transcription factor is similarly regulated and depends on phosphorylation by protein kinase GSK-3β. The Slug phosphorylation sites (Ser-4 and 88) have been identified. Phosphorylation of serine 4 is required for a Slug-mediated induction of EMT [23].

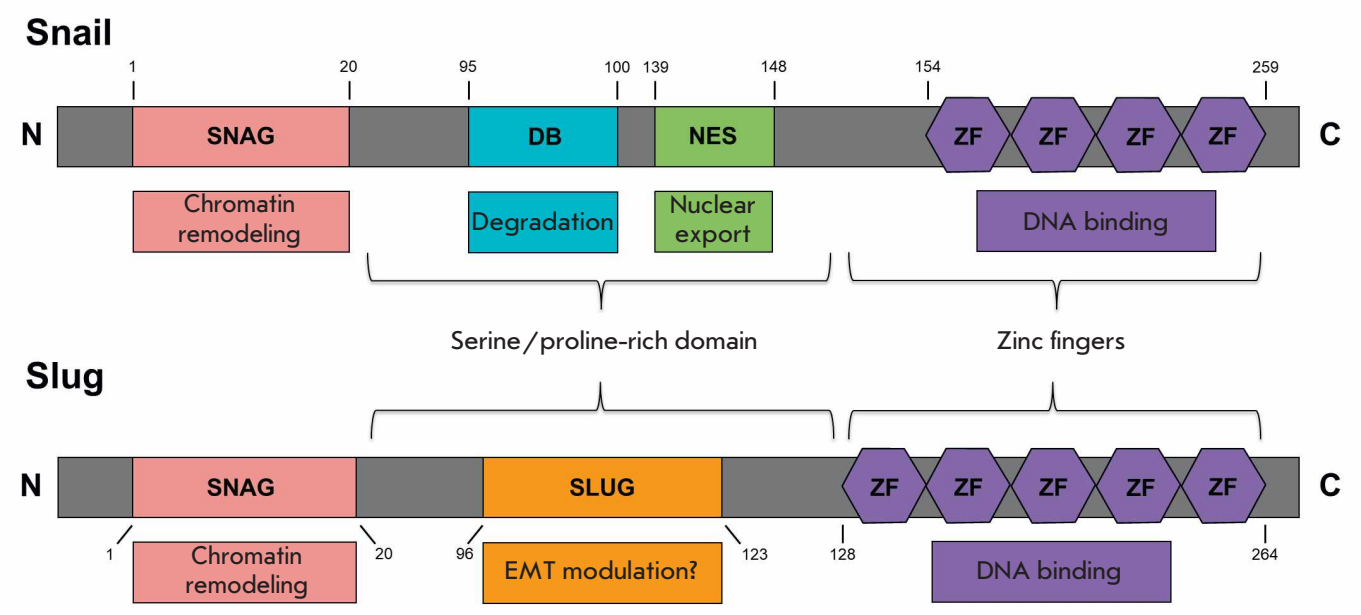


Fig. 1. The structure of Snail and Slug proteins

Stabilization of Snail/Slug involves, apart from phosphorylation by protein kinases, histone acetyltransferases (HATs) that provide nuclear localization of Snail/Slug and their interaction with co-activators [29]. E3 ubiquitin ligase A20 monoubiquitinates Snail at three lysine residues, which reduces the affinity of Snail for GSK-3β and maintains its nuclear localization, facilitating breast cancer (BC) cell EMT induced by transforming growth factor β (TGF-β1). A20 knockdown or increased Snail expression with replacement of monoubiquitinated lysine residues by arginine prevents metastasis in BC models [30].

TARGETS OF Snail FAMILY TRANSCRIPTION FACTORS

Slug and Snail proteins, despite the significant (~70%) homology of their amino acid sequences, are functionally different. For example, Snail activity is necessary in early embryogenesis, because mouse embryos with knockout *snai1* die at the gastrulation stage, due to impaired formation of the mesoderm layer, where cells retain epithelial features such as polarity, tight intercellular junctions, and E-cadherin expression [31]. *Snai2* knockout mice are viable, but they have defects in neural crest cell formation and mesoderm formation [32]. Both Snail and Slug are required for osteogenesis, chondrogenesis [33], and somitogenesis [34].

Snail and Slug are necessary for the regeneration of adult tissues; in particular, for wound healing [15]. The key role in this process is played by Slug that is controlled by the epidermal growth factor (EGF) secreted during healing [35]. In *snai2* knockout mice, there is no migration of keratinocytes into the wound while K6 and Ki-67 proliferation markers and high E-cadherin and K8 levels are retained [36].

In a human colorectal cancer model, ChIP-seq experiments demonstrated that the Snail transcription factor mainly binds to regions located upstream of the transcription start site (within 1 kbp), as well as in intergenic regions and introns distal to the promoter. Therefore, Snail controls transcription mainly through binding to distant regulatory DNA elements [37]. Snail was found to predominantly bind to the genes responsible for differentiation, morphogenesis, organogenesis, signal transduction, and cell junctions, which is in good agreement with its known biological functions [37]. In triple negative BC cells, two more Snail binding sites were identified: the TAL/GATA1 and TGG RREB1/RUNX2/PAX4 motifs, which provide more specific recognition of target genes compared to other transcription factors [38].

Snail and Slug can act both as transcriptional repressors and as activators of transcription of genes encoding mesenchymal proteins: N-cadherin, vimentin, fibronectin, etc. [39, 40]. Snail can also induce transcription by interacting with the transcription factors EGR1 and SP1 [41].

The Snail-mediated mechanism of gene expression repression was studied in detail in the case of E-cadherin, an epithelial cell marker (*Fig. 2*).

The SNAG domain of the Snail protein interacts with the Sin3A protein and the histone deacetylases (HDAC) 1 and 2. The resulting complex binds to the

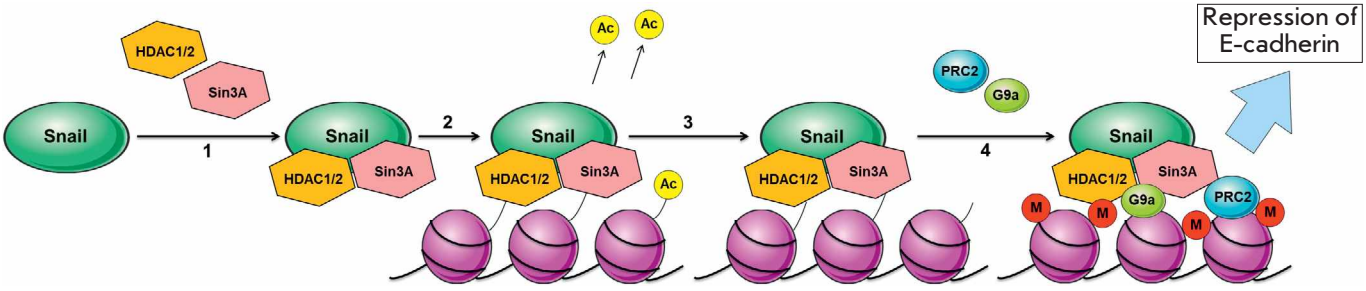


Fig. 2. Snail-mediated repression of E-cadherin. 1 – Formation of the repressor complex (Snail, HDAC1/2, Sin3A); 2 – deacetylation of H3 and H4 histones; 3 – binding of the PRC2 inhibitory complex and methyltransferase G9a; 4 – DNA hypermethylation. Adapted from [42]

E-box region in the E-cadherin gene (*CDH1*) promoter, which leads to de-acetylation of histones H3 and H4. This modification facilitates the binding of the inhibitory complex PRC2 and histone methyltransferase G9a: the second act of E-cadherin expression inhibition occurs via DNA hypermethylation. After the initial suppression of E-cadherin, Snail induces expression of the transcription factor ZEB1, which further inhibits E-cadherin expression, but through a PRC2-independent mechanism, the details of which are still unknown [42].

Snail/Slug-dependent transcription leads not only to the repression of E-cadherin but also to the disassembly of desmosomes and tight intercellular junctions due to repression of occludin, claudin 3, 4, and 7, and desmoplakin genes [43, 44]. Snail and Slug also increase synthesis of matrix metalloproteinases (MMPs), thereby promoting degradation of ECM components [45, 46].

Changes in cell motility during EMT and the development of the locomotor phenotype are associated with the activity of Rho family proteins; small GTPases Rac1, RhoA, RhoV, and Cdc42, which control actin dynamics [47]. Rac1 regulates the TGF-β-dependent activation of Snail: knockdown of Rac1 decreases the activity of Snail and MMP9 [48]. In contrast, inhibition of RhoA increases the Snail level [49]. RhoV, together with Snail, induces Slug in EMT during embryonic development [50]. The increase in the motility of pancreatic cancer cells associated with an elevated Snail level depends on Rac1 [45], and an increase in the Slug level leads to the suppression of ROCK1/2 [46]. Suppression of Snail significantly reduces cell motility because of the lower activity of Cdc42 and increased activity of RhoA [51]. Thus, both proteins, Snail and Slug, are controlled by the small GTPases responsible for cell motility and can regulate GTPase activity, enabling a coordination of changes in cell phenotypes during embryogenesis and tumor progression.

Snail plays an important role in the cell cycle and in cell survival. During embryonic development, Snail represses the transcription of the cyclin D2 gene and increases the expression of the $p21^{Cip1/WAF1}$ gene in order to regulate early-to-late G1 phase transition. An increase in the expression of cyclin-dependent kinases CDK4/6 promotes Snail stabilization through DUB3-mediated deubiquitination [27]. In renal epithelial cells (MDCK line) stably expressing exogenous Snail, about 90% of the cells remain in the G0/G1 phase after 72-h incubation. Overexpression of Snail decreases CDK4, and phosphorylation of Rb and increases the $p21^{Cip1/WAF1}$ level [52]. Thus, Snail can be used to delay or stop the transition of cells in the cell cycle.

Slug is also involved in the regulation of cell-cycle phase alteration. Slug was shown to act in functional cooperation with cyclin D1. Slug knockdown in the MDA-MB-231 triple negative BC cell line reduces the rate of cell proliferation, probably due to a decrease in the cyclin D1 level [53]. According to another study, induced Slug expression can lead to the inhibition of cyclin D1 and arrest of prostate cancer cells in the G0/G1 phase. Thus, the role of Slug varies in cells of different tissue origins [54].

Snail regulates cell survival through decreasing the serum concentration in the culture medium by activating the MAPK (Mek/Erk) and PI3K signaling pathways. Snail and Slug suppress the expression of several pro-apoptotic factors at the transcriptional level; in particular p53, BID, caspase 6, PUMA/BBC3, ATM, DFF40 (DNA fragmentation factor), and PTEN (phosphatase in the PI3K cascade) [52, 55–57]. Interestingly, the Snail protein can directly interact with the tumor suppressor p53, blocking its DNA-binding domain [58].

It is noteworthy that the transcriptional targets of Snail and Slug are similar, but information on mutual regulation of these proteins is insufficient. According to our data, expression of Snail and Slug is interdependent. For example, Snail overexpression in the

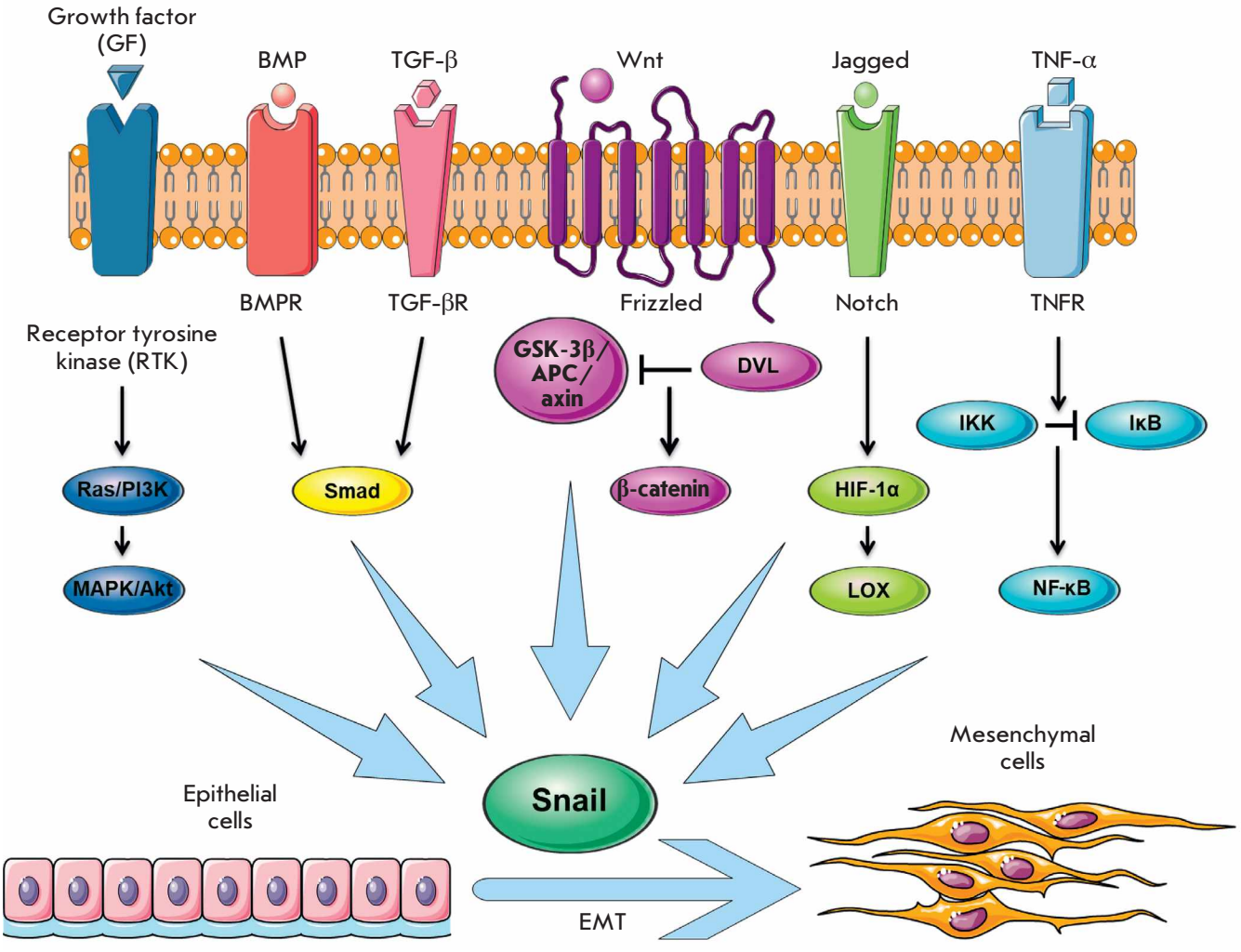


Fig. 3. Mechanisms of Snail-induced EMT

MDA-MB-231 cell line is accompanied by a sharp decrease in the Slug protein level while Snail inhibition by small interfering RNAs is associated with an increase in the Slug level. Probably, Snail and Slug compensate each other under certain conditions [59].

Various exogenous stimuli can activate Snail-family transcription factors. Below, we provide the results of an analysis of the main signaling pathways that regulate Snail and Slug.

REGULATION OF Snail-FAMILY PROTEINS DURING EMT

EMT is a dynamic process that can be initiated by ECM proteins and secreted, soluble growth factors, such as the epidermal growth factor (EGF), hepatocyte growth factor (HGF), fibroblast growth factor (FGF), bone morphogenetic proteins (BMPs), TGF-β, Wnt, Notch, tumor necrosis factor α (TNF-α), and cytokines [60, 61]. Many of these signaling molecules from the tumor cell microenvironment induce the expression of Snail-family proteins (*Fig. 3*).

Signaling cascades initiated by the activation of receptor tyrosine kinases (RTKs) and growth factors cause an increase in the level of Snail, its stabilization, and translocation into the nucleus. MAPK or PI3K signaling cascades cooperate with TGF-β to regulate EMT [62]. Repression of MAPK in some tumor models is sufficient to reduce the expression of Snail and Slug and inhibit EMT [63–65].

The multifunctional protein TGF-β regulates proliferation, differentiation, and apoptosis. TGF-β acts as a tumor growth suppressor at the early stages of carcinogenesis and promotes the formation of a malignant phenotype at later stages [66]. Snail plays an important role in regulating the response of cells to TGF-β, ensuring their resistance to TGF-β-mediated apoptosis and tumor progression. At later stages, TGF-β induces

EMT in a SMAD-dependent manner via Snail. SMAD proteins interact with the *SNAI1* gene promoter and induce Snail expression, which leads to the repression of E-cadherin and an invasive phenotype [4]. Upon TGF-β-induced EMT, Snail was shown to form a complex with SMAD3/4. This complex binds to E-box regions and SMAD-binding elements in the promoters of the genes encoding intercellular junction proteins and represses these genes [67].

Activation of the Notch signaling pathway induces Snail/Slug-mediated EMT, which promotes BC cell invasion and metastasis [68]. Notch controls Snail expression through two synergistic mechanisms: direct activation of transcription and indirect action through lysyl oxidase (LOX) that stabilizes Snail. Notch recruits the hypoxia-inducible factor 1α (HIF-1α) to the *LOX* promoter, activating this gene [67]. In addition, Jagged1-activated Notch stimulates the Slug repressor and suppresses E-cadherin, which leads to the so-called hybrid (intermediate) EMT phenotype. This phenotype is characterized by a partial increase in the expression of mesenchymal markers and a decrease in the expression of epithelial markers. In this case, there are no significant morphological changes in cells, and there is no complete loss of intercellular junctions [69].

Expression of the *SNAI1* gene can also be regulated by the nuclear factor NF-κB/p65. TNF-α-activated NF-κB binds to the *SNAI1* promoter; activation of the transcription of this gene induces EMT [25]. *SNAI1* expression can also be enhanced through the Akt signaling pathway: the protein kinase Akt1 phosphorylates IKKα, which leads to proteolytic degradation of the inhibitory subunit IκB, release of NF-κB dimers and their translocation into the nucleus, and transactivation of *SNAI1* [70]. Simultaneous suppression of Snail and NF-κB was shown to increase the sensitivity of BC cells to antiestrogens [71]. A simultaneous influence on these two transcription factors may be of interest for the development of approaches to anticancer therapy.

Activation of the Wnt signaling pathway is accompanied by the inhibition of β-catenin and Snail phosphorylation by GSK-3β, which leads to the accumulation of β-catenin and Snail in the nucleus. β-Catenin, which acts as a transcription factor in its interaction with TCF/LEF, is required for EMT induction in epithelial cells. The synergistic effect of Snail and β-catenin enables tumor cell survival during invasion and metastasis [72].

The MDM2 protein also plays a role in EMT. Increased expression of MDM2 in MCF7 BC cells leads to an epithelial-to-mesenchymal change in their morphology. On the other hand, knockdown of MDM2 in MDA-MB-231 cells changes the cell morphology from mesenchymal to epithelial (MET). In addition, enhanced expression of MDM2 increases the expression of N-cadherin and vimentin and also decreases the expression of E-cadherin at the mRNA and protein levels. Downregulation of MDM2 expression decreases the expression of N-cadherin and vimentin and increases the expression of E-cadherin. MDM2 increases the level of both mRNA and the Snail protein by activating the TGF-β-SMAD signaling pathway. *SNAI1* knockdown in cells that had entered MDM2-induced EMT was shown to return such cells to their initial epithelial phenotype. Thus, MDM2, like Snail, may be considered a therapeutic target in metastatic BC [73].

It is important that the key EMT-mediating transcription factors can affect the expression of each other. We demonstrated that knockdown of the *TWIST1* and *ZEB1* genes by small interfering RNAs decreases the Slug protein level, with no opposite effect being observed [59].

HYPOXIA AND EMT

One of the EMT regulation factors is hypoxia. Tumor growth leads to a deficiency in oxygen and nutrients in the tumor. This "starvation," on the one hand, inhibits the proliferation of cells and, on the other hand, induces adaptation processes in them, in particular EMT, which enables the tumor cells to migrate to blood vessels. Adaptation of cells to hypoxia involves hypoxia-inducible proteins, such as the HIF-1 transcription factor, a heterodimer composed of the HIF-1α and HIF-1β subunits [74, 75]. Under normoxia conditions, HIF-1α is hydroxylated by prolyl hydroxylase, which leads to the binding of HIF-1α to the Hippel–Lindau protein (VHL), a ubiquitination marker. The VHL–HIF-1α interaction leads to a degradation of HIF-1α in the proteasome. Under oxygen deficiency, the activity of prolyl hydroxylase decreases and HIF-1α fails to undergo rapid degradation because the lack of hydroxylated proline residues stabilizes HIF-1α [76]. HIF-1α accumulates in the cell and dimerizes with HIF-1β, forming an active transcription factor that is translocated into the nucleus, binds there with the hypoxia-responsive element (HRE) sites on DNA and activates the transcription of target genes.

EMT regulation during hypoxia is ensured predominantly by the HIF-1 and Snail/Slug factors. EMT induction under hypoxic conditions was shown in various tumor cell lines [77, 78]. Hypoxia decreases the expression of E-cadherin via a HIF-1α-mediated expression of *SNAI1*. In addition, HIF-1α induces LOX expression, which leads to the stabilization of Snail [79].

In response to hypoxia, the LOX protein level increases in tumor cells, and suppression of LOX expression/activity prevents metastasis. A high LOX level is considered a factor of poor clinical prognosis associated

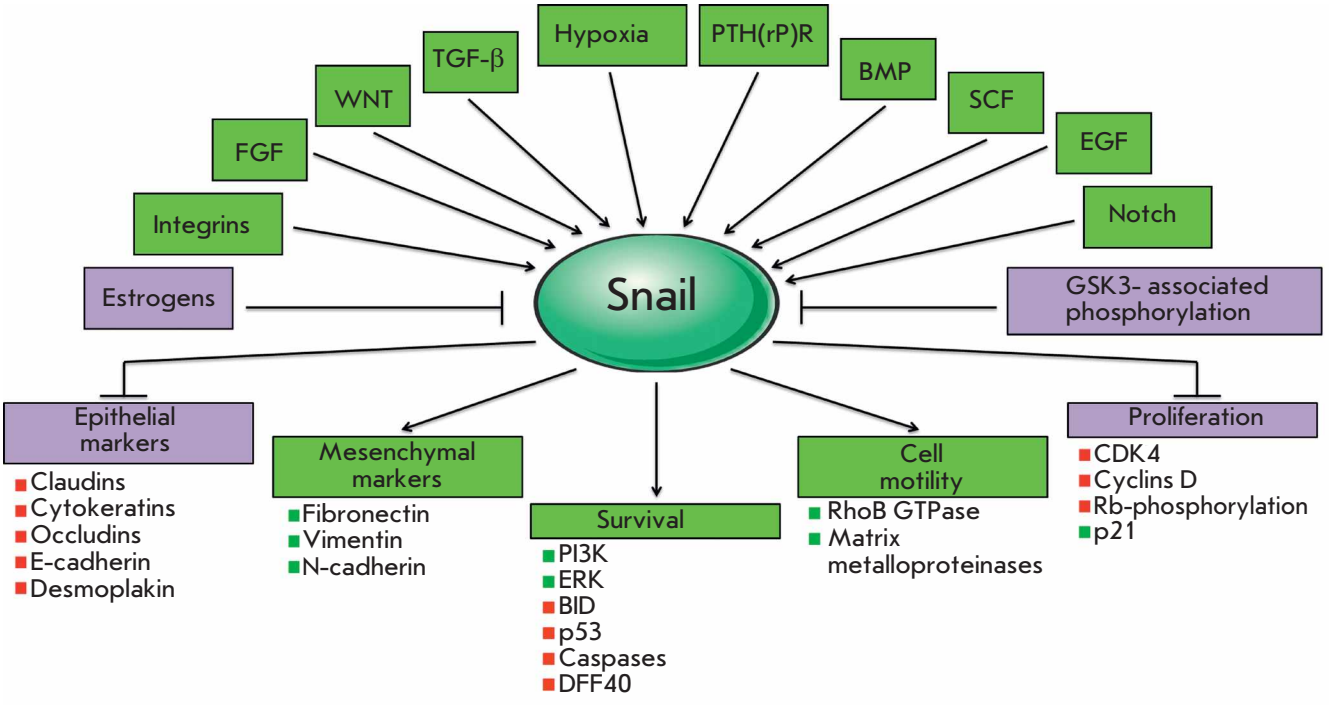


Fig. 4. Regulation and main targets of the Snail transcription factor

with the metastasis of BC and head and neck cancers [80].

Yang and co-authors could demonstrate that HIF-1α regulates the activation of EMT, increasing the Snail level in gastric cancer stem cells. HIF-1α expression in these cells is significantly increased under conditions of hypoxia. As HIF-1α increases, the expression of Snail, vimentin, and N-cadherin is elevated, and the E-cadherin level decreases, which is an indication of EMT initiation. Under hypoxia, the possibility of migration and invasion of gastric cancer stem cells significantly increases [81].

We studied the relationship between β-catenin and Snail-dependent pathways in BC cells during hypoxia and found a Snail-dependent activation of β-catenin. Activated β-catenin regulates the expression of hypoxia-response genes and maintains a resistance of BC cells to reduced partial oxygen pressure. Coordinated activation of the Snail/β-catenin/HIF-1α protein system may be considered as an important factor in determining tumor resistance to hypoxia [82].

We showed that the HBL-100 BC cell line with Snail knockdown is more sensitive to hypoxia, demonstrating blockage of replication and a decrease in the percentage of mitotic cells. In addition, the culture density directly affects the sensitivity of BC cells to hypoxia [83].

Thus, responding to hypoxia, cells acquire a mesenchymal phenotype through EMT induced by HIF-1,2α and Snail/Slug. These phenotypic changes can be regulated by various epigenetic factors [76].

Figure 4 illustrates the regulation of numerous Snail-mediated processes.

EXPRESSION OF Snail-FAMILY PROTEINS IN TUMORS AS A POTENTIAL PROGNOSTIC MARKER

Snail and Slug are aberrantly expressed in many tumors, as well as in tumor-associated fibroblasts and macrophages that colonize damaged tissues [84–86]. Numerous studies have shown that these proteins play different roles in tumor progression.

Expression of both *SNAI1* and *SNAI2* in tumor cells can characterize the degree of malignancy and serve as a prognostic marker of disease. Access to open sequence databases enables the use of various bioinformatics tools for a preliminary assessment of disease prognosis. A similar analysis is performed at the initial stage of the search and validation of new markers and clinically significant criteria. One of these databases, the KM-plotter, contains the gene expression profiles from the GEO, EGA, and TCGA databases [87]. The KM-plotter enables an assessment of the effect of gene expression on the overall survival rate of patients using the Kaplan–Meier method [88]. A total of 54,000 genes can be analyzed in 21 neoplasm types. The summarized data on the expression of *SNAI1* and *SNAI2* in tumors of each type are presented in *Table*. The analysis

Expression of *SNAI1* and *SNAI2* and overall survival rate of cancer patients: analysis of data from the KM-plotter database

Tumor	Indicator			
	*SNAI1**		*SNAI2**	
	Total survival (median)	Statistical significance	Total survival (median)	Statistical significance
Bladder cancer	Low expression (exp) = 42.33 mos, high exp = 28.63 mos	$P = 0.0264$, $q > 0.5$	Low expression (exp) = 47.33 mos, high exp = 20.77 mos	$P = 0.0008$, $q = 0.2$
Squamous cervical cancer	Low exp = 68.4 mos, high exp = 27.9 mos	$P = 0.027$, $q > 0.5$	The difference is statistically insignificant	
Esophageal adenocarcinoma	Low exp = 46.83 mos, high exp = 20.33 mos	$P = 0.0449$, $q > 0.5$	The difference is statistically insignificant	
Squamous cell carcinoma of the head and neck	Low exp = 58.73 mos, high exp = 46.6 mos	$P = 0.0398$, $q > 0.5$	Low exp = 58.73 mos, high exp = 37.77 mos	$P = 0.0174$, $q > 0.5$
Clear cell renal cell carcinoma	Low exp = 73 mos, high exp = 37.03 mos	$P = 0.0058$, $q > 0.5$	Low exp = 52.8 mos, high exp = 37.03 mos	$P = 0.0323$, $q > 0.5$
Papillary renal cell carcinoma	Low exp = 89.47 mos, high exp = 43.53 mos	$P = 8.2e{-}5$, $q = 0.02$	Low exp = 86.97 mos, high exp = 43.8 mos	$P = 0.0014$, $q > 0.2$
Lung adenocarcinoma	Low exp = 50.93 mos, high exp = 40.3 mos	$P = 0.0124$, $q > 0.5$	Low exp = 54.4 mos, high exp = 35.77 mos	$P = 0.0014$, $q > 0.5$
Squamous cell lung cancer	Low exp = 72.33 mos, high exp = 35.83 mos	$P = 0.0002$, $q = 0.05$	The difference is statistically insignificant	
Ovarian cancer	Low exp = 49.97 mos, high exp = 38.97 mos	$P = 0.0089$, $q > 0.5$	Low exp = 46.13 mos, high exp = 38.7 mos	$P = 0.0192$, $q > 0.5$
Pancreatic ductal adenocarcinoma	The difference is statistically insignificant		Low exp = 37.67 mos, high exp = 18.93 mos	$P = 0.0006$, $q = 0.2$
Rectal adenocarcinoma	Low exp = 43.8 mos, high exp = 41.93 mos	$P = 0.0384$, $q > 0.5$	The difference is statistically insignificant	
Sarcoma	The difference is statistically insignificant		Low exp = 86.63 mos, high exp = 48.87 mos	$P = 0.001$, $q = 0.2$
Gastric adenocarcinoma	Low exp = 43.8 mos, high exp = 41.93 mos	$P = 0.0384$, $q > 0.5$	Low exp = 46.9 mos, high exp = 20.23 mos	$P = 0.0013$, $q = 0.2$
Thyroid cancer	Low exp = not achieved, high exp = not achieved	$P = 3.3{-}6$, $q = 0.01$	The difference is statistically insignificant	
Uterine corpus cancer	Low exp = 114.1 mos, high exp = 51.6 mos	$P = 0.0614$, $q = 0.01$	Low exp = 36.87 mos, high exp = 78.4 mos	$P = 0.0113$, $q \geq 0.01$

*The differences in the expression of *SNAI1* and *SNAI2* are statistically insignificant in esophageal squamous cell carcinoma, liver cancer, breast cancer, and uterine corpus endometrial cancer.

involved data on the expression of these genes in 19 neoplasm types; no statistically significant differences (in at least one of the indicators) in the overall survival rate were found for four of the genes. *SNAI1* expression was shown to affect statistically significantly the median overall survival rate in 12 neoplasm types. The greatest difference in the median overall survival rate was found for squamous cervical cancer: the median survival rate was 2.4-fold higher in the group with a low *SNAI1* expression than in the group with a high expression of this gene. These data are consistent with the results reported in a recent publication by Huilun Yang *et al.* [89], who proved the relationship between *SNAI1* and *TWIST1* and active metastasis of cervical cancer. In addition, these data were confirmed by a immunohistochemical analysis [90] of 154 cervical cancer samples. The smallest (significant) difference in the overall survival rate, depending on the *SNAI1* level, was found in gastric and rectal adenocarcinomas. It is noteworthy that *SNAI2* expression does not affect overall survival indicators in rectal adenocarcinoma. The limited use of Snail as an individual (independent) prognostic marker of rectal cancer is indicated by the results of a study [91] that suggested combining EMT markers with stem cell markers to improve the predictive value of each individual indicator. Similar findings were obtained in a study of the relationship between *SNAI2* expression and the overall survival

rate in 10 tumor types. In most tumor types, a change in the *SNAI2* expression has the same tendency as in the *SNAI1* expression: high expression of the marker is considered a poor prognosis factor. An exception to this rule is uterine corpus cancer: high *SNAI2* expression in this neoplasm is associated with longer overall survival. One of the explanations for this may be the low Slug activity in uterine corpus cancer cells. For example, a nuclear localization of Slug was established only in 3.7% of tumor samples; i.e., the clinical significance of this indicator is very limited [92]. Based on other data, 25% of uterine corpus cancer cases had high Slug expression; this indicator is associated with recurrence-free survival; therefore, it may be considered a poor prognosis factor [93]. The prognostic role of Slug (or its absence) in uterine corpus cancer remains to be clarified.

Despite the absence of statistically significant differences in the overall survival of BC patients in groups with different *SNAI1* and *SNAI2* levels (KM-plotter base), a number of studies have shown the clinical significance of EMT markers: in particular Snail, in this disease. In BC cells, there is a high expression of Notch (74%), Slug (36%), Snail (62%), and N-cadherin (77%), while the expression of E-cadherin is increased in just 20% of cases [68]. An analysis of 157 BC samples revealed a statistically significant correlation between the expression of Snail and Slug and their co-activator, the NF-κB factor [94]. According to Cao *et al.*, high expression of Snail and a low level of E-cadherin correlate with the number of BC metastases in lymph nodes. In addition, a high level of Snail is largely associated with a low expression of E-cadherin, and an increased expression of Slug is associated with an increase in N-cadherin in BC patients [63].

The levels of Snail, Slug, and ZEB1 are higher in tumor cells with morphological signs of EMT (the ability to migrate and invade) than in cells without signs of EMT [95]. Knockdown of the *SNAI1* and *SNAI2* genes causes a return to an epithelial morphology and a significant decrease in the number of cells migrating in the Boyden chamber. Feng and co-authors showed that the levels of Snail, E-cadherin, Slug, and Twist – but not N-cadherin – were higher in malignant epithelial cells than in benign neoplasms [96].

A low level of E-cadherin expression and a high level of N-cadherin expression are characteristic of gastric cancer metastases and undifferentiated tumor cells, which correlates with a poor prognosis. A high expression of Snail in the primary tumor and a low expression in metastases correlate with further progression of metastasis and a negative prognosis [97].

A high expression of Snail, but not Slug, and low expression of E-cadherin are associated with poorer survival chances in bladder cancer [98]. In cervical cancer, an increase in Snail and a decrease in E-cadherin are negative prognostic factors. According to recent data, expression of Snail is a more significant predictor of this disease than the expression of other EMT regulators (Slug, ZEB1, and Twist) [99].

Overexpression of the epidermal growth factor receptor Her2/Neu stabilizes Snail, promoting drug resistance in gastric cancer [100] and BC [101]. In an inducible Her2/Neu- expressing BC model, Moody and co-authors found that the rate of tumor recurrence correlates with a high level of Snail [102]. Increased expression of Snail and Twist is associated with a poor prognosis for estrogen-positive BCs [103].

Therefore, the expression of EMT markers, in particular Snail family proteins, is associated with the degree of malignancy and, in general, with disease progression. It is reasonable to believe that the studied EMT markers can be prognostically significant in some cases [96]. But for implementation in clinical practice, it is necessary to choose analytical methods, validate them, and prove the economic feasibility of using new markers.

Snail PROTEINS AND CHEMOTHERAPY RESISTANCE

EMT regulatory proteins can control not only the ability of tumor cells to invade and undergo metastasis, but also their resistance to genotoxic and targeted anticancer drugs. The mechanisms underlying this resistance are mediated by anti-apoptotic effects, decreased proliferation, and the emergence of multidrug resistance. The role played by Slug and Snail in the development of resistance to chemotherapy and radiotherapy has been shown in a number of studies [104].

For example, the Snail protein level is increased in cisplatin-resistant tumors and cell lines [105]. In addition, Snail induces gemcitabine resistance in pancreatic cancer [106] and BC [107] models and etoposide resistance in a small-cell lung cancer model [108].

Haslehurst and co-authors showed that expression of the *SNAI1*, *SNAI2*, *TWIST*, and *ZEB2* genes is increased in the ovarian cancer A2780 cell line resistant to cisplatin. Cisplatin-resistant cells had a mesenchymal phenotype and lacked intercellular junctions, while sensitive cells retained epithelial morphology. Upon knockdown of the genes of key EMT regulators, Snail and Slug, cells returned to their initial epithelial phenotype in [42].

The stability of Snail under the action of cisplatin is due to deubiquitination of Snail by the USP1 protein that is induced upon DNA damage and stabilizes a number of repair and anti-apoptotic proteins [109]. Snail is similarly stabilized by TGF-β-activated USP27x deubiquitinase in a cisplatin resistance model [110]. The repair enzymes PARP-1 and PARP-3 are

another mechanism of the relationship between DNA damage and Snail expression in response to chemotherapy. PARP-1 controls Snail expression at the transcriptional level in cells exposed to doxorubicin, and ABT-888, a PARP-1 inhibitor, is able to enhance the response of BC cells (MDA-MB-231 line) to doxorubicin. Inhibition of PARP-1 can increase tumor cell sensitivity *in vivo* by decreasing the expression of Snail [111]. Similarly, PARP-3 depletion inhibits the TGF-β-dependent EMT of BC cells, preventing the binding of Snail to E-cadherin and increasing their sensitivity to chemotherapy [112].

Snail-family transcription factors also mediate cell resistance to certain targeted drugs. Slug expression is increased in a lung cancer model resistant to gefitinib, an EGFR inhibitor, and in biopsies from patients treated with EGFR inhibitors. In this model, Slug repressed caspase-9 and pro-apoptotic protein Bim and suppression of Slug increased the sensitivity of cells to EGFR inhibitors [113]. Snail determines the resistance of triple-negative BC cells to rapamycin and everolimus, which are mTOR protein kinase inhibitors. In this model, trametinib, a histone deacetylase inhibitor, inhibited Snail-induced EMT in [102].

The role of Snail and Slug in the drug resistance of tumor cells is associated with a repression of the genes of the pro-apoptotic PUMA, ATM, PTEN, p53, BID, and caspase-6 proteins and de-repression of the genes of the proteins associated with the stemness phenotype [52, 55, 57]. Apart from its anti-apoptotic effect, Snail also increases the expression of ABC transporters, which are the most important mechanism of multidrug resistance [114].

Snail can regulate immune responses. For example, TGF-β induces Snail in macrophages migrating to the inflammation site or wound [81]. Tumors with a high level of Snail expression contain few infiltrating $CD8^+$ cytotoxic T-lymphocytes and an increased amount of pro-tumor M2 macrophages [115]. Snail also induces immunosuppression and immunoresistance through cytokine TSP1 and TGF-β-activated regulatory T cells that reduce the expression of stimulating molecules in dendritic cells, which suppresses cytotoxic T lymphocytes [116].

Cancer stem cells (CSCs) are a small population of cells that are characterized by the expression of stemness markers and pluripotency. CSCs are believed to be a source of tumor heterogeneity. In particular, the tumor clonality maintained by the CSC population is a factor of chemo- and radioresistance. There is evidence that CSCs possess an increased metastatic potential, but the mechanisms of this process are not well understood [117–119]. The stemness regulator SOX2 induced by the vascular endothelial growth factor (VEGF-A) was shown to trigger EMT and metastasis. In BC lines and native tumor cells, VEGF-A activates SOX2 expression, which leads to *SNAI2* induction through miR-452, EMT activation, and increased invasion and metastasis. Thus, VEGF-A stimulates SOX2- and Slug-dependent invasion [120]. Therefore, overexpression of the EMT transcription factor Slug increases the migration activity of CSCs [96].

Activation of the SCF/c-Kit signaling pathway leads to an increase in the Slug level, which causes resistance of ovarian cancer cells to radiotherapy and promotes the survival of CSCs [57]. In addition, SCF/c-Kit/Slug mediates drug resistance in human mesothelioma cells. Knockdown of c-Kit/*KIT* or *SNAI2* increases the sensitivity of mesothelioma cells to the chemotherapeutic agents doxorubicin, paclitaxel, and vincristine. Transfection of c-Kit/*KIT* into mesothelioma cells in the presence of SCF enhances Slug activity and increases resistance to these drugs. Mesothelioma cells with high Slug levels are resistant to drug therapy [121].

Therefore, Snail-family proteins can directly participate in the development of resistance to therapy and suppression of antitumor immunity. These properties of Snail, along with their involvement in EMT, indicate a need for pharmacological inhibition of these proteins.

POTENTIAL OF PHARMACOLOGICAL INHIBITION OF Snail

Signaling pathways involving Snail-family proteins are of interest in the search for new approaches in chemotherapy. Direct pharmacological inhibition is hindered by the complexity involved in targeting the protein's functional domain. However, there have been successful attempts (*Fig. 5*).

Vistain *et al.* [122] proposed the E-box, a Snail-binding site, as a target. A Co(III) complex conjugated to a CAGGTG hexanucleotide was synthesized. After entering the cell, the Co(III)–E-box complex binds to Snail and prevents any interaction with DNA. The developed constructs significantly reduced the invasive potential of tumor cells. The authors hope this compound will be highly efficient as a therapeutic inhibitor of tumor progression and BC metastasis.

The search for a chemical inhibitor of Snail was carried out in [123]. The Snail–p53 complex was chosen as a target. A series of compounds were synthesized, and two leader compounds, GN25 and GN29, increasing the expression of p53 and uncoupling it from Snail were identified. Compounds GN25 and GN29 exhibited selectivity for K-Ras mutated cells and low toxicity for non-tumor cells. However, the effect of these compounds on tumor cells remains ambiguous and their mechanism of action is not well understood. So, it is too early to think about clinical trials of these compounds.

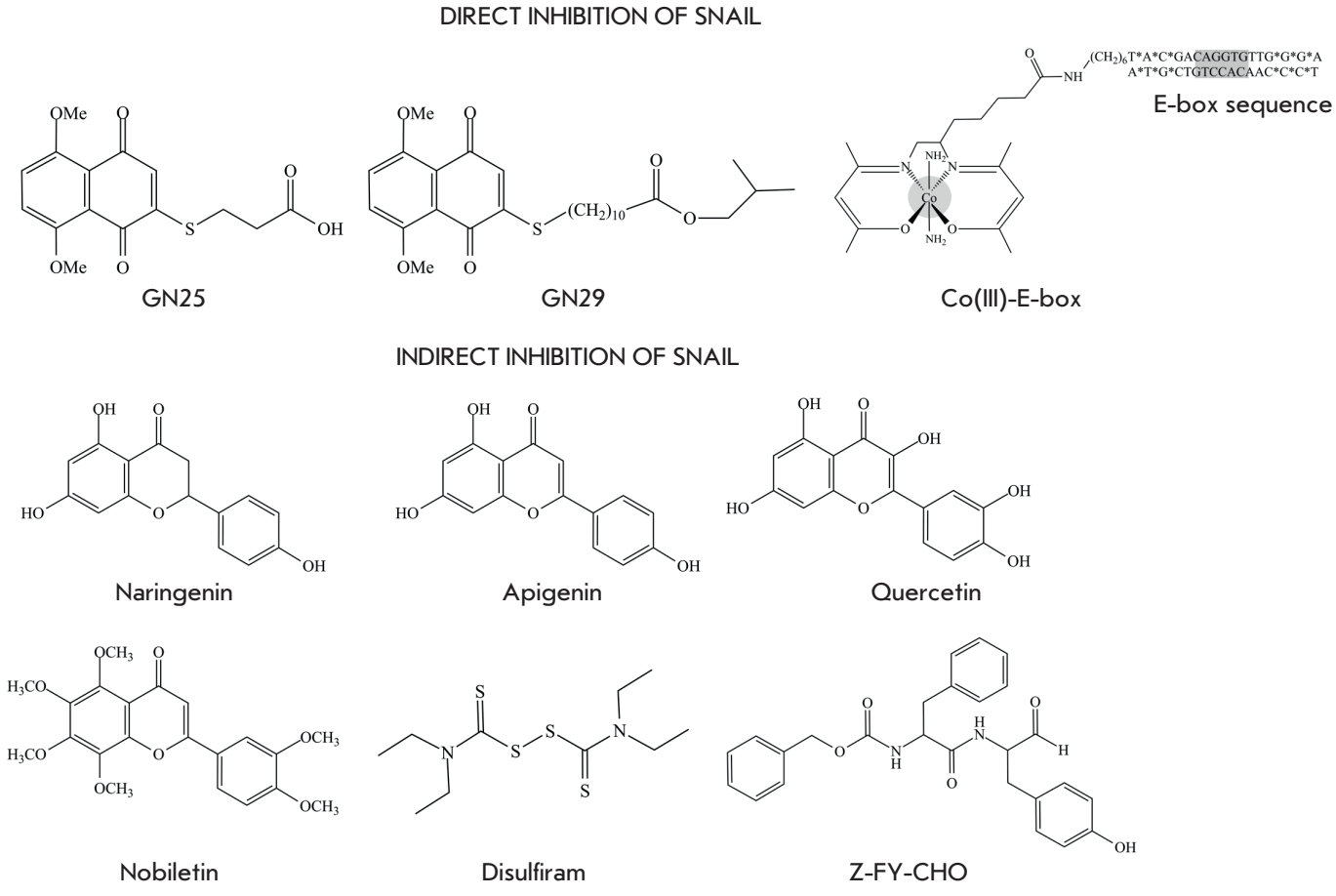


Fig. 5. Pharmacological inhibitors of Snail functions

There are a number of compounds that affect the expression of Snail but are not its direct inhibitors. Disulfiram (DSF), which is used in the treatment of alcohol dependence, inhibits NF-κB. DSF inhibits TGF-β-induced EMT in BC cells, migration and invasion, and growth of tumor grafts. DSF inhibits the ERK/NF-κB/Snail signaling pathway, which leads to MET [124]. DSF is currently under Phase 2 clinical trials to treat patients with stage 4 BC in the Czech Republic. Z-FY-CHO, a selective inhibitor of cathepsin L (ECM component protease), was found to reduce the expression of Snail and trigger MET in prostate cancer cells with a mesenchymal phenotype [125].

MET can be initiated by phytoestrogens that modulate signaling through Snail and Twist1. The flavanone naringenin reduced the invasiveness of prostate cancer cells by blocking Snail and Twist1 [126]. Similar activity was reported for nobiletin, a flavonoid from citrus plants. This compound affects the signaling pathways of TGF-β, ZEB, Slug, and Snail and is capable of suppressing the invasion and migration of tumor cells [127]. The interest of researchers in potential EMT inhibitors of natural origin is justified by the relatively low toxicity of these compounds to non-tumor tissues, as well as by their anticarcinogenic properties [128–130]. Indeed, the flavonoid apigenin exerts an antiproliferative effect on BC cells with mesenchymal traits [131]. This phytoestrogen has been reported to suppress Snail expression, EMT, and cell metastasis [132–134]. Also, the flavonoid quercetin exhibits an antimetastatic effect [135]. Treatment of lung cancer cells with quercetin decreased their invasive and migratory activity. Quercetin affected the Akt-Snail signaling pathway that maintains the survival and metastatic ability of cells. Quercetin is currently under clinical trials as treatment for patients with prostate (phase 2), lung (not specified phase), and kidney (phase 2) cancers. To prevent EMT, it seems relevant to develop compounds that inactivate Snail family proteins and prevent the transactivation of their target genes.

The ability of these compounds to inhibit the functions and activity of Snail suggests that these compounds, after more detailed and thorough investigation of their mechanisms of action, may be included in clini-

cal trials as agents to treat progressive and metastatic tumors.

At the moment, researchers are focused on modifying compounds, finding the best way to deliver them, and developing therapies in combination with other cytotoxic drugs [136].

CONCLUSION

Snail family proteins are key EMT regulators that modulate many ontogenetic and neurobiological processes. A detailed investigation of EMT in tumor cells has revealed the important role played by this process in invasion and metastasis. Snail transcription factors are specific "switches" of the epithelial, more favorable, phenotype of cells to an aggressive prometastatic one. That is why molecular events mediated by these proteins are of interest as targets for therapy of, in particular, resistant metastatic tumors. The development of pharmacological approaches to Snail inhibition is in its infancy. However, chemical classes of synthetic and natural compounds affecting the transcriptional activity and expression of Snail and initiating MEP have already been characterized. Further investigation of EMT and its regulators appears promising for a personalized therapy of tumors. ●

The authors are grateful to A.A. Shtil for discussion of the study and E.A. Varlamova for assistance in preparing the manuscript.
Illustrations were prepared using Servier Medical Art templates (Creative Commons Attribution 3.0 Unported License); https://smart.servier.com.
For preparing the review, we used open access data (https://clinicaltrials.gov/; https://www.uniprot.org/; https://kmplot.com/analysis/).
The study was supported by a grant of the Ministry of Education and Science of the Russian Federation (agreement No. 14.W03.31.0020 with the Institute of Gene Biology, Russian Academy of Sciences).

REFERENCES

1. Heerboth S., Housman G., Leary M., Longacre M., Byler S., Lapinska K., Willbanks A., Sarkar S. // Clin. Transl. Med. 2015. V. 4. P. 1–13.
2. Slattum G., Rosenblatt J. // Cancer. 2014. V. 14. № 7. P. 495–501.
3. Hay E. // Dev. Dyn. 2005. V. 233. № 3. P. 706–720.
4. Wang Y., Shi J., Chai K., Ying X., Zhou B. // Curr. Cancer Drug Targets. 2013. V. 13. № 9. P. 963–972.
5. Murtuza A., Bulbul A., Shen J., Keshavarzian P., Woodward B., Lopez-Diaz F., Lippman S., Husain H. // Cancer Res. 2019. V. 79. № 4. P. 689–698.
6. Vasiliou S., Diamandis E. // Crit. Rev. Clin. Lab. Sci. 2019. V. 56. № 3. P. 200–223.
7. Aubry A., Galiacy S., Allouche M. // Cancers. 2019. V. 11. № 3. P. 275.
8. Ukrainskaya V., Stepanov A., Glagoleva I., Knorre V., Belogurov A., Gabibov A. // Acta Naturae. 2017. V. 9. № 3. P. 55–63.
9. Gonzalez D., Medici D. // Sci. Signal. 2014. V. 7. № 344. P. re8.
10. Peinado H., Olmeda D., Cano A. // Nat. Rev. Cancer. 2007. V. 7. № 6. P. 415–428.
11. Stankic M., Pavlovic S., Chin Y., Brogi E., Padua D., Norton L., Massagué J., Benezra R. // Cell Rep. 2013. V. 5. № 5. P. 1228–1242.
12. Kim N.H., Kim H.S., Li X.Y., Lee I., Choi H.S., Kang S.E., Cha S.Y., Ryu J.K., Yoon D., Fearon E.R., et al. // Cell Biol. 2011. V. 195. № 3. P. 417–433.
13. Siemens H., Jackstadt R., Hunten S., Kaller M., Menssen A., Gotz U., Hermeking H. // Cell Cycle. 2011. V. 10. № 24. P. 4256–4271.
14. Murray S., Gridley T. // Cell Cycle. 2006. V. 5. № 22. P. 2566–2570.
15. Haensel D., Dai X. // Dev. Dyn. Off. Publ. Am. Assoc. Anat. 2018. V. 247. № 3. P. 473–480.
16. Nieto M. // Nat. Rev. Mol. Cell Biol. 2002. V. 3. № 3. P. 155–166.
17. Nishioka R., Itoh S., Gui T., Gai Z., Oikawa K., Kawai M., Tani M., Yamaue H., Muragaki Y. // Exp. Mol. Pathol. 2010. V. 89. № 2. P. 149–157.
18. Veltmaat J., Orelio C., Ward-Van Oostwaard D., van Rooijen M., Mummery C., Defize L.H. // Int. J. Dev. Biol. 2000. V. 44. № 3. P. 297–307.
19. Georgakopoulos-Soares I., Chartoumpekis D., Kyriazopoulou V., Zaravinos A. // Front. Oncol. 2020. V. 10. P. 499.
20. Pastushenko I., Blanpain C. // Trends Cell Biol. 2019. V. 29. № 3. P. 212–226.
21. Katoh M., Katoh M. // Int. J. Mol. Med. 2003. V. 11. № 3. P. 383–388.
22. Phillips S., Kuperwasser C. // Cell Adhes. Migr. 2014. V. 8. № 6. P. 578–587.
23. Molina-Ortiz P., Villarejo A., MacPherson M., Santos V., Montes A., Souchelnytskyi S., Portillo F., Cano A. // PLoS One. 2012. V. 7. № 5. P. e36132.
24. Knorre D., Kudryashova N., Godovikova T. // Acta Naturae. 2009. V. 1. № 3. P. 29–51.
25. Wu Y., Deng J., Rychahou P., Qiu S., Evers B., Zhou B. // Cancer Cell. 2009. V. 15. № 5. P. 416–428.
26. Zhou B., Deng J., Xia W., Xu J., Li Y., Gunduz M., Hung M. // Nat. Cell Biol. 2004. V. 6. № 10. P. 931–940.
27. Liu T., Yu J., Deng M., Yin Y., Zhang H., Luo K., Qin B., Li Y., Wu C., Ren T., et al. // Nat. Commun. 2017. V. 8. № 1. P. 13923.
28. Zhang K., Rodriguez-Aznar E., Yabuta N., Owen R., Mingot J., Nojima H., Nieto M., Longmore G. // EMBO J. 2012. V. 31. № 1. P. 29–43.
29. Chang R., Zhang P., You J. // Open Life Sci. 2016. V. 11. № 1. P. 237–243.
30. Lee J., Jung S., Yang K., Bae E., Ahn S., Park J., Seo D.,

Kim M., Ha J., Lee J., et al. // Nat. Cell Biol. 2017. V. 19. № 10. P. 1260–1273.
31. Carver E., Jiang R., Lan Y., Oram K., Gridley T. // Mol. Cell. Biol. 2001. V. 21. № 23. P. 8184–8188.
32. Jiang R., Lan Y., Norton C., Sundberg J., Gridley T. // Dev. Biol. 1998. V. 198. № 2. P. 277–285.
33. Chen Y., Gridley T. // Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res. 2013. V. 28. № 6. P. 1412–1421.
34. Dale J., Malapert P., Chal J., Vilhais-Neto G., Maroto M., Johnson T., Jayasinghe S., Trainor P., Herrmann B., Pourquié O. // Dev. Cell. 2006. V. 10. № 3. P. 355–366.
35. Arnoux V., Nassour M., L'Helgoualc'h A., Hipskind R., Savagner P. // Mol. Biol. Cell. 2008. V. 19. № 11. P. 4738–4749.
36. Hudson L., Newkirk K., Chandler H., Choi C., Fossey S., Parent A., Kusewitt D. // J. Dermatol. Sci. 2009. V. 56. № 1. P. 19–26.
37. Beyes S., Andrieux G., Schrempp M., Aicher D., Wenzel J., Antón-García V., Boerries M., Hecht A. // Oncogene. 2019. V. 38. P. 6647–6661.
38. Maturi V., Morén A., Enroth S., Heldin C.-H., Moustakas A. // Mol. Oncol. 2018. V. 12. № 7. P. 1153–1174.
39. Strouhalova K., Přechová M., Gandalovičová A., Brábek J., Gregor M., Rosel D. // Cancers. 2020. V. 12. № 1. P. 184. doi: 10.3390/cancers12010184
40. Wang Y., Zhao X., Shuai Z., Li C., Bai Q., Yu X., Wen Q. // Int. J. Clin. Exp. Med. 2015. V. 8. № 5. P. 7388–7393.
41. Wu W., You R., Cheng C., Lee M., Lin T., Hu C. // Sci. Rep. 2017. V. 7. № 1. P. 17753.
42. Haslehurst A.M. The role of the epithelial-mesenchymal transition in aggressive tumour phenotypes. Dissertation. Queen's University. Kingston, Ontario, Canada, 2014.
43. LaBonne C., Bronner-Fraser M. // Dev. Biol. 2000. V. 221. № 1. P. 195–205.
44. Batlle E., Sancho E., Francí C., Domínguez D., Monfar M., Baulida J., García De Herreros A. // Nat. Cell. Biol. 2000. V. 2. № 2. P. 84–89.
45. Shields M.A., Krantz S.B., Bentrem D.J., Dangi-Garimella S., Munshi H.G. // J. Biol. Chem. 2012. V. 287. № 9. P. 6218–6229.
46. Shields M.A., Dangi-Garimella S., Krantz S.B., Bentrem D.J., Munshi H.G. // J. Biol. Chem. 2011. V. 286. № 12. P. 10495–10504.
47. del Barrio M.G., Nieto M.A. // Dev. Camb. Engl. 2002. V. 129. № 7. P. 1583–1593.
48. Fan M., Xu Y., Hong F., Gao X., Xin G., Hong H., Dong L., Zhao X // Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol. 2016. V. 38. № 4. P. 1319–1332.
49. Ungefroren H., Witte D., Lehnert H. // Dev. Dyn. Off. Publ. Am. Assoc. Anat. 2018. V. 247. № 3. P. 451–461.
50. Faure S., Fort P. // Small GTPases. 2011. V. 2. № 6. P. 310–313.
51. Li Y., Zhou C., Gao Y. // Biochem. Biophys. Res. Commun. 2014. V. 452. № 3. P. 490–496.
52. Vega S., Morales A., Ocaña O.H., Valdés F., Fabregat I., Niet M.A. // Genes Dev. 2004. V. 18. № 10. P. 1131–1143.
53. Mittal M.K., Singh K., Misra S., Chaudhuri G. // J. Biol. Chem. 2011. V. 286. № 1. P. 469–479.
54. Assani G., Zhou Y. // Oncol. Lett. 2018. V. 17. № 1. P. 23–30
55. Kajita M., McClinic K.N., Wade P.A. // Mol. Cell. Biol. 2004. V. 24. № 17. P. 7559–7566.
56. Escrivà M., Peiró S., Herranz N., Villagrasa P., Dave N., Montserrat-Sentís B., Murray S.A., Francí C., Gridley T., Virtanen I., et al. // Mol. Cell. Biol. 2008. V. 28. № 5. P. 1528–1540.
57. Kurrey N.K., Jalgaonkar S.P., Joglekar A.V., Ghanate A.D., Chaskar P., Doiphode R.Y., Bapat S.A. // Stem Cells Dayt. Ohio. 2009. V. 27. № 9. P. 2059–2068.
58. Lee S., Lee S, Jung Y., Xu Y., Kang H.S., Ha N., Park B. // Neoplasia. 2009. V. 11. № 1. P. 22–31.
59. Khamidullina A.I., Yastrebova M.A., Scherbakov A.M., Tatarskiy V.V. // Ann. Oncol. 2019. V. 30. Sup. 5. P. v24.
60. Gavert N., Ben-Ze'ev A. // Trends Mol. Med. 2008. V. 14. № 5. P. 199–209.
61. Piedra M.E., Ros M.A. // Dev. Camb. Engl. 2002. V. 129. № 14. P. 3431–3440.
62. Gotzmann J., Mikula M., Eger A., Schulte-Hermann R., Foisner R., Beug H., Mikulits W. // Mutat. Res. 2004. V. 566. № 1. P. 9–20.
63. Wu X., Cai J., Zuo Z., Li J. // Biomed. Pharmacother. 2019. V. 114. P. 108708.
64. Liu J.-Y., Jiang L., He T., Liu J., Fan J., Xu X., Tang B., Shi Y., Zhao Y.-L., Qian F., et al. // Cell Death Dis. 2019. V. 10. № 3. P. 162.
65. Li J., Xu H., Wang Q., Wang S., Xiong N. // Cancer Med. 2019. V. 8. № 2. P. 783–794.
66. Massagué J. // Cell. 2008. V. 134. № 2. P. 215–230.
67. Vincent T., Neve E., Johnson J.R., Kukalev A., Rojo F., Albanell J., Pietras K., Virtanen I., Philipson L., Leopold P.L., et al. // Nat. Cell Biol. 2009. V. 11. № 8. P. 943–950.
68. Cao Y., Wan G.X., Sun J.P., Cui X.B., Hu J.M., Liang W.H., Zheng Y.Q., Li W.Q., Li F. // Kaohsiung J. Med. Sci. 2015. V. 31. № 2. P. 70–76.
69. Leong K.G., Niessen K., Kulic I., Raouf A., Eaves C., Pollet I., Karsan A. // J. Exp. Med. 2007. V. 204. № 12. P. 2935–2948.
70. Julien S., Puig I., Caretti E., Bonaventure J., Nelles L., van Roy F., Dargemont C., de Herreros A.G., Bellacosa A., Larue L. // Oncogene. 2007. V. 26. № 53. P. 7445–7456.
71. Scherbakov A., Andreeva O., Shatskaya V., Krasil'nikov M. // J. Cell. Biochem. 2012. V. 113. № 6. P. 2147–2155.
72. Stemmer V., de Craene B., Berx G., Behrens J. // Oncogene 2008. V. 27. № 37. P. 5075–5080.
73. Lu X., Yan C., Huang Y., Shi D., Fu Z., Qiu J., Yin Y. // Oncotarget. 2016. V. 7. № 24. P. 37177–37191.
74. Al Tameemi W., Dale T.P., Al-Jumaily R., Forsyth N.R. // Front. Cell Dev. Biol. 2019. V. 7. P. 4.
75. Semenza G.L. // Nat. Rev. Cancer 2003. V. 3. № 10. P. 721–732.
76. Yeo C.D., Kang N., Choi S.Y., Kim B.N., Park C.K., Kim J.W., Kim Y.K., Kim S.J. // Korean J. Intern. Med. 2017. V. 32. № 4. P. 589–599.
77. Hill R.P., Marie-Egyptienne D.T., Hedley D.W. // Semin. Radiat. Oncol. 2009. V. 19. № 2. P. 106–111.
78. Kim W.Y., Perera S., Zhou B., Carretero J., Yeh J.J., Heathcote S.A., Jackson A.L., Nikolinakos P., Ospina B., Naumov G., et.al. // J. Clin. Invest. 2009. V. 119. № 8. P. 2160–2170.
79. Krishnamachary B., Zagzag D., Nagasawa H., Rainey K., Okuyama H., Baek J.H., Semenza G.L. // Cancer Res. 2006. V. 66. № 5. P. 2725–2731.
80. Erler J.T., Bennewith K.L., Cox T.R., Lang G., Bird D., Koong A., Le Q.T., Giaccia A.J. // Cancer Cell. 2009. V. 15. № 1. P. 35–44.
81. Yang S.W., Zhang Z.G., Hao Y.X., Zhao Y.L., Qian F., Shi Y., Li P.-A., Liu C.Y., Yu P.W. // Oncotarget. 2017. V. 8. № 6. P. 9535–9545.
82. Scherbakov A., Stefanova L., Sorokin D., Semina S., Berstein L., Krasil'nikov M. // Exp. Cell Res. 2013. V. 319. № 20. P. 3150–3159.
83. Yastrebova M., Khamidullina A., Scherbakov A., Ta-

tarskiy V. Transcription factor Snail leads to resistance of breast cancer cells to hypoxia. // Proc. 44th FEBS congress. july 6–11. 2019. Krakow, Poland.
84. Grubben C., Fryns J.P., De Zegher F., Van Den Berghe H. // Genet. Couns. Geneva Switz. 1990. V. 1. № 2. P. 103–109.
85. Shirley S.H., Greene V.R., Duncan L.M., Torres Cabala C., Grimm E.A., Kusewitt D.F. // Am. J. Pathol. 2012. V. 180. № 6. P. 2479–2489.
86. Zhang T., Chen X.U., Chu X., Shen Y.I., Jiao W., Wei Y., Qiu T., Yan G., Wang X., Xu L. // Oncol. Lett. 2016. V. 11. № 1. P. 306–310.
87. Nagy Á., Lánczky A., Menyhárt O., Győrffy B. // Sci. Rep. 2018. V. 8. № 1. P. 9227.
88. Lira R.P.C., Antunes-Foschini R., Rocha E.M. // Arq. Bras. Oftalmol. 2020. V. 83. № 2. P. 5–7.
89. Yang H., Hu H., Gou Y., Hu Y., Li H., Zhao H., Wang B., Li P., Zhang Z. // Int. J. Clin. Oncol. 2018. V. 23. № 2. P. 321–328.
90. Gong X., Tao Y., Zhou L., Yu L., Wu S., Song W., Wang D., Cheng Z. // J. Southern Med. Univ. 2015. V. 35. № 12. P. 1733–1738.
91. Choi J.E., Bae J.S., Kang M.J., Chung M.J., Jang K.Y., Park H.S., Moon W.S. // Oncol. Rep. 2017. V. 38. № 3. P. 1695–1705.
92. Tanaka Y., Kawaguchi Y.T., Fujiwara S., Yoo S., Tsunetoh S., Takai M., Kanemura M., Tanabe A., Ohmichi M. // Cancer Biol. Ther. 2013. V. 14. № 1. P. 13–19.
93. Kihara A., Wakana K., Kubota T., Kitagawa M. // Histopathology. 2016. V. 69. № 3. P. 374–382.
94. Scherbakov A., Gershtein E., Korotkova E., Ovchinnikova L., Ovsii O., Ermilova V., Gens G., Kushlinskii N. // Bull. Exp. Biol. Med. 2016. V. 160. № 6. P. 802–806.
95. Noh H.S., Hah Y.S., Ha J.H., Kang M.Y., Zada S., Rha S.Y., Kang S.S., Kim H.J., Park J.Y., Byun J.H., et al. // Oncotarget. 2016. V. 7. № 4. P. 4632–4646.
96. Feng X., Zhao L., Shen H., Liu X., Yang Y., Lv S., Niu Y. // Oncotarget. 2017. V. 8. № 20. P. 33365–33374.
97. Okubo K., Uenosono Y., Arigami T., Yanagita S., Matsushita D., Kijima T., Amatatsu M., Uchikado Y., Kijima Y., Maemura K., et al. // Gastric Cancer Off. J. Int. Gastric Cancer Assoc. Jpn. Gastric Cancer Assoc. 2017. V. 20. № 5. P. 802–810.
98. Yu Q., Zhang K., Wang X., Liu X., Zhang Z. // J. Exp. Clin. Cancer Res. CR. 2010. V. 29. P. 119.
99. Tian Y., Qi P., Niu Q., Hu X. // Front. Mol. Biosci. 2020. V. 7. P. 22.
100. Huang D., Duan H., Huang H., Tong X., Han Y., Ru G., Qu L., Shou C., Zhao Z. // Sci. Rep. 2016. V. 6. P. 20502.
101. Desai K., Aiyappa R., Prabhu J.S., Nair M.G., Lawrence P.V., Korlimarla A., Ce A., Alexander A., Kaluve R.S., Manjunath S., et al. // Tumour Biol. 2017. V. 39. № 3. P. 1010428317695028. Doi: 10.1177/1010428317695028.
102. Moody S.E., Perez D., Pan T., Sarkisian C.J., Portocarrero C.P., Sterner C.J., Notorfrancesco K.L., Cardiff R.D., Chodosh L.A. // Cancer Cell. 2005. V. 8. № 3. P. 197–209.
103. van Nes J.G., de Kruijf E.M., Putter H., Faratian D., Munro A., Campbell F., Smit V., Liefers G.J., Kuppen P., van de Velde C.J.H., et al. // Breast Cancer Res. Treat. 2012. V. 133 (1). P. 49–59.
104. Du B., Shim J.S. // Molecules. 2016. V. 21. № 7. P. 965. Doi: 10.3390/molecules21070965.
105. Cao L., Wan Q., Li F., Tang C.E. // BMB Rep. 2018. V. 51. № 9. P. 456–461.
106. Zheng X., Carstens J.L., Kim J., Scheible M., Kaye J., Sugimoto H., Wu C.C., LeBleu V.S., Kalluri R. // Nature. 2015. V. 527. № 7579. P. 525–530.
107. Olmeda D., Moreno-Bueno G., Flores J.M., Fabra A., Portillo F., Cano A. // Cancer Res. 2007. V. 67. № 24. P. 11721–11731.
108. Cañadas I., Rojo F., Taus Á., Arpí O., Arumí-Uría M., Pijuan L., Menéndez S., Zazo S., Dómine M., Salido M., et al. // Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2014. V. 20. № 4. P. 938–950.
109. Sonego M., Pellarin I., Costa A., Vinciguerra G.L., Coan M., Kraut A., D'Andrea S., Dall'Acqua A., Castillo-Tong D.C., Califano D., et al. // Sci. Adv. 2019. V. 5. № 5. P. eaav3235.
110. Lambies G., Miceli M., Martínez-Guillamon C., Olivera-Salguero R., Peña R., Frías C.P., Calderón I., Atanassov B.S., Dent S.Y., Arribas J., et al. // Cancer Res. 2019. V. 79. № 1. P. 33–46.
111. Mariano G., Ricciardi M.R., Trisciuoglio D., Zampieri M., Ciccarone F., Guastafierro T., Calabrese R., Valentini E., Tafuri A., Del Bufalo D., et al. // Oncotarget 2015. V. 6. № 17. P. 15008–15021.
112. Karicheva O., Rodriguez-Vargas J.M., Wadier N., Martin-Hernandez K., Vauchelles R., Magroun N., Tissier A., Schreiber V., Dantzer F. // Oncotarget. 2016. V. 7. № 39. P. 64109–64123.
113. Chang T.H., Tsai M.F., Su K.Y., Wu S.G., Huang C.P., Yu S.L., Yu Y.L., Lan C.C., Yang C.H., et al. // Am. J. Respir. Crit. Care Med. 2011. V. 183. № 8. P. 1071–1079.
114. Saxena M., Stephens M. A., Pathak H., Rangarajan A. // Cell Death Dis. 2011. V. 2. P. 179.
115. Dongre A., Rashidian M., Reinhardt F., Bagnato A., Keckesova Z., Ploegh H.L., Weinberg R.A. // Cancer Res. 2017. V. 77. № 15. P. 3982–3989.
116. Kudo-Saito C., Shirako H., Takeuchi T., Kawakami Y. // Cancer Cell. 2009. V. 15. № 3. P. 195–206.
117. Lee K.S., Choi J.S., Cho Y.W. // Biochem. Biophys. Res. Commun. 2019. V. 512. № 3. P. 511–516.
118. Srivastava A.K., Banerjee A., Cui T., Han C., Cai S., Liu L., Wu D., Cui R., Li Z., Zhang X., et al. // Cancer Res. 2019. V. 79. № 9. P. 2314–2326.
119. Izumi D., Toden S., Ureta E., Ishimoto T., Baba H., Goel A. // Cell Death Dis. 2019. V. 10. № 4. P. 267.
120. Kim M., Jang K., Miller P., Picon-Ruiz M., Yeasky T.M., El-Ashry D., Slingerland, J.M. // Oncogene. 2017. V. 36. № 36. P. 5199–5211.
121. Catalano A., Rodilossi S., Rippo M.R., Caprari P., Procopio A. // J. Biol. Chem. 2004. V. 279. № 45. P. 46706–46714.
122. Vistain L.F., Yamamoto N., Rathore R., Cha P., Meade T.J. // Chembiochem. Eur. J. Chem. Biol. 2015. V. 16. № 14. P. 2065–2072.
123. Lee S.H., Shen G.N., Jung Y.S., Lee S.J., Chung J.Y., Kim H.S., Xu Y., Choi Y., Lee J.W., Ha N.C., et al. // Oncogene. 2010. V. 29. № 32. P. 4576–4587.
124. Han D., Wu G., Chang C., Zhu F., Xiao Y., Li Q., Zhang T., Zhang L. // Oncotarget. 2015. V. 6. № 38. P. 40907–40919.
125. Burton L.J., Dougan J., Jones J., Smith B.N., Randle D., Henderson V., Odero-Marah V.A. // Mol. Cell. Biol. 2017. V. 37. № 5. P. e00297-16.
126. Han K.Y., Chen P.N., Hong M.C., Hseu Y.C., Chen K.M., Hsu L.S., Chen W.J. // Anticancer Res. 2018. V. 38. № 12. P. 6753–6758.
127. Ashrafizadeh M., Zarrabi A., Saberifar S., Hashemi F., Hushmandi K., Hashemi F., Moghadam E.R., Mohammadinejad R., Najafi M., Garg M. // Biomedicines. 2020. V. 8. № 5. P. 110.
128. Lamartiniere C.A., Wang J., Smith-Johnson M., Eltoum I.E. // Toxicol. Sci. Off. J. Soc. Toxicol. 2002. V. 65. № 2. P. 228–238.

129. Martinović L.S., Peršurić Ž., Pavelić K. // Molecules. 2020. V. 25. № 9. P. 2222.
130. Amawi H., Ashby C.R., Samuel T., Peraman R., Tiwari A.K. // Nutrients. 2017. V. 9. № 8. P. 911.
131. Scherbakov A., Andreeva O. // Acta Naturae. 2015. V. 7. № 3. P. 133–139.
132. Qin Y., Zhao D., Zhou H., Wang X.H., Zhong W.L., Chen S., Gu W.G., Wang W., Zhang C.H., Liu Y.R., et al. // Oncotarget. 2016. V. 7. № 27. P. 41421–41431.
133. Chang J.H., Cheng C.W., Yang Y.C., Chen W.S., Hung W.Y., Chow J.M., Chen P.S., Hsiao M., Lee W.J., Chien M.H. // Exp. Clin. Cancer Res. CR. 2018. V. 37. № 1. P. 199.
134. Erdogan S., Doganlar O., Doganlar Z.B., Serttas R., Turkekul K., Dibirdik I., Bilir A. // Life Sci. 2016. V. 162. P. 77–86.
135. Chang J.H., Lai S.L., Chen W.S., Hung W.Y., Chow J., Hsiao M., Lee W., Chien M. // Biochim. Biophys. Acta BBA – Mol. Cell Res. 2017. V. 1864. № 10. P. 1746–1758.
136. Kothari A.N., Mi Z., Zapf M., Kuo P.C. // Clin. Transl. Med. 2014. V. 3. P. 35.

Mechanisms of Action of the PGLYRP1/Tag7 Protein in Innate and Acquired Immunity

D. V. Yashin*, L. P. Sashchenko, G. P. Georgiev
Institute of Gene Biology RAS, Moscow, 119334 Russia
*E-mail: yashin_co@mail.ru
Received July 29, 2020; in final form, October 19, 2020
DOI: 10.32607/actanaturae.11102

ABSTRACT **One of the promising fields of modern molecular biology is the search for new proteins that regulate the various stages of the immune response and the investigation of the molecular mechanisms of action of these proteins. Such proteins include the multifunctional protein PGLYRP1/Tag7, belonging to the PGRP-S protein family, whose gene was discovered in mice at the Institute of Gene Biology, Russian Academy of Sciences, in 1996. PGLYRP1/Tag7 is classified as a protein of innate immunity; however, it can also participate in the regulation of acquired immunity mechanisms. In this paper, we consider the involvement of PGLYRP1/Tag7 in the triggering of antimicrobial defense mechanisms and formation of subsets of cytotoxic lymphocytes that kill tumor cells. The paper emphasizes that the multifaceted functional activity of Tag7 in the immune response has to do with its ability to interact with various proteins to form stable protein complexes. Hsp70-associated Tag7 can induce the death of tumor cells carrying the TNFR1 receptor. Tag7, associated with the Mts1 (S100A4) protein, can stimulate the migration of innate and adaptive immune cytotoxic lymphocytes to a lesion site. Involvement of Tag7 in the regulation of immunological processes suggests that it may be considered as a promising agent in cancer therapy. These properties of Tag7 were used to develop autologous vaccines that have passed the first and second phases of clinical trials in patients with end-stage melanoma and renal cancer. The C-terminal peptide of Tag7, isolated by limited proteolysis, was shown to protect the cartilage and bone tissue of the ankle joint in mice with induced autoimmune arthritis and may be a promising drug for suppressing the development of inflammatory processes.**
KEYWORDS **PGLYRP1/Tag7, cytotoxicity, antimicrobial effect, antitumor therapy, Mts1, Hsp70, HspBP1, TNFR1.**
ABBREVIATIONS **Tag7 – tumor antagonistic gene 7 protein product; PGRP – peptidoglycan recognition protein; FCA – Freud's complete adjuvant; Mts1 – metastasin 1; Hsp70 – major 70-kDa heat shock protein.**

INTRODUCTION

Understanding the mechanisms of activation and action of regulatory and effector lymphocytes is necessary in order to identify the pathways of host defense against cells with foreign or modified antigens. Understanding these processes is important in anti-inflammatory and anticancer therapy. In this regard, one of the areas of focus in modern immunology is the investigation of the proteins involved in innate and adaptive immunity, which allows for a deeper understanding of the immune response principles and the causes of the dysfunctions in various pathologies. The search for new proteins that regulate the activity of the cells involved in the immune response and the investigation of the molecular mechanisms underlying the action of these proteins seem promising. PGLYRP1/Tag7 is one such protein.

The gene for this protein was discovered in mice by subtracting cDNA libraries obtained from metastatic and non-metastatic mouse tumor cell lines at the Institute of Gene Biology, Russian Academy of Sciences, in the laboratory headed by one of the authors (Georgy Pavlovich Georgiev), in 1996. The protein was given the working name Tag7 [1]. Tag7 turned out to be playing an important role in antitumor defense [2]: so, its name signifies the tumor antagonistic gene protein product.

In 1998, Kang *et al.* [3] found a gene in the insect hemolymph whose structure was highly homologous to that of the *tag7* gene. The product of this gene was shown to bind to peptidoglycans of the bacterial cell wall and was termed the peptidoglycan recognition protein (PGRP) [3]. The structure of a mouse *PGRP* homologue is identical to that of the previously described *tag7* gene [4, 5], which means that Tag7 and

PGRP are the same protein. Later, when the gene family was discovered, the term PGRP was changed to PGRP-S (where S stands for "small").

Further functional studies of Tag7/PGRP-S were conducted in two directions. While European and U.S. researchers have focused on the role of Tag7/PGRP-S in innate antimicrobial immunity, we have mainly studied the mechanisms of antitumor action of this protein and related issues (Institute of Biochemistry, Russian Academy of Sciences).

PGRP/Tag7 PROTEIN FAMILY

The Tag7/PGRP-S protein belongs to a small protein family. Members of this family differ in their transcript lengths: extracellular PGRP-S (short form) [1, 3], long transmembrane PGRP-L [5–7], and intermediate PGRP-I [6]. Structural studies revealed a highly conserved region of 160 amino acid residues at the C-terminus of all the proteins of this family. This region contains three adjacent PGRP domains connected by segments with a conserved amino acid sequence [6]. Only PGRP-S has a signal peptide in front of the PGRP domain, indicating that PGRP-S can be secreted by the cell [5].

In humans, there are four PGRP family proteins designated as PGLYRP1, PGLYRP2, PGLYRP3, and PGLYRP4. The first one corresponds to Tag7/PGRP-S [8]. It consists of 196 amino acids and has a signal peptide and a PGRP domain. Its gene is expressed in the bone marrow, thymus, fetal liver, polymorphonuclear leukocytes, lymphoid cells of the duodenum, spleen, and lymph nodes, alveolar epithelium, and pulmonary endothelium [6, 9].

Analysis of the crystal structure of PGRP proteins revealed a ligand-binding site recognizing a specific peptidoglycan sequence. Also, there was a protein–protein recognition site formed by a unique hydrophobic groove and the conserved amino acid residues Leu65, Arg18, Thr90, Glu93, Phe94, and Leu133 [10].

The PGLYRP1/Tag7 structure determines its functional activity. The protein can participate in antimicrobial defense activation by binding to peptidoglycan. The protein–protein interaction sites are responsible for the association of PGLYRP1/Tag7 with other proteins, which is followed by the formation of the stable complexes involved in immune response triggering. PGLYRP1/Tag7 is usually referred to as an innate immunity protein, which is not entirely true (see below). Its involvement in the regulation of the immune defense has been extensively studied. There exist three main areas of investigation of the PGLYRP1/Tag7 functional activity: (1) participation of PGLYRP1/Tag7 in antimicrobial defense; (2) role of Tag7 in human lymphocyte activation; (3) use of Tag7 in antitumor therapy. This article discusses in detail these areas characterizing PGLYRP1/Tag7 as an active immune response regulator.

INVOLVEMENT OF PGRP FAMILY PROTEINS IN INNATE ANTIMICROBIAL IMMUNITY IN INSECTS

Insect PGRP proteins can induce an antimicrobial immune response through either the Toll receptor or the Imd pathway [11–13].

After peptidoglycan recognition, insect PGRP proteins interact with Grass serine protease that initiates a proteolytic cascade, leading to cleavage of the Spatzle protein. One of the resulting fragments, Spatzle, forms a homodimer, causing dimerization and activation of the Toll receptor that further induces an antimicrobial response [14]. The PGRP-L protein interacts with Imd upon activation of the Toll-independent immune response pathway. Imd, in turn, induces a second signaling pathway, also resulting in the secretion of antimicrobial peptides [11, 15–17].

INVOLVEMENT OF PGRP FAMILY PROTEINS IN INNATE ANTIMICROBIAL IMMUNITY IN MAMMALS

All four PGRP family members in humans and other mammals are soluble secreted proteins possessing both recognition and effector functions [18, 19]. PGLYRP1, PGLYRP3, and PGLYRP4 can directly lyse both gram-positive and gram-negative bacteria [20–23]. PGLYRP3 is a peptidoglycan amidase [24, 25].

Each of these proteins contains one or two PGRP domains with a binding site specific for a muramyl peptide fragment of bacterial peptidoglycan [18, 19]. In addition, PGRPs can interact with lipoteichoic acid and lipopolysaccharide [22, 26]. Thus, PGRPs interact with the entire outer membrane of gram-negative bacteria [27].

PGRP uses three cytotoxic mechanisms to lyse bacteria. Firstly, PGRP induces oxidative stress because of increased formation of hydrogen peroxide (H_2O_2) and hydroxyl radicals ($HO^{\cdot}$) [27, 28]. Secondly, PGRP triggers thiol stress, leading to the depletion of more than 90% of intracellular thiols. The third antibacterial effect is metal stress that results in increased concentrations of intracellular Zn^{2+} and Cu^{+} ions [27, 28]. Each stress response alone has only a bacteriostatic effect, while combined induction of all three stress responses simultaneously exerts a bactericidal effect [27].

The antimicrobial effect of PGRP is enhanced through cooperation with innate immune cells. For instance, during the phagocytosis of bacteria, phagocytic cells pump not only oxygen radicals, but also Cu^{+} and Zn^{2+} ions into phagolysosomes to enhance the antimicrobial effect [29, 30]. In response, bacteria increase the

expression of Cu^{+} and Zn^{2+} ion exporters [27]. PGRP proteins prevent these changes by promoting bacterial lysis [28]. PGRPs were also shown to act synergistically with antimicrobial peptides [31]. Also, PGRP-S was shown to interact with the innate immune receptor TREM-1 that triggers a pro-inflammatory immune response. This interaction will be discussed below. Synergistic interaction with other host defense mechanisms further enhances the antimicrobial efficacy of PGRP and prevents the development of resistance, thus making PGRP an important component of the innate antimicrobial immunity.

THE Hsp70–PGLYRP1/Tag7 COMPLEX KILLS VARIOUS TUMOR CELL TYPES

The first studies showed that a conditioned medium of VMR-0 tumor cells transfected with a Tag7-encoding construct has a cytotoxic effect on VMR-0 cells. Antibodies to Tag7 neutralize this effect, indicating that Tag7 is cytotoxic [2].

However, another group of researchers demonstrated that PGRP-S expressed in *Escherichia coli* cells has no cytotoxic activity [4].

Later, Tag7 produced in a yeast system was also found to lack any toxic effect. However, it can form a stable equimolar complex with the major heat shock protein Hsp70, which is highly cytotoxic [32]. The Tag7–Hsp70 complex at a concentration of 10^{-10} M can induce cell death in a wide range of tumor cell lines.

Two Hsp70 domains are required to form a stable complex. Tag7 can bind to the peptide-binding domain of Hsp70 and even to the 14-mer peptide of this domain, which is located on the tumor cell surface and plays an essential role in NK cell activation [33]. However, complexes of Tag7 with these Hsp70 fragments show low cytotoxic activity, and the presence of the Hsp70 ATP-binding domain leads to the formation of a highly active cytotoxic complex [32].

COS-1 cells transfected with *tag7* were shown to release the Tag7–Hsp70 complex, which kills tumor cells, into a conditioned medium. The complex is secreted via the Golgi apparatus [32]. Apparently, VMR-0 cells transfected with *tag7* also secrete the Tag7–Hsp70 complex, which explains the Tag7-dependent cytotoxic activity of a conditioned medium of these cells.

An intratumoral injection of the Tag7–Hsp70 complex was shown to inhibit tumor growth. For instance, administration of the Tag7–Hsp70 complex to mice subcutaneously inoculated with aggressive M3 melanoma cells suppressed tumor growth and increased animals' life span more than two-fold [34].

LAK cells obtained by 6-day cultivation with cytokine IL-2 released the Tag7–Hsp70 cytotoxic complex into a conditioned medium after incubation with target tumor cells. A Golgi apparatus inhibitor suppressed the secretion of this complex by lymphocytes [32].

INTERACTION OF THE Tag7–Hsp70 CYTOTOXIC COMPLEX WITH THE TNFR1 RECEPTOR INDUCES INTRACELLULAR CELL DEATH SIGNALS

A detailed study of the cytotoxic effect of this complex showed that different cells in heterogeneous tumor cell cultures died at different time intervals, and that different cell death mechanisms were induced in the cells. The cells incubated with the Tag7–Hsp70 complex underwent apoptotic death 3 h after incubation, while RIP1 kinase-mediated necroptosis was activated in them only 20 h later [35].

Both cytolytic processes are induced upon interaction of Tag7–Hsp70 with the same cellular receptor, TNFR1, which is specific to cytokine TNF-α.

TNFR1 is a member of the death receptor family; it can induce alternative cytotoxic pathways of programmed cell death: caspase-dependent apoptosis and RIP1-kinase-dependent necroptosis [36, 37]. Necroptosis pathways are induced in tumor cells with suppressed caspase activity through any of the pathways [38].

Tag7–Hsp70 binds to TNFR1 on the plasma membrane of tumor cells and interacts with its extracellular domain (sTNFR1) both in solution and on an affinity column. Antibodies to TNFR1 suppress this process in all cases. The Tag7–Hsp70 complex may be considered as a new ligand for the TNFR1 receptor, which induces various apoptotic and necroptotic pathways in tumor cells [35].

In apoptotic cell death, the cytotoxic effect of the Tag7–Hsp70 complex has to do with sequential activation of caspase-8 and caspase-3. No intracellular apoptosis mechanisms involving mitochondria and caspase-9 are activated [35].

Necroptosis begins with necrosome formation, mediated by RIP1 and RIP3 kinases. The cytotoxic signal is further transmitted to cellular organelles: lysosomes and mitochondria. Accumulation of reactive oxygen species on mitochondrial membranes plays a key role in necroptotic cell death. There exists a relationship between lysosome activation and mitochondria. Inhibition of the catalytic activity of lysosomal cathepsins released into the intracellular space hinders both changes in the mitochondrial membrane potential and the accumulation of reactive oxygen species [39] (*Fig. 1*).

Tag7 and Hsp70 play different roles in the activation of cytotoxic pathways. Activation of a cytotoxic signal is known to be a two-stage process. At the first stage, a cytotoxic ligand binds to the receptor's extracellular domain. At the second stage, the TNFR1 intracellular

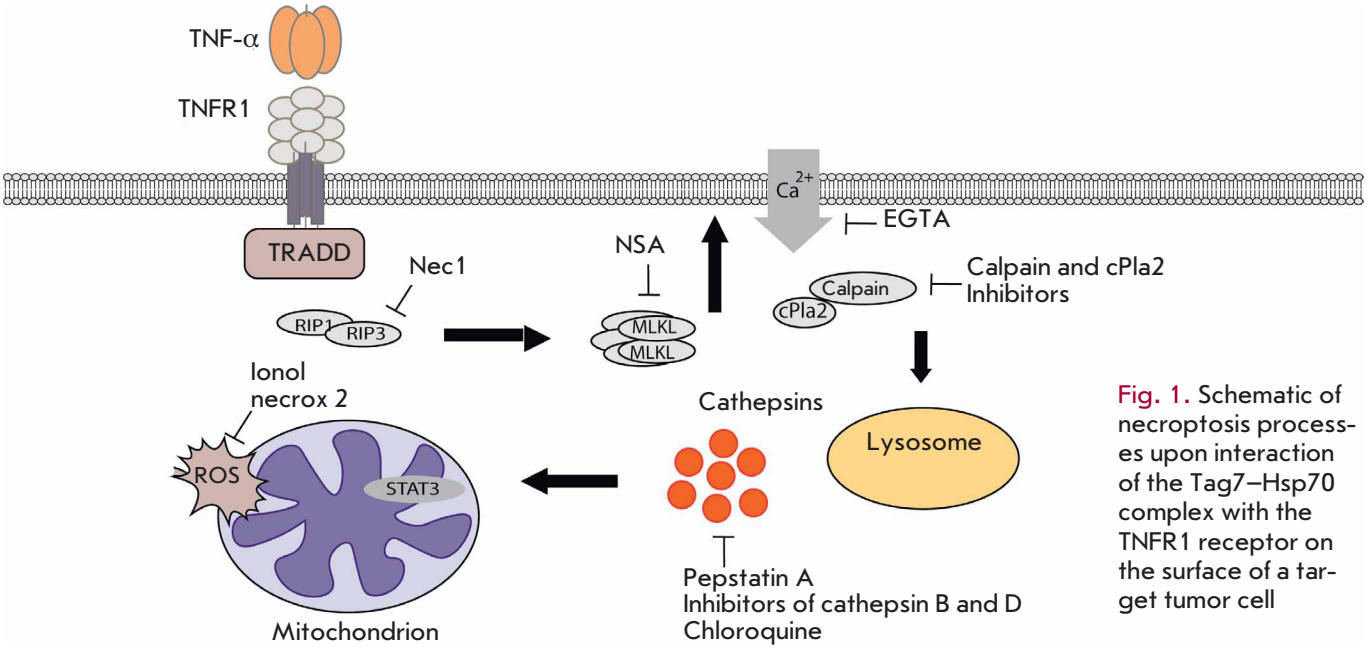


Fig. 1. Schematic of necroptosis processes upon interaction of the Tag7–Hsp70 complex with the TNFR1 receptor on the surface of a target tumor cell

domain changes its structure to form the death domain that activates intracellular cytotoxic processes [40].

A necessary condition for death domain formation is the trimerization of the TNFR1 receptor [40]. In the absence of Hsp70, Tag7 is able to bind to TNFR1 as a monomer, but unable to induce receptor trimerization on the cell surface and, hence, trigger cell death. Tag7 inhibits the cytotoxic effect of both TNF-α and the Tag7–Hsp70 complex by competing with cytotoxic ligands for the TNFR1-binding site. Hsp70 cannot bind to TNFR1, but its interaction with Tag7 is necessary to induce cytotoxicity [35].

A 12-mer peptide at the Tag7 C-terminus was isolated by limited trypsinolysis. This peptide can bind to the TNFR1 receptor both in solution and on the cell surface [41]. The peptide was designated as "17.1" when first obtained by synthesis. Like the full-length Tag7, the 17.1 peptide did not induce cell death but inhibited the cytotoxic activity of both TNF-α and Tag7–Hsp70 [41]. Interestingly, the same peptide can interact with the heat shock protein Hsp70 and form the 17.1–Hsp70 cytotoxic complex that induces cell death [41].

The 17.1 peptide inhibited the functional activity of TNF-α not only in a cell model, but also in a mouse model. The anti-inflammatory effect of the 17.1 peptide was studied in a model of autoimmune arthritis induced by Freund's complete adjuvant (FCA) stimulating tissue TNF-α secretion. This peptide was found to protect the cartilage and bone tissue of the ankle joint in mice [41]. We suggest that the 17.1 peptide may be a promising agent for preventing inflammatory processes.

METASTATIC PROTEIN Mts1/S100A4 DESTROYS THE Tag7–Hsp70 COMPLEX

Many metastatic cancer cell lines are insensitive to the effect of Tag7–Hsp70. One of the key metastasis-stimulating proteins, metastasin 1 (Mts1/S100A4), belongs to the S100 family of Ca^{2+}-binding proteins [42, 43]. Mts1 can form stable complexes with both Tag7 and Hsp70. Interestingly, Mts1 binds to the same region in the Tag7 protein as Hsp70. When Mts1 interacts with the Tag7–Hsp70 complex, the latter dissociates, with further formation of Mts1–Tag7 and Mts1–Hsp70 complexes that lack cytotoxicity [44]. Thus, Mts1 secretion by tumor cells protects them from the toxic effect of Tag7–Hsp70 [45].

Indeed, cells with a high Mts1 level are not targeted by the Tag7–Hsp70 cytotoxic complex. This complex usually induces a cytotoxic signal in tumor cells with a low metastatic potential [45]. Obviously, this is due to the secretion of high levels of the Mts1 protein by active metastatic cells, which leads to dissociation of the cytotoxic complex. Therefore, Mts1 secretion appears to be one of the ways for tumor cells to escape the action of cytotoxic agents.

THE Tag7–Mts1 COMPLEX IS A CHEMOKINE

Investigation of Tag7 chemotactic activity has yielded contradictory data: some researchers have argued that

Tag7 is unable to induce chemotaxis of lymphocytes [4], while others have found that neutrophil-secreted Tag7 is able to induce cell movement [2]. As in the case of Tag7 cytotoxicity, both research groups are partially right.

Tag7 lacks chemotactic activity, but the Tag7–Mts1 complex causes directed migration of NK cells and CD4+ and CD8+ T lymphocytes along the complex concentration gradient [46]. Interestingly, the Tag7–Mts1 complex has a number of features atypical of classical chemokines. Tag7–Mts1 is a two-component complex with a high molecular weight, and none of its constituent proteins possesses the Greek key structure typical of most chemokines. Nevertheless, this complex induces a chemotactic signal through the chemotactic receptor CCR5 specific for ligands with a classical chemokine structure [47].

Apparently, protein components of the Tag7–Mts1 complex play different roles in inducing chemotaxis. Mts1 can bind to the CCR5 extracellular domain and inhibit the interaction between this receptor and ligands. However, this binding is insufficient to induce cell migration. Tag7 cannot interact with the CCR5 receptor; however, it participates in the transduction of a chemotactic signal by binding to Mts1 [47].

The Tag7–Mts1 complex can be secreted by both innate and adaptive immune cells [46]. Interestingly, secretion of the Tag7–Mts1 complex and, hence, induction of directed lymphocyte migration occur without preliminary activation of immunocompetent cells. Hence, effector lymphocytes start migrating along the gradient of the Tag7–Mts1 complex concentration before the immune response onset, which provides a rapid immune reaction to pathogen invasion.

Thus, Mts1, on the one hand, destroys the Tag7–Hsp70 cytotoxic complex and, on the other hand, forms the Tag7–Mts1 complex recruiting different types of T lymphocytes to the tumor to attack tumor cells.

Tag7 AND Mts1 PARTICIPATE IN THE ACTIVITY OF A NEW TYPE OF CD4+ LYMPHOCYTES DIRECTED AGAINST TUMOR CELLS LACKING HLA ANTIGENS

Tag7 and Mts1 also interact with each other in another process: the killing of tumor cells lacking the HLA complex by CD4+ lymphocytes.

CD4+ T lymphocytes are mostly immune regulatory cells involved in the activation of effector T cells by secreting a wide range of cytokines [48]. In addition, they can kill various cells, including tumor cells carrying major histocompatibility complex (MHC) class II proteins on their surface. Cell death is induced through the classical pathway by interaction between the TCR receptor and antigens in complex with MHC II, alongside with secretion of perforin and granzymes [48].

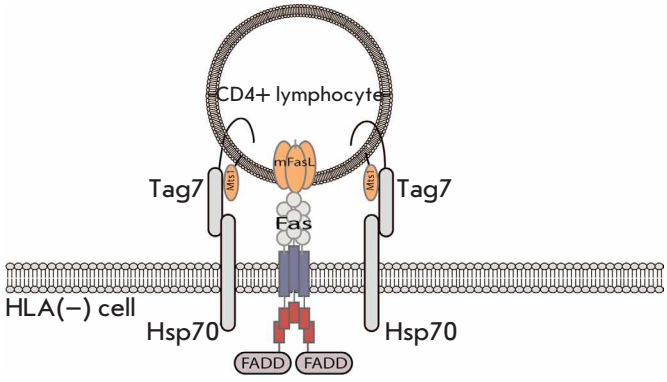


Fig. 2. The Tag7–Mts1 complex is involved in the recognition of HLA-negative tumor cells by CD4+ cytotoxic lymphocytes

A new subset of cytotoxic CD4+ T lymphocytes has recently been identified. These lymphocytes kill tumor cells lacking MHC I and MHC II proteins but carrying the major heat shock protein Hsp70 on their surface [49].

IL-2 is shown to induce the generation of cytotoxic CD4+ and CD8+ T lymphocytes in LAK cells; these lymphocytes kill HLA$^-$ tumor cells upon interaction of FasL on the lymphocyte surface with the Fas receptor of target cells [50]. Tag7 is present on the plasma membrane of both subsets but has different functions.

Not only Tag7 and FasL but also Mts1 are present on the plasma membrane of cytotoxic CD4+ T lymphocytes. Mts1 is involved in the formation of an intercellular ternary complex between lymphocytic Tag7 and Mts1 proteins and Hsp70 on the target cell membrane. Along with Tag7, lymphocytic Mts1 is also required for the cytotoxic activity of these lymphocytes [44, 45].

Thus, a sufficiently stable intercellular complex Tag7–Mts1–Hsp70 is formed, which allows the cytotoxic lymphocyte to anchor on the target cell surface. As a result, lymphocytic FasL interacts with the Fas receptor of the target cell and induces cell death. Both Tag7 and Mts1 are essential for cytotoxic activity [44] (*Fig. 2*).

Neither TCR nor granzymes are involved in the cytolysis of these cells; a cytotoxic signal is induced through the interaction between lymphocytic FasL and the Fas receptor of the target cell [49].

The CD127 antigen was detected on the surface of these CD4+ T lymphocytes, which is atypical of regulatory T cells (Treg) [50].

It is noteworthy that the described subset of T lymphocytes exposing the CD3, CD4, CD25, and CD127 antigens and Tag7, Mts1, and FasL proteins on their surface are present in the blood of healthy donors and

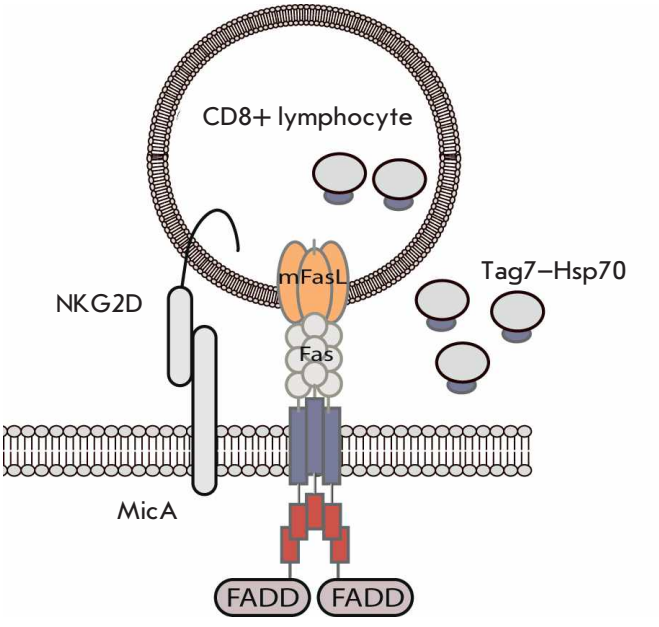


Fig. 3. Schematic representation of the recognition and killing of HLA-negative tumor cells by a CD8+ cytotoxic lymphocyte. Contact with the target tumor cell leads to the secretion of the Tag–Hsp70 cytotoxic complex

accounts for about 1% of all T lymphocytes. Probably, the subset plays an essential role in fighting tumor cells that have lost their HLA complex during tumor progression.

CD8+ T LYMPHOCYTES SECRETE THE Tag7–Hsp70 CYTOTOXIC COMPLEX

IL-2-activated CD8+ T cells can kill tumor cells that have lost surface antigens in a complex with MHC II and thus escaped the classical immune response. CD8+ T lymphocytes interact with these tumor cells via binding of the lymphocyte receptor NKG2D to the non-canonical MHC molecule MicA on the tumor cell. IL-2-activated CD8+ T lymphocytes form an intercellular NKG2D–MicA complex [51]. Although Tag7 is present on the membrane of these lymphocytes, and both MicA and Hsp70 are expressed on the membrane of the investigated tumor cells, no Tag7–Hsp70 complex forms between CD8+ T lymphocytes and target cells. This is probably due to the absence of Mts1 on the lymphocyte membrane (see above).

The interaction of NKG2D with MicA underlies two of the activities of cytotoxic lymphocytes. Firstly, it is the induction of a cytotoxic signal, followed by the death of tumor target cells due to binding of the lymphocyte FasL to the Fas receptor of the tumor cell. Secondly, it is the secretion of a soluble Tag7–Hsp70 cytotoxic complex to the cell-cell contact area [52]. Binding of the Fas receptor to FasL on the lymphocyte surface is supposed to induce an accumulation of the Tag7–Hsp70 complex in the intracellular membranes of lymphocytes. Additional binding of MicA on the target cell to NKG2D on the lymphocyte surface is required for secretion of this complex, presumably for the formation of an intercellular contact area [52] (*Fig. 3*).

HspBP1 CO-CHAPERONE IS INVOLVED IN THE REGULATION OF Tag7–Hsp70 CYTOTOXICITY

HspBP1 co-chaperone is an inhibitor of the ATPase activity of Hsp70; it can also bind to Tag7 and inhibit the cytotoxic activity of the Tag7–Hsp70 complex [53]. Various mechanisms can cause this. For instance, HspBP1 can bind to Tag7 and Hsp70, thus forming a ternary complex, followed by irreversible aggregation and formation of large conglomerates lacking cytotoxic activity. In addition, HspBP1 can competitively displace Hsp70 from the Tag7–Hsp70 complex. The resulting Tag7–HspBP1 complex has no cytotoxic effect on tumor cells. This complex is quite stable; it has been found in conditioned media of some tumor cells and human serum [54]. In the presence of high Hsp70 concentrations, Tag7–HspBP1 dissociates, with further formation of the Tag7–Hsp70 cytotoxic complex [53].

Interestingly, cytotoxic CD8+ T lymphocytes secrete Tag7–Hsp70 simultaneously with its inhibitor HspBP1; the cytotoxic activity of this complex persists for no more than 30 h. Addition of HspBP1 antibodies prevents inactivation of the secreted Tag7–Hsp70 complex during storage [53].

Thus, the inhibitor is present in lymphocytes containing the Tag7–Hsp70 complex and is secreted via the same mechanisms. Induction of its secretion also requires the formation of a contact area between the lymphocyte and the target cell [53].

PGLYRP1/Tag7 BINDS TO THE TREM-1 RECEPTOR AND INDUCES MECHANISMS OF INNATE AND ACQUIRED IMMUNITY

It has been recently established that Tag7 is a ligand for the innate immunity receptor TREM-1 that belongs to the immunoglobulin superfamily and is expressed on monocytes and neutrophils [55]. TREM-1 is believed to be involved in the activation of monocytes and the pro-inflammatory immune response [56]. The interaction between Tag7 and TREM-1 leads to the activation of the genes encoding pro-inflammatory cytokines (TNF-α, IL-6, and IL-1β) and the secretion of their products [55, 57]. This is most likely one of the ways of Tag7 involvement in antimicrobial defense in mammals, which is associated with the secretion of these cytokines.

However, lymphocyte activation, followed by cytokine secretion, observed during the interaction between Tag7 and TREM-1 is not limited to the stimulation of antimicrobial defense mechanisms solely. An activation signal induced by the innate immunity protein Tag7 is transmitted to adaptive immune regulatory and effector lymphocytes and further promotes the formation of subsets of cytotoxic lymphocytes, killing tumor and virus-infected cells that have escaped immune surveillance [57]. As in the case of IL-2-activated lymphocytes, Tag7-activated CD4+ and CD8+ T cells were shown to recognize stress proteins (Hsp70 and the non-canonical molecules HLA and MicA) on the target cell surface and kill these cells through the FasL–Fas interaction via either apoptosis or necroptosis.

A low-molecular-weight immunity activator, Tilorone, was shown to induce production of the same cytotoxic lymphocytes, which indicates a common mechanism for the formation of these cytotoxic populations [58].

Tag7 AND CANCER THERAPY

The data herein suggest that Tag7 is a promising anticancer agent. In fact, studies in this area have already been started. The very first studies on Tag7 functions assessed its effect on the growth of grafted VMR-0 tumors in mice [59, 60]. Like the vast majority of tumors, these cells do not synthesize Tag7. The cells were transfected, with genetic constructs providing a moderate expression of *tag7* since its more active production results in cell death.

Control VMR-0 tumors grew rapidly and caused the death of mice after about a month. Tumors expressing Tag7 grew much more slowly and disappeared after several months. Next, the mice were administered a mixture of control and transfected cells. Growth of these tumors was intermediate between the growth of control and transfected cells. However, the tumors disappeared again after several months. Interestingly, Tag7-producing tumors were heavily infiltrated with NK cells, in contrast to the control tumors.

Given the obtained results, the first-phase clinical trials of autologous vaccines based on *tag7* were carried out at the N.N. Blokhin Russian Cancer Research Centre (Moscow) and N.N. Petrov Research Institute of Oncology (St. Petersburg) [61, 62]. The trials were carried out in patients with either stage IV melanoma or stage IV renal cancer for whom all mandatory therapies had failed. Cell cultures were obtained from surgical samples. The cells were transfected with a construct carrying the human *tag7* gene expressing the Tag7 protein. Following inactivation of the cells by X-ray irradiation, they were subcutaneously injected to the same patient from whom the tumor was obtained. The vaccine was shown to be completely safe; some positive effect was noted in 20–25% of cases, which was observed in the form of either tumor growth stabilization or its partial regression up to a complete reduction of large metastases.

Phase 2 clinical trials of these vaccines were carried out in 80 patients with the same tumor types at the N.N. Petrov Research Institute of Oncology [63]. The number of vaccine injections was increased (up to 26 injections). Some of the patients did not respond to therapy. Contact with the remaining patients was lost at different time points for reasons unrelated to the disease. Only those patients who were followed up for up to five years were taken into account. A total of 12 out of these 74 patients survived for more than five years: Contact with them was lost after 5–15 years. Moreover, the patients had no signs of tumor progression at the time of the last follow-up. The *Table* shows that the fate of some patients can be followed up for

Tag7 therapeutic effect

Tumor staging	Last follow-up (years after therapy)	Tumor progression	Baseline age
MELANOMA (63)			
3	15.4	No cases	33
3	15.2	Same	39
4	14.9	«–»	40
3	12.1	«–»	67
4	8.9	«–»	56
4	8.8	«–»	65
3	8.6	«–»	59
4	7.1	«–»	62
3	6.9	«–»	41
3	5.3	«–»	35
RENAL CANCER (11)			
4	9.9	There were no cases. The patient died 10 years later due to another cause	58
4	5.2	No cases	65

up to 15 years. Unfortunately, these results were not formally approved, because the trials were carried out according to the previous regulations, when preclinical studies were performed by a research and development laboratory, and vaccines were prepared not at a certified institution but either in a laboratory or in a clinic.

Complete cure of 16% of fatal patients is of certain interest, especially because there are suggestions as to why other cases failed. On the one hand, one of the most important factors in the described therapeutic approach seems to be the recruitment of different T lymphocyte types to the tumor. On the other hand, a number of mechanisms are known through which tumor cells become unrecognizable to protective T lymphocytes [64]. One of the important mechanisms is synthesis of DP-L1, a DP1 receptor ligand, by tumor cells [64]. Antibodies to DP-L1 or DP1 were shown to cause a strong therapeutic effect in patients with melanoma and other tumors, due to disrupted DP-L1–DP1 interaction [64]. There exist several commercial drugs of this type. A strong synergistic effect may be expected from a combined use of the two technologies, because each of them complements the other.

In addition, autologous vaccines should be substituted for allogeneic ones, which are much more technologically convenient. This switch requires a number of genetic, technological manipulations. There exist some studies in this area.

Thus, of 74 patients followed for ≥ 5 years (from the time of the last follow-up), 12 (16.2%) patients remained alive and had no signs of tumor progression at the last follow-up after > 5 years, while nine and three patients remained alive after seven and 15 years, respectively. The follow-up was terminated for reasons unrelated to the disease.

CONCLUSIONS

In conclusion, we would like to emphasize that PGLYRP1/Tag7 is one of the key regulatory proteins involved in immune responses. PGLYRP1/Tag7 is classified as an innate immunity protein, but it can participate in the regulation of the immune mechanisms of both innate and acquired immunity. Tag7 induces antimicrobial defense mechanisms and the formation of subsets of cytotoxic lymphocytes killing cells that have escaped the antitumor immune response. The Tag7–Hsp70 complex causes the death of tumor cells carrying the TNFR1 receptor.

Investigation of the Tag7 crystal structure revealed the presence of a protein-protein interaction site in it. Apparently, Tag7 can interact with various proteins and this interaction determines its multiple functional activities. To date, the ability of proteins to change their function after interacting with other proteins and forming stable complexes is well known and referred to as moonlighting [65].

The above data indicate that Tag7 can bind to five proteins: TREM-1, TNFR1, Hsp70, HspBP1, and Mts. Two of these proteins are receptors exposed on the plasma membrane of immune and tumor cells and involved in the induction of the immune response. The interaction of Tag7 with these proteins triggers the innate and adaptive responses involved in the host defense against pathogens (*Fig. 4*).

The antimicrobial effect of Tag7 in insects is associated with activation of the serine protease cascade, which converts Spatzle, a Toll receptor ligand, into an active form, followed by the release of antimicrobial peptides. The antimicrobial activity of PGRP in mammals is associated with three cytotoxic mechanisms: induction of oxidative, thiol, and metal stress. However, PGRP also functions in cooperation with other immune defense mechanisms and antimicrobial peptides.

The interaction between Tag7 and the human innate immune receptor TREM-1 at an early stage of monocyte activation results in the secretion of pro-inflammatory cytokines, inducing one of the antimicrobial defense pathways. Further transmission of the activation signal to regulatory cells activates subsets of cytotoxic lymphocytes, eliminating tumor and virus-containing cells that have lost their surface HLA antigens.

These lymphocytes can kill tumor cells through both the contact mechanism of lysis through the FasL–Fas interaction and the secretory mechanism through the release of the Tag7–Hsp70 cytotoxic complex into the contact area.

Secretion of the HspBP1 co-chaperone regulates the cytotoxic effect of the Tag7–Hsp70 complex. HspBP1 is secreted by lymphocytes simultaneously with the cytotoxic complex and can inhibit its activity through either disordered aggregation of the ternary Tag7–Hsp70–HspBP1 complex or dissociation of the Tag7–Hsp70 complex.

By binding to the extracellular domain of the receptor, Tag7 alone inhibits transduction of the cytotoxic signal to tumor cells. It is not only unable to cause cell death, but also inhibits the cytotoxic effect of other TNFR1 ligands, mainly TNF-α activity. Both Hsp70 and the formation of the cytotoxic complex on the cell surface are required for cytolysis induction by Tag7. In this regard, identification of a Tag7 peptide fragment modulating its functions is of particular interest. Expanding the spectrum of these functional peptides may be relevant in the development of drugs that inhibit acute inflammatory processes.

Involvement of Tag7 in the immune response is not limited to the activation of cytotoxic lymphocytes and

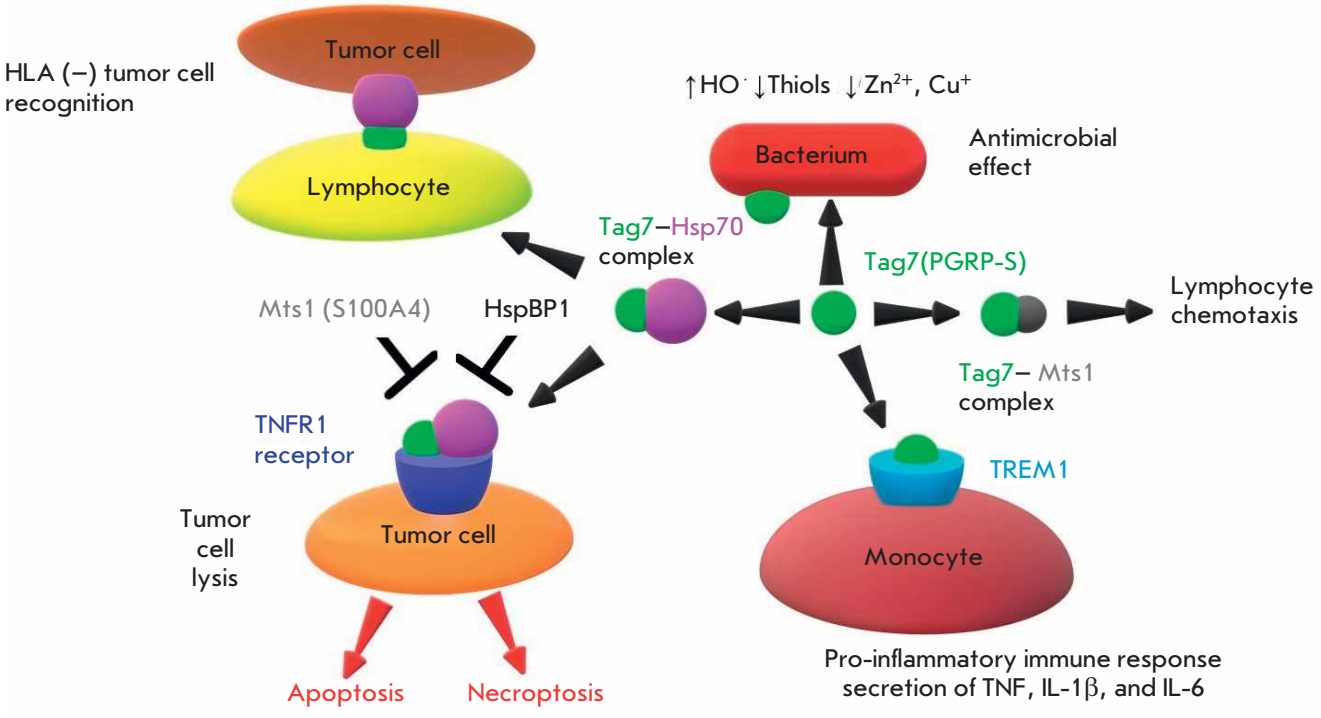


Fig. 4. Functions of the PGLYRP1/Tag7 protein

the cytotoxic effect, together with Hsp70, on tumor cells. Tag7 can also interact with the Mts1 (S100A4) protein present in a wide range of metastatic tumors. Soluble Mts1 competes with Hsp70 for binding to Tag7, displacing the latter from the cytotoxic Tag7–Hsp70 complex to form an inactive Tag7–Mts1 complex. However, the Tag7–Mts1 complex has chemotactic activity and induces directed migration of innate and adaptive immune effector lymphocytes along the complex concentration gradient.

The Tag7–Mts1 complex is secreted by immune system cells, mainly neutrophils and monocytes, without pre-activation, which can yield rapid development of immune responses upon pathogen infection.

Experiments on mice using a number of tumor cell lines showed that an injection of tumor cells transfected with a construct producing Tag7 inhibits the growth of a grafted tumor of the same cell line. Autologous vaccines have been created based on these data; they have passed the first and second phases of clinical trials in fatal patients with melanoma or kidney cancer. Complete cure was observed in 12 out of 74 cases. There exist a number of opportunities to significantly improve the treatment effectiveness.

The above facts indicate that Tag7 is a multifunctional protein that is involved in the regulation of various stages of the immune response and is a promising agent for practical use in oncology. ●

REFERENCES

1. Kustikova O.S., Kiselev S.L., Borodulina O.R., Senin V.M., Afanas'eva A.V., Kabishev A.A. // Genetika. 1996. V. 32. № 5. P. 621–628.
2. Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhortchouk E.B., Kabishev A.A., Lukanidin E.M., Georgiev G.P. // J. Biol. Chem. 1998. V. 273. № 29. P. 18633–18639.
3. Kang D., Liu G., Lundström A., Gelius E., Steiner H. // Proc. Natl. Acad. Sci. USA. 1998. V. 95. № 17. P. 10078–10082.
4. Liu C., Gelius E., Liu G., Steiner H., Dziarski R. // J. Biol. Chem. 2000. V. 275. № 32. P. 24490–24499.
5. Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D. // Proc. Natl. Acad. Sci. USA. 2000. V. 97. № 25. P. 13772–13777.
6. Liu C., Xu Z., Gupta D., Dziarski R. // J. Biol. Chem. 2001. V. 276. № 37. P. 34686–34694.
7. Kibardin A.V., Mirkina I.I., Baranova E.V., Zakeyeva I.R., Georgiev G.P., Kiselev S.L. // J. Mol. Biol. 2003. V. 326. № 2. P. 467–474.
8. Lu X., Wang M., Qi J., Wang H., Li X., Gupta D., Dziarski R. // J. Biol. Chem. 2006. V. 281. № 9. P. 5895–5907.
9. Mirkina I.I., Kiselev S.L., Sashchenko L.P., Sadchikova E.R., Gnuchev N.V. // Dokl. Akad. Nauk. 1999. V. 367. № 4. P. 548–552.

10. Kim M.S., Byun M., Oh B.H. // Nat. Immunol. 2003. V. 4. № 8. P. 787–793.
11. Hultmark D. // Curr. Opin. Immunol. 2003. V. 15. № 1. P. 12–19.
12. Söderhäll K., Cerenius L. // Curr. Opin. Immunol. 1998. V. 10. № 1. P. 23–28.
13. Gottar M., Gobert V., Michel T., Belvin M., Duyk G., Hoffmann J.A., Ferrandon D., Royet J. // Nature. 2002. V. 416. № 6881. P. 640–644.
14. Wang L, Weber A.N., Atilano M.L., Filipe S.R., Gay N.J., Ligoxygakis P. // EMBO J. 2006. V. 25. № 20. P. 5005–5014.
15. Khush R.S., Leulier F., Lemaitre B. // Trends Immunol. 2001. V. 22. № 5. P. 260–264.
16. Hoffmann J.A., Kafatos F.C., Janeway C.A., Ezekowitz R.A. // Science. 1999. V. 284. № 5418. P. 1313–1318.
17. Takehana A., Katsuyama T., Yano T., Oshima Y., Takada H., Aigaki T., Kurata S. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. № 21. P. 13705–13710.
18. Dziarski R., Park S.Y., Kashyap D.R., Dowd S.E., Gupta D. // PLoS One. 2016. V. 11. № 1. e0146162.
19. Royet J., Dziarski R. // Nat. Rev. Microbiol. 2007. V. 5. № 4. P. 264–277.
20. Lu X., Wang M., Qi J., Wang H., Li X., Gupta D., Dziarski R. // J. Biol. Chem. 2006. V. 281. № 9. P. 5895–5907.
21. Tydell C.C., Yount N., Tran D., Yuan J., Selsted M.E. // J. Biol. Chem. 2002. V. 277. № 22. P. 19658–19664.
22. Tydell C.C., Yuan J., Tran P., Selsted M.E. // J. Immunol. 2006. V. 176. № 2. P. 1154–1162.
23. Li X., Wang S., Qi J., Echtenkamp S.F., Chatterjee R., Wang M., Boons G.J., Dziarski R., Gupta D. // Immunity. 2007. V. 27. № 3. P. 518–529.
24. Gelius E., Persson C., Karlsson J., Steiner H. // Biochem. Biophys. Res. Commun. 2003. V. 306. № 4. P. 988–994.
25. Wang Z.M., Li X., Cocklin R.R., Wang M., Wang M., Fukase K., Inamura S., Kusumoto S., Gupta D., Dziarski R. // J. Biol. Chem. 2003. V. 278. № 49. P. 49044–49052.
26. Sharma P., Dube D., Sinha M., Mishra B., Dey S., Mal G., Pathak K.M., Kaur P., Sharma S., Singh T.P. // J. Biol. Chem. 2011. V. 286. № 36. P. 31723–31730.
27. Kashyap D.R., Rompca A., Gaballa A., Helmann J.D., Chan J., Chang C.J., Hozo I., Gupta D., Dziarski R. // PLoS Pathog. 2014. V. 10. № 7. e1004280.
28. Kashyap D.R., Kuzma M., Kowalczyk D.A., Gupta D., Dziarski R. // Mol. Microbiol. 2017. V. 105. № 5. P. 755–776.
29. Chandrangsu P., Rensing C., Helmann J.D. // Nat. Rev. Microbiol. 2017. V. 15. № 6. P. 338–350.
30. German N., Doyscher D., Rensing C. // Future Microbiol. 2013. V. 8. № 10. P. 1257–1264.
31. Cho J.H., Fraser I.P., Fukase K., Kusumoto S., Fujimoto Y., Stahl G.L., Ezekowitz R.A. // Blood. 2005. V. 106. № 7. P. 2551–2558.
32. Sashchenko L.P., Dukhanina E.A., Yashin D.V., Shatalov Y.V., Romanova E.A., Korobko E.V., Demin A.V., Lukyanova T.I., Kabanova O.D., Khaidukov S.V., et al. // J. Biol. Chem. 2004. V. 279. № 3. P. 2117–2124.
33. Multhoff G., Pfister K., Gehrmann M., Hantschel M., Gross C., Hafner M., Hiddemann W. // Cell Stress Chaperones. 2001. V. 6. № 4. P. 337–344.
34. Dukhanina E.A., Yashin D.V., Lukjanova T.I., Romanova E.A., Kabanova O.D., Shatalov Y.V., Sashchenko L.P., Gnuchev N.V. // Dokl. Biol. Sci. 2007. V. 414. P. 246–248.
35. Yashin D.V., Ivanova O.K., Soshnikova N.V., Sheludchenkov A.A., Romanova E.A., Dukhanina E.A., Tonevitsky A.G., Gnuchev N.V., Gabibov A.G., Georgiev G.P., Sashchenko L.P. // J. Biol. Chem. 2015. V. 290. № 35. P. 21724–21731.
36. Christofferson D.E., Yuan J. // Curr. Opin. Cell Biol. 2010. V. 22. № 2. P. 263–268.
37. Holler N., Zaru R., Micheau O., Thome M., Attinger A., Valitutti S., Bodmer J.L., Schneider P., Seed B., Tschopp J. // Nat. Immunol. 2000. V. 1. № 6. P. 489–495.
38. Nagata S. // Cell. 1997. V. 88. № 3. P. 355–365.
39. Yashin D.V., Romanova E.A., Ivanova O.K., Sashchenko L.P. // Biochimie. 2016. V. 123. P. 32–36.
40. Wingfield P., Pain R.H., Craig S. // FEBS Lett. 1987. V. 211. № 2. P. 179–184.
41. Romanova E.A., Sharapova T.N., Telegin G.B., Minakov A.N., Chernov A.S., Ivanova O.K., Bychkov M.L., Sashchenko L.P., Yashin D.V. // Cells. 2020. V. 9. № 2. P. 488.
42. Ebralidze A., Tulchinsky E., Grigorian M., Afanasyeva A., Senin V., Revazova E., Lukanidin E. // Genes Dev. 1989. V. 3. № 7. P. 1086–1093.
43. Kriajevska M.V., Cardenas M.N., Grigorian M.S., Ambartsumian N.S., Georgiev G.P., Lukanidin E.M. // J. Biol. Chem. 1994. V. 269. № 31. P. 19679–19682.
44. Dukhanina E.A., Kabanova O.D., Lukyanova T.I., Shatalov Y.V., Yashin D.V., Romanova E.A., Gnuchev N.V., Galkin A.V., Georgiev G.P., Sashchenko L.P. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 33. P. 13963–13967.
45. Dukhanina E.A., Yashin D.V., Galkin A.V., Sashchenko L.P. // Cell Cycle. 2010. V. 9. № 4. P. 676–682.
46. Dukhanina E.A., Lukyanova T.I., Romanova E.A., Guerriero V., Gnuchev N.V., Georgiev G.P., Yashin D.V., Sashchenko L.P. // Cell Cycle. 2015. V. 14. № 22. P. 3635–3643.
47. Sharapova T.N., Romanova E.A., Sashchenko L.P., Yashin D.V. // Acta Naturae. 2018. V. 10. № 4. P. 115–120.
48. Appay V., Zaunders J.J., Papagno L., Sutton J., Jaramillo A., Waters A., Easterbrook P., Grey P., Smith D., McMichael A.J., et al. // J. Immunol. 2002. V. 168. № 11. P. 5954–5958.
49. Sashchenko L.P., Dukhanina E.A., Shatalov Y.V., Yashin D.V., Lukyanova T.I., Kabanova O.D., Romanova E.A., Khaidukov S.V., Galkin A.V., Gnuchev N.V., Georgiev G.P. // Blood. 2007. V. 110. № 6. P. 1997–2004.
50. Sharapova T.N., Romanova E.A., Sashchenko L.P., Yashin D.V. // J. Immunol. Res. 2018. V. 2018. P. 4501273.
51. Ivanova O.K., Sharapova T.N., Romanova E.A., Soshnikova N.V., Sashchenko L.P., Yashin D.V. // J. Cell. Biochem. 2017. V. 118. № 10. P. 3359–3366.
52. Sashchenko L.P., Romanova E.A., Ivanova O.K., Sharapova T.N., Yashin D.V. // IUBMB Life. 2017. V. 69. № 1. P. 30–36.
53. Yashin D.V., Dukhanina E.A., Kabanova O.D., Romanova E.A., Lukyanova T.I., Tonevitskii A.G., Raynes D.A., Gnuchev N.V., Guerriero V., Georgiev G.P., Sashchenko L.P. // J. Biol. Chem. 2011. V. 286. № 12. P. 10258–10264.
54. Yashin D.V., Dukhanina E.A., Kabanova O.D., Romanova E.A., Lukyanova T.I., Tonevitskii A.G., Belogurov A.A., Raynes D.A., Sheludchenkov A.A., Gnuchev N.V., et al. // Biochimie. 2012. V. 94. № 1. P. 203–206.
55. Read C.B., Kuijper J.L., Hjorth S.A., Heipel M.D., Tang X., Fleetwood A.J., Dantzler J.L., Grell S.N., Kastrup J., Wang C., et al. // J. Immunol. 2015. V. 194. № 4. P. 1417–1421.
56. Gibot S., Kolopp-Sarda M.N., Béné M.C., Bollaert P.E., Lozniewski A., Mory F., Levy B., Faure G.C. // J. Exp. Med. 2004. V. 200. № 11. P. 1419–1426.
57. Sharapova T.N., Ivanova O.K., Soshnikova N.V., Romanova E.A., Sashchenko L.P., Yashin D.V. // J. Innate Immun. 2017. V. 9. № 6. P. 598–608.
58. Sharapova T.N., Romanova E.A., Sashchenko L.P., Yashin D.V. // IUBMB Life. 2019. V. 71. № 3. P. 376–384.
59. Larin S.S., Korobko E.V., Kustikova O.S., Borodulina O.R.,

Raikhlin N.T., Brisgalov I.P., Georgiev G.P., Kiselev S.L. // J. Gene Med. 2004. V. 6. № 7. P. 798–808.
60. Kiselev S.L., Larin S.S., Gnuchev N.V., Georgiev G.P. // Genetika. 2000. V. 36. № 11. P. 1431–1435.
61. Moiseenko V.M., Danilov A.O., Baldueva I.A., Danilova A.B., Tiukavina N.V., Larin S.S., Kiselev S.L., Orlova R.V., Semenova A.I., Turkevich E.A., et al. // Vopr. Onkol. 2004. V. 50. № 3. P. 293–303.
62. Moiseyenko V.M., Danilov A.O., Baldueva I.A., Danilova A.B., Tyukavina N.V., Larin S.S., Kiselev S.L., Orlova R.V., Anisimov V.V., Semenova A.I., et al. // Ann. Oncol. 2005. V. 16. № 1. P. 162–168.
63. Novik A.V., Danilova A.B., Sluzhev M.I., Nehaeva T.L., Larin S.S., Girdyuk D.V., Protsenko S.A., Semenova A.I., Danilov A.O., Moiseyenko V.M., et al. // Oncologist. 2020. doi: 10.1634/theoncologist.2020-0160. Online ahead of print.
64. Ingram J.R., Blomberg O.S., Rashidian M., Ali L., Garforth S., Fedorov E., Fedorov A.A., Bonanno J.B., Gall C.L., Crowley S., et al. // Proc. Natl. Acad. Sci. USA. 2018. V. 115. № 15. P. 3912–3917.
65. Jeffery C.J. // Protein Sci. 2019. V. 28. № 7. P. 1233–1238.

COVID-19 in Russia: Clinical and Immunological Features of the First-Wave Patients

T. V. Bobik[1], N. N. Kostin[1], G. A. Skryabin[1], P. N. Tsabai[1], M. A. Simonova[1], V. D. Knorre[1*], O. N. Stratienko[2], N. L. Aleshenko[2], I. I. Vorobiev[1], E. N. Khurs[3], Yu. A. Mokrushina[1], I. V. Smirnov[1], A. I. Alekhin[2], A. E. Nikitin[2], A. G. Gabibov[1]

[1]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia

[2]Central Clinical Hospital of the Russian Academy of Sciences, Moscow, 117593 Russia

[3]Engelhardt Institute of Molecular Biology, RAS, Moscow, 119991 Russia

[*]E-mail: vera.knorre@gmail.com

Received December 20, 2020; in final form, February 16, 2021

DOI: 10.32607/actanaturae.11374

ABSTRACT **The coronavirus disease outbreak in 2019 (COVID-19) has now achieved the level of a global pandemic and affected more than 100 million people on all five continents and caused over 2 million deaths. Russia is, needless to say, among the countries affected by SARS-CoV-2, and its health authorities have mobilized significant efforts and resources to fight the disease. The paper presents the result of a functional analysis of 155 patients in the Moscow Region who were examined at the Central Clinical Hospital of the Russian Academy of Sciences during the first wave of the pandemic (February–July, 2020). The inclusion criteria were a positive PCR test and typical, computed tomographic findings of viral pneumonia in the form of ground-glass opacities. A clinical correlation analysis was performed in four groups of patients: (1) those who were not on mechanical ventilation, (2) those who were on mechanical ventilation, and (3) those who subsequently recovered or (4) died. The correlation analysis also considered confounding comorbidities (diabetes, metabolic syndrome, hypertension, etc.). The immunological status of the patients was examined (levels of immunoglobulins of the M, A, G classes and their subclasses, as well as the total immunoglobulin level) using an original SARS-CoV-2 antibody ELISA kit. The ELISA kit was developed using linear S-protein RBD-SD1 and NTD fragments, as well as the N-protein, as antigens. These antigens were produced in the prokaryotic *E. coli* system. Recombinant RBD produced in the eukaryotic CHO system (RBD CHO) was used as an antigen representing conformational RBD epitopes. The immunoglobulin A level was found to be the earliest serological criterion for the development of a SARS-CoV-2 infection and it yielded the best sensitivity and diagnostic significance of ELISA compared to that of class M immunoglobulin. We demonstrated that the seroconversion rate of "early" N-protein-specific IgM and IgA antibodies is comparable to that of antibodies specific to RBD conformational epitopes. At the same time, seroconversion of SARS-CoV-2 N-protein-specific class G immunoglobulins was significantly faster compared to that of other specific antibodies. Our findings suggest that the strong immunogenicity of the RBD fragment is for the most part associated with its conformational epitopes, while the linear RBD and NTD epitopes have the least immunogenicity. An analysis of the occurrence rate of SARS-CoV-2-specific immunoglobulins of different classes revealed that RBD- and N-specific antibodies should be evaluated in parallel to improve the sensitivity of ELISA. An analysis of the immunoglobulin subclass distribution in sera of seropositive patients revealed uniform induction of N-protein-specific IgG subclasses G1–G4 and IgA subclasses A1–A2 in groups of patients with varying severity of COVID-19. In the case of the S-protein, G1, G3, and A1 were the main subclasses of antibodies involved in the immune response.**

KEYWORDS **serological analysis of patients with COVID-19, SARS-CoV-2-specific antibody subclasses.**

ABBREVIATIONS **COVID-19 – coronavirus disease 2019; SARS-CoV-2 – severe acute respiratory syndrome coronavirus 2; PCR test – polymerase chain reaction test; ELISA – enzyme-linked immunosorbent assay; RBD-SD1 – receptor-binding domain–subdomain 1; NTD – N-terminal domain; RBD – receptor binding domain; CHO cells – Chinese hamster ovary cells.**

INTRODUCTION

The pandemic, officially declared by the WHO on March 11, 2020, after the rapid spread of the new coronavirus disease in 2019 (COVID-19), has proved a challenge for the global medical and scientific communities. By February 2021, more than 100 million people had been infected with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) across the world, and more than 2 million people had died. The infection spread rather quickly in the regions of Russia. According to the Ministry of Health of the Russian Federation, as of February 10, 2021, a total of more than 4 million people have been infected across the country; of these, more than 80,000 people have died.

The new coronavirus SARS-CoV-2, which belongs to the genus *Betacoronavirus*, is a cytopathic single-stranded RNA virus assigned to the II pathogenicity group. This virus infects cells carrying angiotensin-converting enzyme 2 (ACE2) receptors on their surface, mainly type II alveolar pneumocytes and, to a lesser extent, other epithelial cells [1]. Infection with the SARS-CoV-2 coronavirus leads to a wide range of manifestations, from asymptomatic to severe acute respiratory distress syndrome (ARDS) leading to death. According to our statistical analysis, about 80% of patients have a mild form of the disease not requiring hospitalization, with clinical signs of an acute respiratory tract infection with typical catarrhal symptoms, and they usually develop spontaneous recovery. The disease course usually resembles that of an acute respiratory viral infection (ARVI) caused by the influenza A and B viruses, rhinoviruses, adenoviruses, and seasonal coronaviruses; however, in some cases, the SARS-CoV-2 virus infection can lead to a very rapid acute inflammation with the development of severe bilateral pneumonia, hemorrhagic fever, and organ dysfunctions. A dramatic course of the disease is accompanied by severe pneumonia and affects 15% of patients; about 5% of patients develop ARDS and multiple organ failure. The mortality rate varies from country to country and, according to recent data, amounts to 1.04 to 8.5% of confirmed disease cases. Over the past year, many attempts have been made to establish a relationship between various factors (e.g., gender, age, race, comorbidities, various indicators and markers (including genetic ones), etc.) and the severity of the disease [2–8].

However, despite the large amount of data accumulated to date, most of the identified correlations remain inconsistent. In most publications, the genetic predisposition to the development of complications is associated with the structural features of ACE2, antigen presentation system, and the genes responsible for the innate immune system [9].

Humoral responses were used as the main markers of disease severity in other viral lung infections, including SARS-CoV and influenza virus infections [10–13].

The SARS-CoV-2 genome encodes four structural proteins: spike (S-protein), nucleocapsid (N-protein), envelope (E), and membrane (M) [14]. The S and N proteins are the two main coronavirus antigens that induce production of immunoglobulins [15]. Anti-N-protein antibodies are often induced in relatively higher amounts than other proteins used as the main targets of serological assays [15, 16].

The receptor binding domain (RBD), which is situated in the spike protein S1 subunit, is the main target of neutralizing antibodies (NAbs) and is also used in the design of vaccines [17–20].

According to reported data (WHO statistics), the mortality rate in Russia (1.89%) remains one of the lowest in the world. This fact still requires a detailed investigation. Of course, factors related to the healthcare organization in the Russian Federation may play a role in this phenomenon; however, we may suggest that the explanation for this phenomenon is related to demographic factors, as well as factors associated with risk groups and markers of inflammation severity. It was of interest to characterize in detail the humoral responses of adaptive immunity in cohorts of patients in the Russian population in response to coronavirus infection.

EXPERIMENTAL

Materials

In this study, we used reagents from Sigma, Bio-Rad, Thermo Scientific (USA), Pharmacia (Sweden), Difco (England), Panreac (Spain), and Reakhim (Russia).

Preparation of recombinant proteins and SARS-CoV-2 S- and N-protein fragments

Artificially synthesized DNA fragments encoding S-protein RBD (330–538 aa), RBD-SD1 (330–590 aa), NTD (17–305 aa) fragments, and the N-protein sequence (1–420 aa) of the SARS-CoV-2 virus were cloned into pET22b plasmids using NdeI and XhoI restriction endonucleases. The correctness of the produced constructs was confirmed by sequencing.

Escherichia coli BL21 (DE3) cells transformed with the produced genetic constructs were cultured in a 2YT medium with ampicillin at 37°C and vigorous stirring until OD_{600} = 0.4, induced with 1 mM IPTG, and cultured at 30°C for 6 h. Isolation and purification of recombinant proteins from inclusion bodies was performed using metal chelate chromatography (HiTrap FF, GE Healthcare, USA) under denaturing conditions.

Expression and purification of the recombinant RBD (amino acid residues 320–537) produced in the eukaryotic system of CHO cells was performed according to the previously described method [21].

ELISA

The purified recombinant RBD, RBD-SD1, NTD, and N-protein produced in the prokaryotic *E. coli* system were adsorbed to plate wells using buffer containing 50 mM sodium bicarbonate and 4 M urea, pH 10.6. For the detection of SARS-CoV-2-specific antibodies, a mixture of RBD-SD1, NTD, and N antigens at a ratio of 40/20/40 ng per well for each antigen (100 ng) was placed into Nun MaxiSorp flat-bottom 96-well plates (Nunc, USA) and incubated at 4°C without stirring for 16 h. After incubation, the solution was removed from the wells, the wells were washed with distilled water, and a blocking solution (phosphate buffered saline, 0.1% Tween 20, 3% BSA) was added. The plates were incubated at room temperature without stirring for 1 h. At the end of the incubation, the blocking solution was removed and the plates were dried to dryness at room temperature and stored at 6 ± 2°C.

In experiments with biological samples, sera were diluted at a ratio of 1 : 100 (for the analysis of total N-protein-specific immunoglobulin G) or 1 : 10 (in the other cases) in a washing solution (phosphate buffered saline, 0.1% Tween 20), placed into the wells, and incubated at 37°C and stirring (700 rpm, 30 min). The plates were washed 5 times with a washing solution, and antibodies to the appropriate classes and subclasses of human antibodies were added in a conjugate dilution solution (phosphate buffered saline, 0.1% Tween 20, 0.1% BSA) and incubated at 37°C and stirring (700 rpm, 30 min) [22]. The plates were washed 5 times with a washing solution, and anti-species antibodies conjugated with horseradish peroxidase in a conjugate dilution solution were added and incubated at 37°C and stirring (700 rpm, 30 min). After washing the plates with a washing solution (5 times), the TMB substrate was added and incubated in the dark for 15 min. The reaction was stopped with a 10% phosphoric acid solution, and the absorbance was measured at a wavelength of 450 nm on a plate reader.

For a correct interpretation of ELISA results, the threshold ODcrit was calculated based on OD values of a panel of sera from healthy donors (OD-ref), using the average optical density in these wells according to the following formula:

$$\text{ODcrit} = \text{OD-ref} + 3 \times \text{standard deviations}.$$

The resulting ODcrit value was used to calculate the positivity index for each test sample using the following formula:

$$\text{PIsamp} = \text{ODsamp}/\text{ODcrit}.$$

At PIsamp ≥ 1, a blood serum sample was considered positive (the sample contains SARS-CoV-2 coronavirus-specific antibodies); at PIsamp < 0.9, a serum sample was considered negative.

The data were statistically processed using the GraphPad Prism 8 software.

RESULTS AND DISCUSSION

This study was based on clinical material collected at the Central Clinical Hospital of the Russian Academy of Sciences during April–May, 2020. A total of 155 patients diagnosed with COVID-19 were examined. The criteria for inclusion in the group of COVID-19-positive patients were a positive PCR test and pulmonary lesions identified in CT scans as ground-glass opacities. We searched for statistically significant differences in the disease course among groups of patients different in gender, age, and comorbidities. The number of days spent in the hospital was used as an indicator to indirectly assess the disease severity.

An analysis (*Fig. 1A*) using the nonparametric Mann–Whitney test revealed that the number of days after the onset of symptoms to hospitalization did not differ among patients of two age groups, and that there was no correlation between the number of days and the age of patients within the two age groups. However, the length of hospital stay was longer in the group of patients over 51 years of age ($p < 0.0001$), which indicates a more severe course of the disease in older patients. An analysis of the effect of gender on the length of hospital stay in patients of the two age groups (*Fig. 1B*) revealed slight differences between the groups of males under 50 and those over 51 years of age ($p = 0.0177$). There was a significant difference ($p < 0.0001$) in the length of hospital stay in females under 50 and those over 51 years of age. There were no statistically significant differences in the length of hospital stay in patients of different gender within the two age groups and regardless of age. The identified dependencies were confirmed by a correlation analysis; the Spearman's correlation coefficient (r) value was statistically significant in the group of females ($r = 0.65$) and insignificant in the group of males ($r = 0.31$). An analysis of variance on the relationship between the length of hospital stay and a linear combination of age and gender factors showed that gender was not a significant factor ($p = 0.719$), while age, on the con-

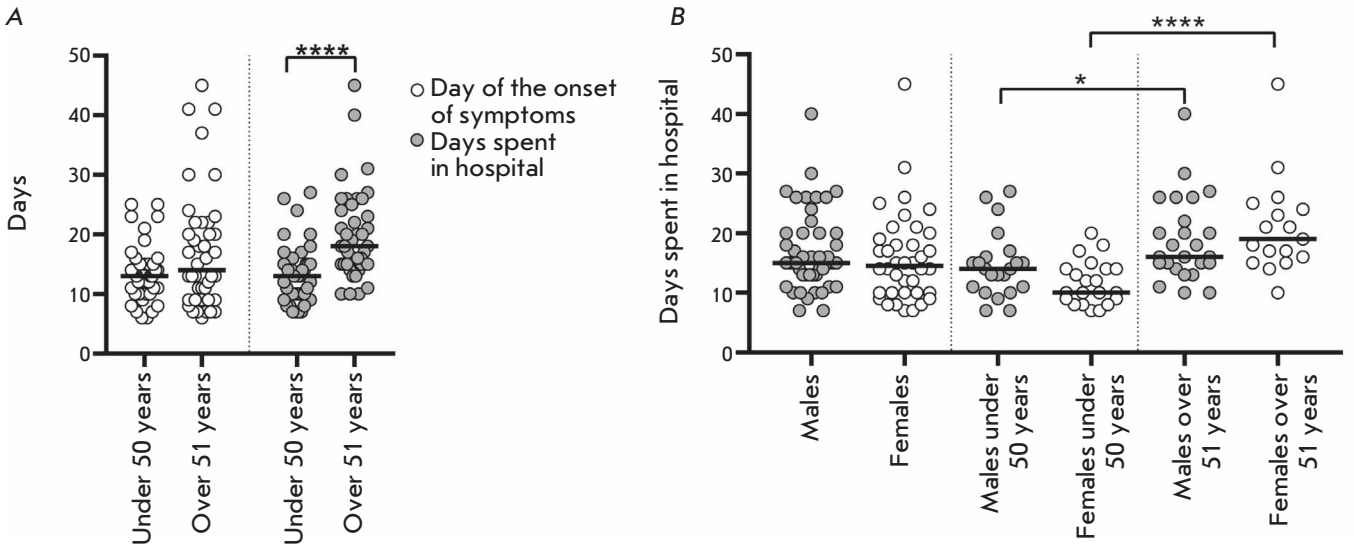


Fig. 1. Distribution of the number of days from the onset of symptoms to hospitalization (A) and days spent in a hospital (*A*, *B*) among patients of different age groups

trary, affected the length of hospital stay ($p < 0.0001$), and this relationship was more significant in the female group than in the male group ($p << 0.0001$ and $p = 0.0278$, respectively). Therefore, this difference in the length of hospital stay of patients of different age groups is significantly associated with the difference between females of older (over 51 years) and younger (under 50 years) ages.

According to the published data [4, 5, 23, 26], comorbidities, such as hypertension, diabetes mellitus, and obesity, are risk factors for a severe course of COVID-19. We investigated the effect of these comorbidities on the length of hospital stay. We analyzed the length of hospital stay in three groups of patients: 1) with hypertension, diabetes mellitus, or obesity (HDO), 2) with other comorbidities, 3) without comorbidities.

The nonparametric Mann–Whitney test results revealed a significant excess ($p = 0.01$) in the length of hospital stay of patients in the group with comorbidities, such as hypertension, diabetes mellitus, and obesity, compared to that in the group without comorbidities (*Fig. 2A*). However, further analysis showed that the age median of patients in these groups was very different: 62 and 43 years, respectively. After matching the median between the groups by excluding patients of the maximum and minimum age, respectively, from the samples, there was no significant difference in the length of hospital stay between the groups (*Fig. 2B*). Comparison of the length of hospital stay in the HDO patients of the two age groups and in patients without comorbidities (*Fig. 2C*) revealed no effect of the diseases under consideration. The effect of age is statistically significant both in the group of patients without comorbidities ($p = 0.0001$) and in the group with these diseases ($p = 0.0076$). In this case, the length of hospital stay in patients of the same age groups, differing in the presence/absence of comorbidities, did not differ statistically significantly. Thus, there was no effect of comorbidities on the severity of COVID-19 in the studied cohort of patients. Perhaps, previously published data on a correlation between disease severity and some comorbidities did not consider the age imbalance in the compared groups.

To identify differences in some hematological characteristics among groups of patients with differing severity of COVID-19, the cohort of hospitalized patients was divided into those who needed and did not need mechanical ventilation. The results of clinical studies of patients requiring mechanical ventilation were analyzed either in total or in two groups, depending on the disease outcome (recovery or death). The nonparametric Mann–Whitney test revealed a significant ($p < 0.0001$) increase in the leukocyte count in patients of the older age group (over 51 years of age) compared to that in the group under 50 years of age (*Fig. 3A*). These data are consistent with the results of other studies [25–27].

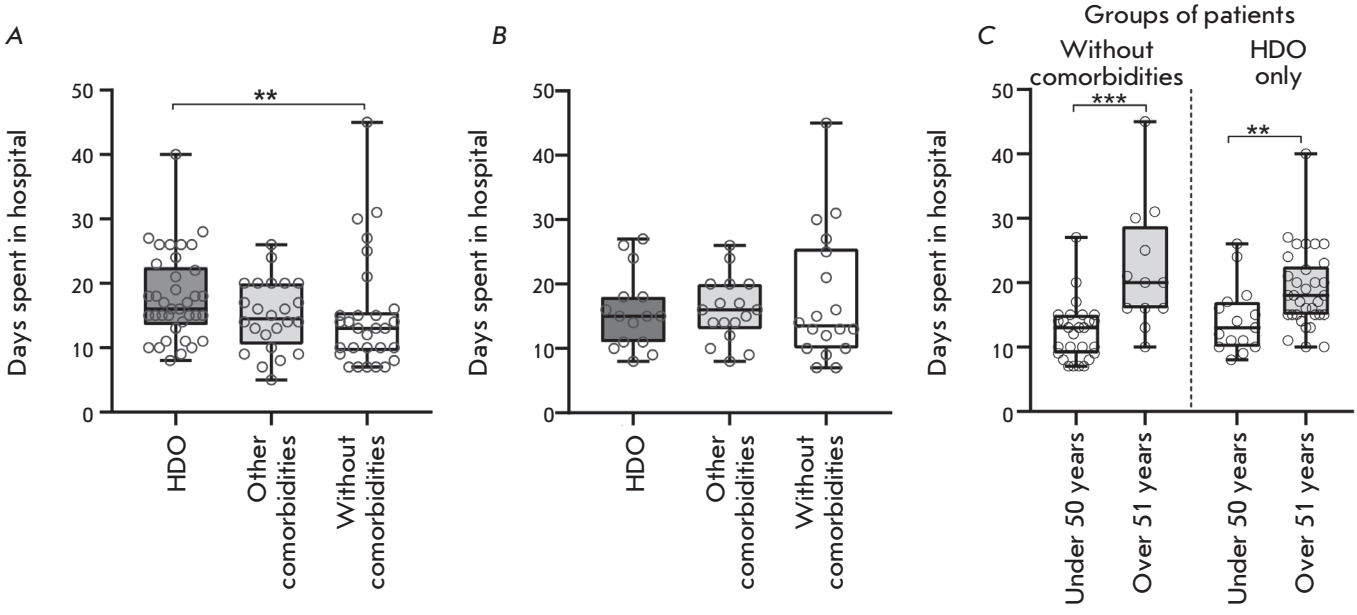


Fig. 2. Distribution of the number of days spent in a hospital by patients with/without comorbidities (*A*–*C*) and patients of different age groups (*C*). (*A*) – groups of patients regardless of age; (*B*) – groups of patients with the same median age. (*C*) – comparison of the mean length of hospital stay in patients of age groups, with/without comorbidities. HDO – hypertension, diabetes mellitus, and obesity

An increase in the leukocyte count was also detected in patients on mechanical ventilation ($p = 0.006$), which is consistent with previous data indicating that leukocytosis is associated with a severe course of COVID-19 and a high risk of death [25, 28, 29]. However, there was no significant differences in the leukocyte count between the groups of patients who recovered after mechanical ventilation therapy and those who died; therefore, it is erroneous to consider an increase in the leukocyte count as a prognostic factor of disease outcome. However, a number of researchers have proposed using leukocytosis in combination with idiopathic lymphopenia as a prognostic marker of disease severity. According to various sources, lymphopenia is detected in 40–80% of COVID-19 cases [30–32] and is pronounced in patients in critical condition [5, 33]. However, despite numerous studies indicating a close relationship between lymphopenia and disease severity, we did not find significant differences in the lymphocyte count in groups of patients with differing severity of COVID-19 in our cohort. Many researchers have suggested considering the C-reactive protein (CRP) as a prognostic factor [3, 36], a high level of which is associated with a worsening of the disease. A number of studies [3, 5, 37] have reliably demonstrated a significant increase in the blood CRP level in critical-condition patients. However, some researchers have found a slight [38], or even no, difference [39] in the CRP level at different severities of the disease. Among the groups in our study cohort, a weak but statistically significant difference ($p = 0.04$) in the mean CRP concentration was found only between the groups of patients who underwent (or not) mechanical ventilation therapy (*Fig. 3B*).

A correlation analysis within four groups of patients: 1 – not on mechanical ventilation (*Fig. 3C*); 2 – on mechanical ventilation (*Fig. 3E*); 3 – those of them who subsequently recovered (*Fig. 3D*); or (4) died (*Fig. 3F*), revealed a correlation (from moderate to strong) in all groups between the CRP level and the degree of lung involvement assessed by CT. The CRP concentration in inflammatory diseases, including various pneumonias, was shown to correlate with the inflammation level and unaffected by factors such as age, gender, and the physical condition of the patient. CRP can be used to diagnose COVID-19 because the diagnostic sensitivity of CT alone is 76.4%, and CRP can detect inflammation in early pneumonia [40].

In groups of patients on mechanical ventilation, we found a significant correlation between the leukocyte

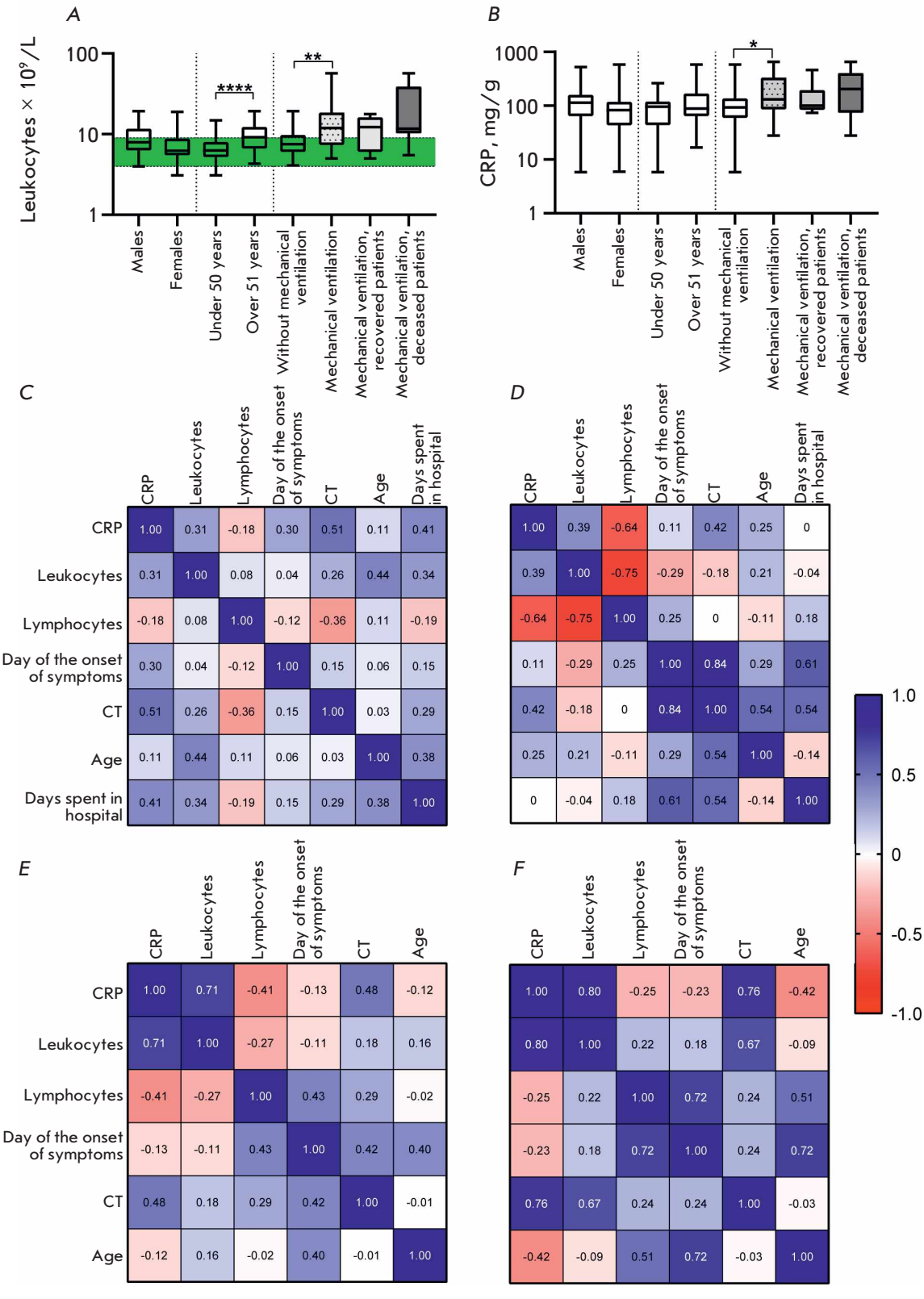
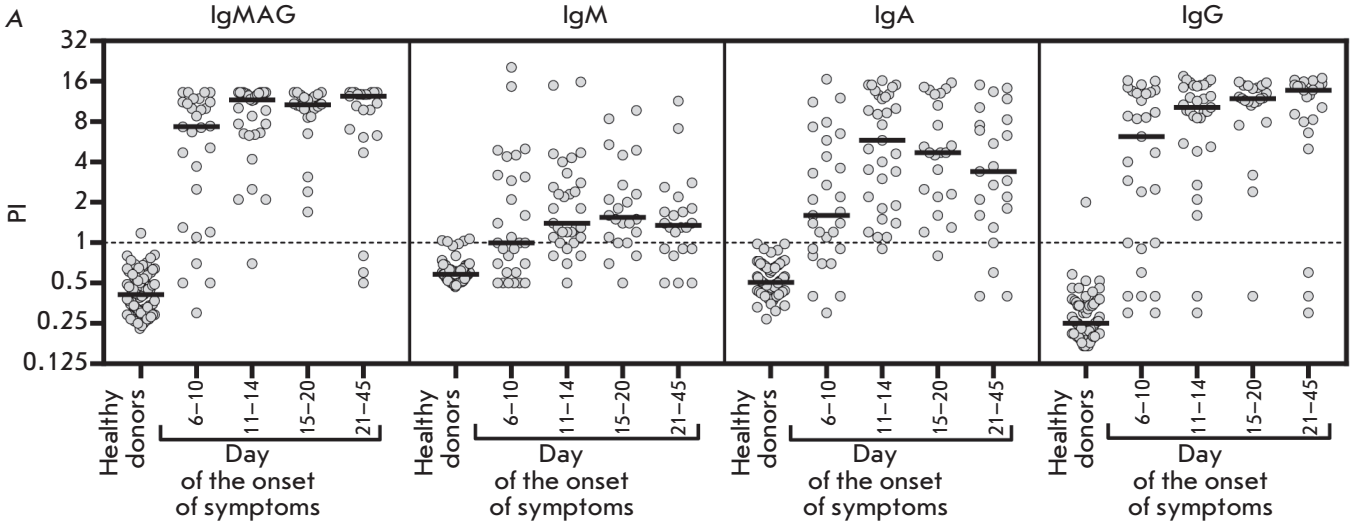


Fig. 3. Distribution of the leukocyte count (*A*) and C-reactive protein level (*B*) within groups of patients of different ages, gender, and disease severity. The interval of normal values is marked in green. (*C–F*) – correlations for groups of patients who did not need mechanical ventilation (*C*), who needed mechanical ventilation ($n = 16$) (*E*), with subsequent recovery (n = 7) (*D*) or death (n = 9) (*F*)

A



B

Serum samples		Classes of detected antibodies					
		IgMAG	IgGM	IgGA	IgM	IgA	IgG
Healthy donors (n = 70)		1.4	5.6	0.0	1.4	1.4	6.9
Patients with a verified diagnosis of COVID-19, days after the onset of symptoms (n = 105)	6–10	83.0	72.4	79.3	51.7	68.9	72.4
	11–14	96.8	93.5	96.8	80.6	96.8	93.5
	15–20	100.0	100.0	100.0	81.8	95.5	95.5
	> 20	86.9	86.9	86.9	48.2	82.6	86.9
	> 6	93.8	91.5	92.9	58.8	79.0	90.5

Fig. 4. Results of serodiagnostic ELISA tests of blood sera from healthy donors and patients with a confirmed diagnosis of COVID-19 at different times after the onset of symptoms. (*A*) – individual values of the positivity index of test samples calculated upon detection of SARS-CoV-2-specific IgM, IgA, and IgG antibodies, separately or simultaneously. A mixture of recombinant proteins, SARS-CoV-2 S-protein RBD-SD1 and NTD fragments, and the recombinant N-protein was used as antigens. The sample positivity index was calculated as the sample signal to mean signal ratio for healthy donor samples ($n = 70$) + 3 standard deviations. The threshold value (PI = 1) is marked with a dashed line. (*B*) – number of samples exceeding the threshold value (expressed as %) for one or more of the indicated SARS-CoV-2-specific classes of antibodies

count and CRP, moderate upon further recovery and strong upon death, which may indicate an intense inflammatory process.

The probability of two diametrically opposite outcomes of COVID-19 in severe patients on ventilation may be assessed through a correlation analysis. In recovered patients, there is further a strong correlation ($r = 0.84$) between the number of days after symptoms onset and the severity of lung involvement. Perhaps, due to impaired early antiviral immunity in these patients, a SARS-CoV-2 infection persists for a while and gradually increases the degree of damage to the lung tissue, until the patient is hospitalized due to symptoms associated with lung damage. In-hospital treatment, including mechanical ventilation, helps resolve the viral infection.

Currently, studies on humoral responses of adaptive immunity in a SARS-CoV-2 coronavirus infection are under way to determine whether there is a connection between the body's immunological reactions and different scenarios of disease course, as well as the influence of various factors on them (gender, age, comorbidities, etc.). The inconsistency of data obtained over the past year necessitates further accumulation and a large-scale analysis. We compared qualitative and quantitative parameters of the B-cell immune response in different groups of patients diagnosed with COVID-19. Changes in the immune response

were assessed by ELISA of blood serum samples from 155 patients with a confirmed diagnosis of COVID-19; of these, 105 patients were hospitalized at different times after the onset of symptoms. As antigens, we used a mixture of recombinant proteins, SARS-CoV-2 S-protein fragments (RBD-SD1 and NTD), and the recombinant N-protein, which were produced in the prokaryotic *E. coli* system and adsorbed in denatured state to plate wells.

The assay results (*Fig. 4A*), expressed as the distribution of a calculated sample positivity index (PI) depending on the number of days after the onset of symptoms, revealed differences in the timing of the emergence of antibodies specific to the used SARS-CoV-2 fragments, depending on the time after the onset of symptoms. For class M, G, and A immunoglobulins, the median positivity index exceeding the threshold value (PI = 1) was reached on day 6 after the onset of symptoms. The maximum values were detected on day 11–14 for class A immunoglobulins, day 15–20 for class M immunoglobulins, and day 20 for class G immunoglobulins, which is consistent with the data obtained using other test systems [8, 41]. The maximum sensitivity of ELISA detection of IgG antibodies using our test system reached 95.5% in a range of 15–20 days after the onset of symptoms (*Fig. 4B*). In the case of the IgM and IgA antibodies, the maximum sensitivity of 81.8 and 96.7% was observed within 11–14 and 15–20 days after the onset of symptoms, respectively, and then it decreased, remaining significantly higher in the case of immunoglobulins A. A decrease in the sensitivity of detection of IgM and IgA antibodies by ELISA may be explained by a gradual decline in the levels of these antibodies in the bloodstream at a later follow-up period [42, 43]. The highest ELISA sensitivity (more than 93.8%) and specificity (98.6%) of detection of SARS-CoV-2-specific antibodies throughout the study period was achieved upon determination of total immunoglobulins M, G, and A. The sensitivity of detection of IgM, IgA, and IgG antibodies was slightly lower and amounted to more than 58.8, 79, and 90.5%, respectively. A ROC analysis was used to compare the diagnostic value of the tests at selected threshold levels. The AUC indicator was 0.93 (95% CI: 0.90–0.96) for a IgA analysis, 0.87 (95% CI: 0.83–0.92) for a IgM analysis, and 0.95 (95% CI: 0.93–0.98) for a IgG analysis. Since IgM and IgA antibodies have a similar timing of emergence and disappearance in the bloodstream, and the absolute values of sample positivity indices and the calculated sensitivity and diagnostic significance of ELISA are significantly higher for IgA antibodies than for IgM antibodies, it may be argued that detection of class A immunoglobulins is more reasonable for a diagnosis of COVID-19.

To determine the contribution of each antigen to the ELISA sensitivity at different times after the onset of symptoms, we evaluated the level of antibodies specific to each of the antigens separately. As antigens in the analysis, we used S-protein RBD-SD1 and NTD fragments and the N-protein produced in the prokaryotic *E. coli* system and adsorbed in denatured state to plate wells. Similarly, the produced RBD fragment (RBD *E. coli*) was used to assess the contribution of SD1-specific immunoglobulins to the ELISA sensitivity. Recombinant RBD produced in the eukaryotic CHO system (RBD CHO) was used as an antigen representing the conformational RBD epitopes. The assay results (*Fig. 5A,C,E*) reveal a different timing of the emergence of antibodies, which depends on the antigen nature and the time after the onset of symptoms. The median positivity indices of N- and RBD (CHO)-specific class M and A immunoglobulins exceeded the threshold values on day 6 after the onset of symptoms, reached maximum values by day 11–14 in the case of RBD (CHO)-specific IgM antibodies and day 15–20 in other cases, and decreased after 3 weeks of observation. In the case of the antigens representing linear epitopes of the S-protein (RBD *E. coli*, RBD-SD1, and NTD), the number of seropositive patients in each time range did not exceed 10%, which did not allow the median positivity indices of immunoglobulins specific to these antigens to exceed the threshold. The seroconversion rate of SARS-CoV-2 N-protein-specific class G immunoglobulins is significantly higher than that of antibodies of other specificity; the median level of N-specific antibodies significantly exceeded the threshold value as early as on day 6 after the onset of symptoms, reaching a maximum on the second week. At the same time, the median level of RBD (CHO)-specific conformation-dependent antibodies exceeded the threshold by the second week after the onset of symptoms, reaching its maximum within 21–45 days.

For IgG antibodies specific to NTD and RBD-SD1 antigens containing linear epitopes, the threshold value was exceeded only on the third week after the onset of symptoms. Thus, the seroconversion rate of early IgM and IgA antibodies is somewhat higher for antibodies specific mainly to the conformational RBD fragment epitopes than for N-specific antibodies. Conversely, the seroconversion rate of IgG antibodies decreased in the series of N-, conformation-dependent RBD (CHO)-, and conformation-independent RBD-SD1/NTD-specific antibodies. According to the obtained data (*Fig. 5*), the N-protein has the highest immunogenicity, as described earlier [44], while the linear RBD and NTD epitopes have the least immunogenicity. Thus, strong immunogenicity of the RBD fragment, reported

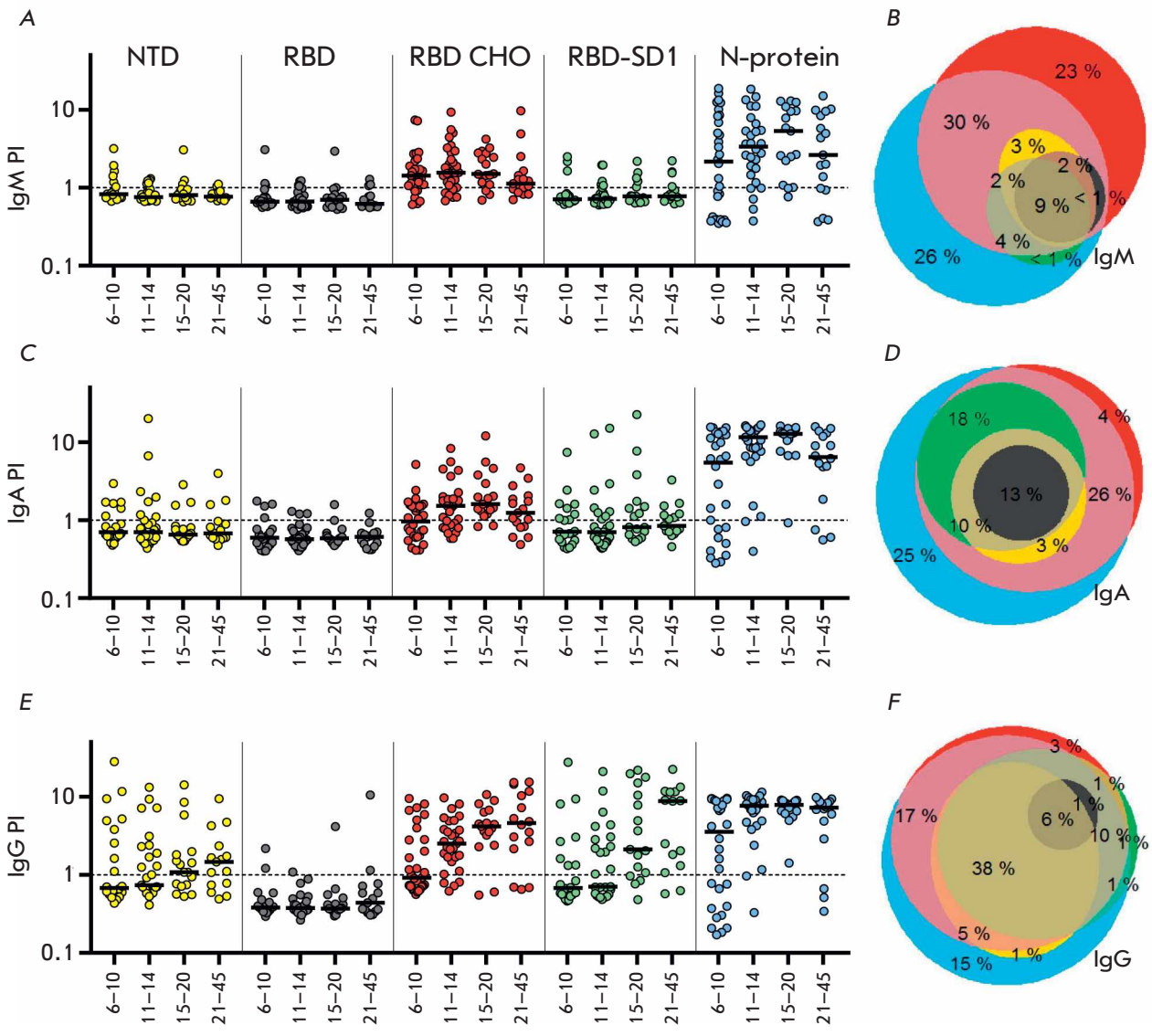


Fig. 5. Results of serodiagnostic ELISA tests of blood serum samples from patients with a confirmed diagnosis of COVID-19 hospitalized at various times after the onset of symptoms. (*A*, *C*, *E*) – sample positivity index calculated upon detection of SARS-CoV-2 S-protein NTD, RBD, and RBD-SD1 fragment- and N-protein-specific IgM (*A*), IgA (*C*), and IgG (*E*) antibodies. (*B*, *D*, *F*) – Venn diagrams representing antigen-specificity spectra of IgM (*A*), IgA (*C*), and IgG (*E*) immunoglobulins in samples

previously [45], is mainly associated with conformational epitopes. Linear SD1 subdomain epitopes have strong but slowly developing immunogenicity, which may be especially important in light of the data on the existence of neutralizing antibodies specific to a linear epitope located in this region [45]. The spectra of antigen specificity were found to differ for class M, A, and G immunoglobulins (*Fig. 5B,D,E*). The number of seropositive patients with blood antibodies specific to only one "strong" immunogen was found to decrease in the series of class M, A, and G immunoglobulins and accounted for 49%, 29%, and 19% of the total seropositive patients, respectively. These data indicate the need to use at least two antigens in ELISA for the diagnosis of COVID-19 to improve assay sensitivity, especially at an early stage of the disease.

The available data demonstrating the influence of age on the B-cell immune response (in particular, on

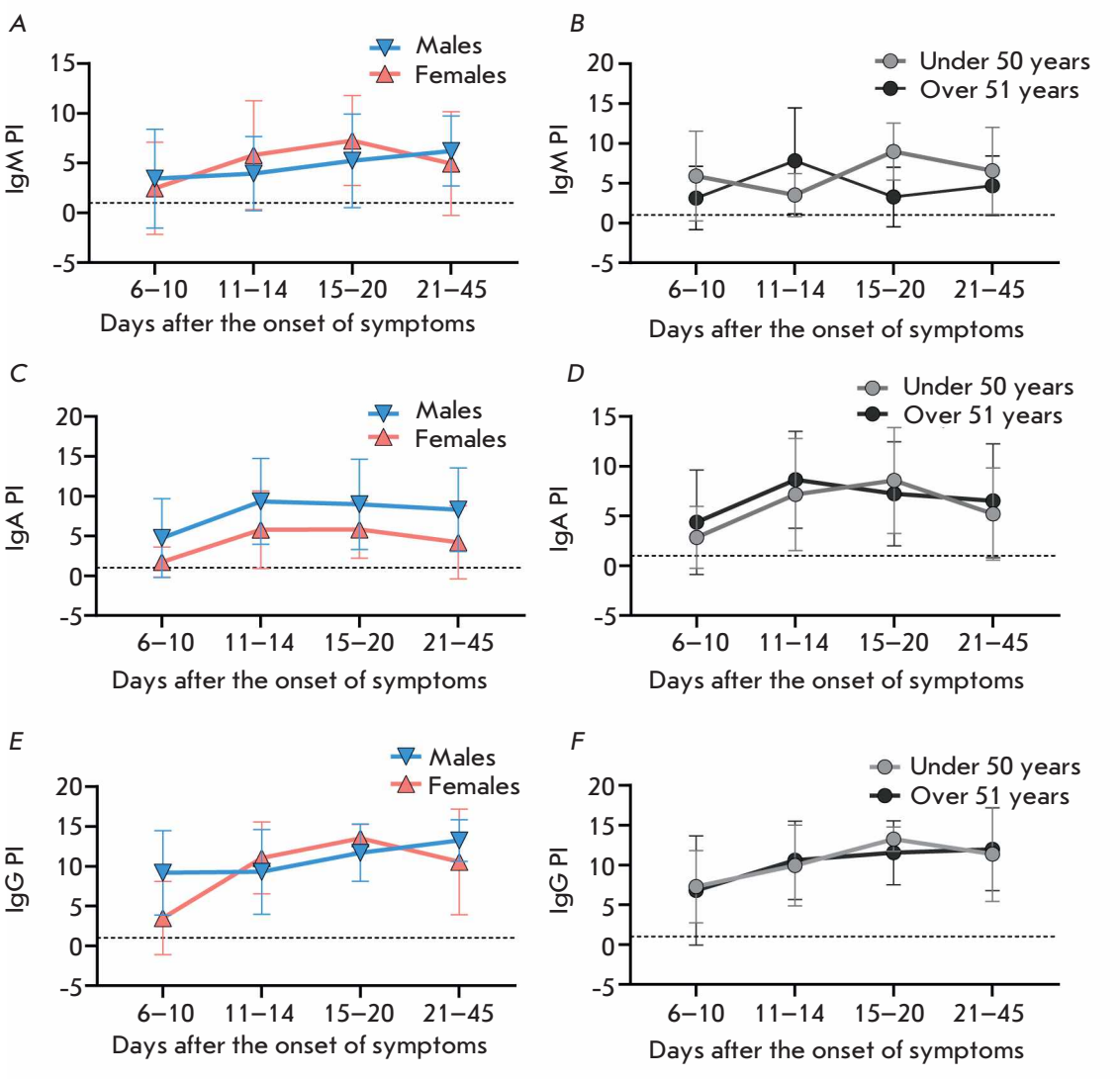


Fig. 6. Distribution of the positivity index for the blood samples of patients, depending on days after the onset of symptoms in groups of males and females (*A*, *C*, *E*) and groups of patients of different ages (under 50 years and over 51 years) (*B*, *D*, *F*), calculated upon detection of SARS-CoV-2-specific class M (*A*, *B*), A (*C*, *D*), and G (*E*, *F*) immunoglobulins

the rate of seroconversion and the titer of immunoglobulins) in COVID-19 patients remain inconclusive. A number of studies in elderly patients have reported a higher titer of antibodies of all classes [8, 46, 47]; however, there are studies that have reported no relationship between age and the B-cell response [48, 49]. There is no evidence of an effect of gender on the level of SARS-CoV-2-specific antibodies [47, 48].

We compared the blood levels of class M, A, and G immunoglobulins in patients of different age and gender groups from the cohort of hospitalized patients at different times after the onset of symptoms. For this purpose, we used ELISA and a mixture of recombinant proteins, SARS-CoV-2 S-protein RBD-SD1 and NTD fragments, and the recombinant N-protein as antigens. Comparison of PI values at each time interval using the nonparametric Mann–Whitney test did not reveal significant differences in the seroconversion rate between the study groups (*Fig. 6*).

To identify differences in the levels of class M, A, and G antibodies specific to different SARS-CoV-2 virus fragments in patients with differing severity of COVID-19, a group of outpatients (n = 50) who had had mild symptoms was additionally included in the

A

Disease severity	IgM					IgA					IgG				
	NTD	RBD *E. coli*	RBD CHO	RBD-SD1	N	NTD	RBD *E. coli*	RBD CHO	RBD-SD1	N	NTD	RBD *E. coli*	RBD CHO	RBD-SD1	N
Mild symptoms	5	0	19	5	19	0	2	35	16	40	23	5	88	56	84
Hospitalization	15	15	74	26	79	18	6	77	44	85	59	9	82	77	88
Mechanical ventilation	6	13	38	6	75	25	19	44	25	75	25	0	50	38	69
Mechanical ventilation, recovered patients	14	29	43	14	100	43	29	86	43	86	43	0	86	57	86
Mechanical ventilation, deceased patients	0	0	33	0	56	22	11	11	11	67	11	0	22	11	56

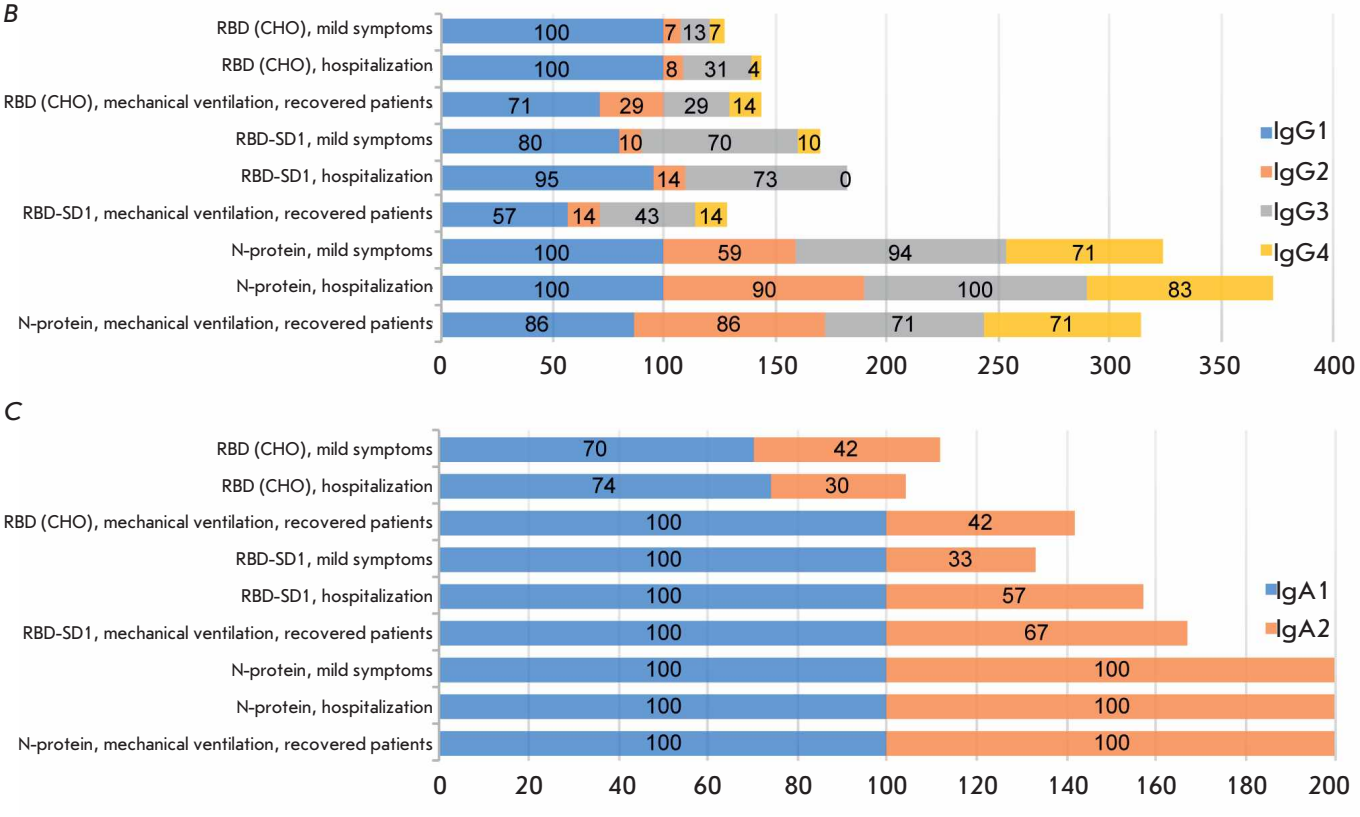


Fig. 7. Changes in the occurrence rate of patients seropositive for immunoglobulins of various classes and subclasses specific to SARS-CoV-2 antigens in groups of patients with differing severity of COVID-19. (*A*) – occurrence rate (%) of patients seropositive for class M, A, and G immunoglobulins specific to NTD, RBD, RBD-SD1 antigens and the N-protein in the groups of patients with differing severity of COVID-19. (*B*) – occurrence rate (%) of patients seropositive for subclass G and A (*C*) immunoglobulins specific to RBD (CHO) and RBD-SD1 antigens and the N-protein in groups of patients with differing severity of COVID-19

cohort of patients who needed (or did not need) mechanical ventilation (mean time to hospitalization 21.2 days). Since the blood level of antibodies depends on the time from the onset of symptoms, for an accurate comparison, the hospitalized group included patients admitted 15–45 days (mean 21.8 days) after the onset of symptoms.

An analysis of the occurrence rate of patients seropositive for class M or A immunoglobulins specific to one or more of the used antigens revealed a significant

decrease in the rate in the group with mild symptoms compared to that of hospitalized patients seropositive for each antigen (*Fig. 7A*). In this group, there was also a decrease in the occurrence rate of class G immunoglobulins specific to linear RBD, NTD, and RBD-SD1 epitopes, while the occurrence rate of patients seropositive for RBD (CHO)- and N-protein-specific class G immunoglobulins did not change. In the group of patients on mechanical ventilation, the rate of patients seropositive for one or more classes of the antibodies under consideration was also reduced. However, this decrease was associated with a significant reduction in the number of seropositive patients in the subgroup of fatal cases, while these characteristics were similar in the subgroup of recovered patients and in the group of hospitalized patients.

An analysis of the levels of specific IgM, IgA, and IgG antibodies in serum-positive blood sera using the nonparametric Mann–Whitney test did not reveal a statistically significant effect of disease severity on the levels of SARS-CoV-2-specific antibodies. These results contradict some data indicating that the blood levels of immunoglobulins of various classes in severe patients are increased, while the content of antibodies is reduced in the group of asymptomatic or mild-symptoms patients [5, 8, 48, 50].

An analysis of the occurrence rate ratio of IgA and IgG subclasses in sera of appropriate seropositive samples (*Fig. 7B,C*) reveals a uniform induction of N-protein-specific immunoglobulin G subclasses G1–G4 and immunoglobulin A subclasses A1–A2 in groups of patients with differing severity of COVID-19, while G1, G3, and A1 are the main subclasses in the immune response to the S antigen. At more severe symptoms, the occurrence rate of S antigen-specific IgG1 antibodies is decreased, while that of IgA2, on the contrary, is increased. However, a correlation between the levels of the studied SARS-CoV-2-specific antibodies and disease severity has not been reliably established.

CONCLUSION

By using the developed ELISA diagnostic kit based on recombinant antigens – SARS-CoV-2 virus protein fragments, we have reliably established the advantage of class A immunoglobulins as an early immunological criterion of the development of the disease. The spectrum of specificity of SARS-CoV-2-induced immunoglobulins in each patient depends on the time after infection and varies in the series of M, A, and G immunoglobulins from narrow to wide. We have also shown uneven induction of immunoglobulin subclasses, which depends on the antigen nature. The N-protein induces immunoglobulins G1–G4 and A1–A2 in equal proportions, while G1, G3, and A1 are the main subclasses in the immune response to the S-antigen. The ratio between N-specific subclasses remains almost unchanged in groups of patients with differing severity of COVID-19, but with a more severe course of the disease, the occurrence rate of S-specific IgG1 antibodies decreases, while that of IgA2, on the contrary, increases. However, no reliable correlation between the levels of the studied SARS-CoV-2-specific antibodies and disease severity has been revealed. ●

This work was supported by a grant RSF 17-74-30019 and by a grant ICGEB CRP/RUS18-01

REFERENCES

1. Xiuyuan Ou, Yan Liu, Xiaobo Lei, Pei Li, Dan Mi, Lili Ren, Li Guo, Ruixuan Guo, Ting Chen, Jiaxin Hu, Zichun Xiang, et al. // Nat. Commun. 2020. V. 1. № 1. P. 1620. doi: 10.1038/s41467-020-15562-9.
2. Xiaofan Liu, Hong Zhou, Yilu Zhou, Xiaojun Wu, Yang Zhao, Yang Lu, Weijun Tan, Mingli Yuan, Xuhong Ding, Jinjing Zou, et al. // J. Infect. 2020 V. 81. № 1. P. e95-e97. doi: 10.1016/j.jinf.2020.04.008. Epub. 2020 Apr. 17.
3. Ahnach M., Zbiri S., Nejjari S., Ousti F., Elkettani C. // J. Med. Bioch. 2020. V. 39. P. 1–8.
4. Barek A., Aziz A., Islam M.S. // Heliyon. 2020. V. 6. P. e05684.
5. Chuan Qin, Luoqi Zhou, Ziwei Hu, Shuoqi Zhang, Sheng Yang, Yu Tao, Cuihong Xie, Ke Ma, Ke Shang, Wei Wang, et al. // China Clin. Infect. Dis. 2020. V. 10. 1093/cid/ciaa248.
6. Lavillegrand J.R., Garnier M., Spaeth A., Mario N., Hariri G., Pilon A., Berti E., Fieux F., Thietart S., Urbina T., et al. // Ann. Intensive. Care. 2021. V. 11. P. 9.
7. Yamada T., Wakabayashi M., Yamaji T., Chopra N., Mikami T., Miyashita H., Miyashita S. // Clin. Chim. Acta. 2020. V. 509. P. 235.
8. Huan Ma, Weihong Zeng, Hongliang He, Dan Zhao, Dehua Jiang, Peigen Zhou, Linzhao Cheng, Yajuan Li, Xiaoling Ma, Tengchuan Jin. // Cell. Mol. Immunol. 2020. V. 17. P. 773–775.
9. Debnath M., Banerjee M., Berk M. // FASEB J. 2020. V. 34. № 7. P. 8787–8795. doi: 10.1096/fj.202001115R.
10. Mei-Shang Ho, Wei-Ju Chen, Hour-Young Chen, Szu-Fong Lin, Min-Chin Wang, Jiali Di, Yen-Ta Lu, Ching-Lung Liu, Shan-Chwen Chang, Chung-Liang Chao, et al. // Emerg. Infect. Dis. 2005. V. 11. № 11. P. 1730–1737. doi: 10.3201/eid1111.040659.
11. Li Guo, Xi Zhang, Lili Ren, Xuelian Yu, Lijuan Chen, Hongli Zhou, Xin Gao, Zheng Teng, Jianguo Li, Jiayu Hu, et al. // Emerg. Infect. Dis. 2014. V. 20. № 2. P. 192–200. doi: 10.3201/eid2002.131094.
12. Kelvin Kai-Wang To, Owen Tak-Yin Tsang, Wai-Shing

Leung, Anthony Raymond Tam, Tak-Chiu Wu, David Christopher Lung, Cyril Chik-Yan Yip, Jian-Piao Cai, Jacky Man-Chun Chan, Thomas Shiu-Hong Chik, et al. // Lancet Infec. Dis. 2020. V. 20. № 5. P. 565–574. doi: 10.1016/S1473-3099(20)30196-1.
13. Belogurov A., Kozyr A., Ponomarenko N., GabibovA. // Bioessays. 2009. V. 31. № 11. P. 1161–1171. doi: 10.1002/bies.200900020
14. LiLi Ren, Ye-Ming Wang, Zhi-Qiang Wu, Zi-Chun Xiang, Li Guo, Teng Xu, Yong-Zhong Jiang, Yan Xiong, Yong-Jun Li, Xing-Wang Li, et al. // Chin. Med. J. (Engl.) 2020. V. 133. № 9. P. 1015–1024. doi: 10.1097/CM9.0000000000000722.
15. Kuldeep Dhama, Khan Sharun, Ruchi Tiwari, Maryam Dadar, Yashpal Singh Malik, Karam Pal Singh Wanpen Chaicumpa // Hum. Vaccin. Immunother. 2020. V. 16. № 6. P. 1232–1238. doi: 10.1080/21645515.2020.1735227.
16. Li Guo, Lili Ren, Siyuan Yang, Meng Xiao, De Chang, Fan Yang, Charles S Dela Cruz, Yingying Wang, Chao Wu, Yan Xiao, et al. // Clin. Infect Dis. 2020. V. 71. № 15. P. 778–785. doi: 10.1093/cid/ciaa310.
17. Logunov D.Y., Dolzhikova I.V., Zubkova O.V., Tukhvatulin A.I., Shcheblyakov D.V., Dzharullaeva A.S., Grousova D.M., Erokhova A.S., Kovyrshina A.V., Botikov A.G., et al. // Lancet. 2020. 396. V. 10. № 255. P. 887–897. doi: 10.1016/S0140-6736(20)31866-3.
18. Baden L.R., El Sahly H.M, Essink B., Kotloff K., Frey S., Novak R., Diemert D., Spector S.A., Rouphael N., Creech C.B., et al. // N. Engl. J. Med. 202. V. 384. № 5. P. 403–416. doi: 10.1056/NEJMoa2035389.
19. van Doremalen N., Lambe T., Spencer A., Belij-Rammerstorfer S., Purushotham J.N., Port J. R., Avanzato V., Bushmaker T., Flaxman A., Ulaszewska M. // bioRxiv. 2020. V. 2020.05.13.093195. doi: 10.1101/2020.05.13.093195. Preprint
20. Calina D., Docea A.O., Petrakis D., Egorov A.M., Ishmukhametov A.A., Gabibov A.G., Shtilman M.I., Kostoff R., Carvalho F., Vinceti M., et al. // Int. J. Mol. Med. 2020. V. 46. № 1. P. 3–16. doi: 10.3892/ijmm.2020.4596.
21. Sinegubova M.V., Orlova N.A., Kovnir S.V., Dayanova L.K., Vorobiev I.I. // PLoS One. 2021. V. 16. № 2. P. e0242890. doi: 10.1371/journal.pone.0242890. eCollection 2021.
22. Durova O., Vorobiev I., Smirnov I., Reshetnyak A., Telegin G., Shamborant O., Orlova N., Genkin D., Bacon A., Ponomarenko N., et al. // Mol. Immunol. 2009. V. 47. № 1. P. 87–95. doi: 10.1016/j.molimm.2008.12.020.
23. Ahnach M., Zbiri S., Nejjari S., Ousti F., Elkettani C. // J. Med. Biochem. 2020. V. 9. № 4. P. 500–507. doi: 10.5937/jomb0-27554.
24. Stefan N., Birkenfeld A.L., Schulze M.B. // Nat. Rev. Endocrinol. 2021. V. 17. № 3. P. 135–149. doi: 10.1038/s41574-020-00462-1.
25. Kaochang Zhao, Ruiyun Li, Xiaojun Wu, Yang Zhao, Tao Wang, Zhishui Zheng, Shaolin Zeng, Xuhong Ding, Hanxiang Nie. // Eur. J. Clin. Microbiol. Infect. Dis. 2020. V. 39. № 12. P. 2279–2287. doi: 10.1007/s10096-020-03976-8.
26. Riley L.K., Rupert J. // Am. Fam. Physician. 2015. V. 92. № 11. P. 1004–1011.
27. Shapiro M.F., Greenfield S. // Ann. Intern. Med. 1987. V. 106. № 1. P. 65–74. doi: 10.7326/0003-4819-106-1-65.
28. Mei Y., Weinberg S.E., Zhao L., Frink A., Qi C., Behdad A., Ji P. // E. Clin. Med. 2020. V. 26. P. 100475. doi: 10.1016/j.eclinm.2020.100475. eCollection 2020 Sep.
29. Takayuki Yamada, Mako Wakabayashi, Takahiro Yamaji, Nitin Chopra, Takahisa Mikami, Hirotaka Miyashita, Satoshi Miyashita. // Clin. Chim. Acta. 2020. V. 509. P. 235–243. doi: 10.1016/j.cca.2020.06.008.
30. Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, et al. // Lancet. 2020. V. 395. № 10223. P. 497–506. doi: 10.1016/S0140-6736(20)30183-5.
31. Barnaby Edward Young, Sean Wei Xiang Ong, Shirin Kalimuddin, Jenny G Low, Seow Yen Tan, Jiashen Loh, Oon-Tek Ng, Kalisvar Marimuthu, Li Wei Ang, Tze Minn Mak, et al. // JAMA. 2020. V. 323. № 15. P. 1488–1494. doi: 10.1001/jama.2020.3204.
32. Wei-Jie Guan, Zheng-Yi Ni, Yu Hu, Wen-Hua Liang, Chun-Quan Ou, Jian-Xing He, Lei Liu, Hong Shan, Chun-Liang Lei, David S C Hui, et al. // N. Engl. J. Med. 2020. V. 382. № 18. P. 1708–1720. doi: 10.1056/NEJMoa2002032.
33. Yan Deng, Wei Liu, Kui Liu, Yuan-Yuan Fang, Jin Shang, Ling Zhou, Ke Wang, Fan Leng, Shuang Wei, Lei Chen, et al. // Chin. Med. J. (Engl.). 2020. V. 133 № 11. P. 1261–1267. doi: 10.1097/CM9.0000000000000824.
34. Sharifpour M., Rangaraju S., Liu M., Alabyad D., Nahab F.B., Creel-Bulos C. M., Jabaley C.S.// PLoS One. 2020. V. 15. № 11. P. e0242400. doi: 10.1371/journal.pone.0242400. eCollection 2020.
35. Fang Liu, Lin Li, Meng Da Xu, Juan Wu, Ding Luo, Yu Si Zhu, Bi Xi Li, Xiao Yang Song, Xiang Zhou. // J. Clin. Virol. 2020. V. 127. P. 104370. doi: 10.1016/j.jcv.2020.104370.
36. Zhang Yitao, Chen Mu, Zhou Ling, Cheng Shiyao, Xue Jiaojie, Chen Zhichong, Peng Huajing, Ou Maode, Cheng Kanglin, Ou Yang Mao, et al. // Curr. Med. Res. Opin. 2021. V. 4. P. 1–14. doi: 10.1080/03007995.2021.1876005. Online ahead of print.
37. Chaochao Tan, Ying Huang, Fengxia Shi, Kui Tan, Qionghui Ma, Yong Chen, Xixin Jiang, Xiaosong Li. // J. Med. Virol. 2020. V. 92. № 7. P. 856–862. doi: 10.1002/jmv.25871.
38. Takatoshi Higuchi, Tsutomu Nishida, Hiromi Iwahashi, Osamu Morimura, Yasushi Otani, Yukiyoshi Okauchi, Masaru Yokoe, Norihiro Suzuki, Masami Inada, Kinya Abe. // J. Med. Virol. 2021. V. 93. P. 2141–2148. doi: 10.1002/jmv.26599.
39. Jialin Xiang, Jing Wen, Xiaoqing Yuan, Shun Xiong, Xue Zhou, Changjin Liu1, Xun Min // MedRxiv. 2020. Preprint doi: https://doi.org/10.1101/2020.03.19.20034447
40. Ling W. // Med. Mal. Infect. 2020. V. 50. № 4. P. 332–334. doi: 10.1016/j.medmal.2020.03.007.
41. Grzelak L., Temmam S., Planchais C., Demeret C., Tondeur L., Huon Ch., Guivel-Benhassine F., Staropoli I., Chazal M., Dufloo J., et al. // Sci. Transl. Med. 2020. V. 12. № 559. P. eabc3103. doi: 10.1126/scitranslmed.abc3103.
42. Lili Ren, Lulu Zhang, De Chang, Junwen Wang, Yongfeng Hu, Hong Chen, Li Guo, Chao Wu, Conghui Wang, Yingying Wang, et al. // Commun. Biol. 2020. V. 3. № 1. P. 780. doi: 10.1038/s42003-020-01526-8.
43. Bin Lou, Ting-Dong Li, Shu-Fa Zheng, Ying-Ying Su, Zhi-Yong Li, Wei Liu, Fei Yu, Sheng-Xiang Ge, Qian-Da Zou, Quan Yuan, et al. // Eur. Respir. J. 2020. V. 56. № 2. P. 2000763. doi: 10.1183/13993003.00763-2020.
44. Burbelo P.D., Riedo F.X., Morishima Ch., Rawlings S., Smith D., Das S., Strich J.R., Chertow D.S., Davey R.T., Cohen J. // J. Infect. Dis. 2020. V. 222. № 2. P. 206–213. doi: 10.1093/infdis/jiaa273.
45. Chek Meng Poh, Guillaume Carissimo, Bei Wang, Siti Naqiah Amrun, Cheryl Yi-Pin Lee, Rhonda Sin-Ling Chee, Siew-Wai Fong, Nicholas Kim-Wah Yeo, Wen-Hsin Lee, Anthony Torres-Ruesta, et al. // Nat. Commun. 2020. V. 11. № 1. P. 2806. doi: 10.1038/s41467-020-16638-2.
46. Madariaga M.L.L., Guthmiller J.J., Schrantz S., Jansen M.O., Christensen C., Kumar M., Prochaska M., Wool G.,

Durkin-Celauro A., Oh W.H., et al. // J. Intern. Med. 2020. Oct 9; 10.1111/joim.13185. doi: 10.1111/joim.13185.
47. Garcia-Basteiro A.L., Moncunill G., Tortajada M., Vidal M., Guinovart C., Jiménez A., Santano R., Sanz S., Méndez S., Llupià A., et al.// Nat. Commun. 2020. V. 11. № 1. P. 3500. doi: 10.1038/s41467-020-17318-x.
48. Nilsson C.J., Zurbuchen Y., Valaperti A., Schreiner J., Wolfensberger A., Raeber M.E., Adamo S., Weigang S., Emmenegger M., et al. // J. Allergy Clin. Immunol. 2021. V. 147. № 2. P. 545–557.e9. doi: 10.1016/j.jaci.2020.10.040. Carlo
49. Juan Chen, Zhen-Zhen Zhang, Yao-Kai Chen, Quan-Xin Long, Wen-Guang Tian, Hai-Jun Deng, Jie-Li Hu, Xian-Xiang Zhang, Pu-Liao, Jiang-Lin Xiang, et al. // Genes Dis. 2020. V. 7. № 4. P. 535–541. doi: 10.1016/j.gendis.2020.03.008.
50. Lynch K.L, Whitman J.D., Lacanienta N.P., Beckerdite E.W., Kastner Sh.A, Shy B.R., Goldgof G.M., Levine A.G., Bapat S.P., Stramer S.L., et al. // Clin. Infect. Dis. 2020. ciaa979. doi: 10.1093/cid/ciaa979.

Suppression of the Immune Response by Syngeneic Splenocytes Adoptively Transferred to Sublethally Irradiated Mice

A. A. Kalinina[1], L. M. Khromykh[1], D. B. Kazansky[1], A. V. Deykin[2,3], Yu. Yu. Silaeva[2*]

[1]Federal State Budgetary Institution "N.N. Blokhin National Medical Research Center of Oncology" of the Ministry of Health of the Russian Federation, Moscow, 115478 Russia

[2]Core Facility Centre, Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia

[3]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia

*E-mail: silaeva@genebiology.ru, yulya.silaeva@gmail.com

Received November 05, 2020; in final form, January 20, 2021

DOI: 10.32607/actanaturae.11252

ABSTRACT **The peripheral T-cell pool consists of several, functionally distinct populations of $CD8^+$ T cells. CD44 and CD62L are among the major surface markers that allow us to define T-cell populations. The expression of these molecules depends on the functional status of a T lymphocyte. Under lymphopenic conditions, peripheral T cells undergo homeostatic proliferation and acquire the memory-like surface phenotype $CD44^{hi}CD62L^{hi}$. However, the data on the functional activity of these cells remains controversial. In this paper, we analyzed the effects of the adoptive transfer of syngeneic splenocytes on the recovery of $CD8^+$ T cells in sublethally irradiated mice. Our data demonstrate that under lymphopenia, donor lymphocytes form a population of memory-like $CD8^+$ T cells with the phenotype $CD122^+CD5^+CD49d^{hi}CXCR3^+$ that shares the phenotypic characteristics of true memory cells and suppressive $CD8^+$ T cells. *Ex vivo* experiments showed that after adoptive transfer in irradiated mice, T cells lacked the functions of true effector or memory cells; the allogeneic immune response and immune response to pathogens were greatly suppressed in these mice.**

KEYWORDS **memory-like T cell, lymphopenia, homeostatic proliferation, CD44, CD62L.**

ABBREVIATIONS **MHC – major histocompatibility complex; T_{ML} – memory-like T cell; MLR – mixed lymphocyte reaction; cpm – counts per minute; AT – adoptive transfer; TCR – T-cell receptor.**

INTRODUCTION

The peripheral T-cell pool is comprised of several functionally distinct $CD8^+$ T-cell populations. The major surface markers of these populations are CD44 and CD62L, whose expression defines the activation phenotype and the migration properties of a T cell. CD62L mediates the interaction between a T lymphocyte and cells of the high endothelium venules, as well as its migration within the lymphoid system. CD44, the receptor for hyaluronic acid in the extracellular matrix, allows T lymphocytes to leave the lymphoid system and migrate to the peripheral tissues [1]. The expression profile of these markers varies depending on the functional state of T lymphocytes. Naive T cells have the surface phenotype $CD62L^{hi}CD44^{lo}$; CD8 clones activated during the primary immune response lose the CD62L expression and become $CD62L^{lo}CD44^{hi}$. Most CD8 effectors die after completion of their role in the immune response; a small portion of them forms a population of long-living memory T cells capable of maintaining a stable pool in the absence of the specific antigen and accelerated immune response to the specific antigen.

Long-living memory CD8 T cells have the $CD44^{hi}CD62L^{hi}$ phenotype; however, this does not always correlate with the "antigenic experience" of T cells. Indeed, the peripheral T-cell pool in non-immunized gnotobiotic animals contains virtual memory T cells specific to the model antigen [2, 3]. Under lymphopenia, the peripheral T lymphocytes undergo homeostatic proliferation and acquire the surface phenotype of memory T cells: $CD44^+CD62L^+$ (T_{ML},

"memory-like" T cells) [4–7]. The T_{ML} population cannot down-regulate the expression of surface activation molecules and acquire a naive phenotype [8, 9]. Thus, this population is phenotypically similar to true memory T cells.

Our vast pool of experimental data on the functional properties of T_{ML} cells remains controversial. Several studies have shown that adoptive transfer of naive $CD8^+$ T cells under lymphopenic conditions leads to the formation of a T-cell population with the functional features of true memory cells [10, 11]. However, the localization of this population and the expression profile of the chemokine receptors on these cells differ from those of true memory cells [12]. The T_{ML} population, with immunosuppressive activity, was reported as well [13]. Moreover, under lymphopenic conditions, T-cell clones with high affinity to self MHC molecules (i.e., autoreactive T cells) proliferate and acquire a memory phenotype [14, 15]. A population of $CD8^+CD44^+CD122^+$ T cells with suppressive activity was reported in several studies [13, 16–18].

These data suggest that the surface phenotype of T lymphocytes may not reflect their actual functional status, and that the population in question could be incorrectly assigned to long-living memory $CD8^+$ T cells. In this work, we investigated the relationship between the expression of the surface markers CD44 and CD62L and the functional properties of $CD8^+$ T cells under lymphopenia. We observed that the adoptive transfer of syngeneic lymphocytes to sublethally irradiated mice suppressed the immune response in the mice, and that the effect could be at least partially mediated by T_{ML} $CD8^+$ T cells with the phenotype $CD122^+CD5^+CD49d^{hi}CXCR3^+$ acquired from the donor lymphocytes.

MATERIALS AND METHODS

Mice

C57BL/6 ($K^bI\text{-}A^bD^b$), B10.D2(R101) ($K^dI\text{-}A^dI\text{-}E^dD^b$), FVB ($K^qI\text{-}A^qI\text{-}E^qD^q$), and C57BL/6-TgN(ACTbEGFP)1Osb ($K^bI\text{-}A^bD^b$) (hereafter referred to as B6.GFP) strains were obtained from the breeding facility of the N.N. Blokhin National Medical Research Center of Oncology of the Ministry of Health of the Russian Federation (N.N. Blokhin NMRCO, Moscow, Russia). All the experimental procedures were approved by the Ethics Committee on Animal Experimentation of N.N. Blokhin NMRCO and of the Institute of Gene Biology of the Russian Academy of Sciences (Moscow, Russia).

Cell lines

The EL4 lymphoma cells were obtained from the collection of N.N. Blokhin NMRCO. The EL4 cells were transplanted intraperitoneally (i.p.) into syngeneic C57BL/6 mice ($3.0–5.0 \times 10^6$ cells/mouse) and grown as ascites for 10–14 days. Tumor cells were aseptically aspirated from the peritoneal ascites and washed three times by centrifugation (200 *g*) in a phosphate buffered saline (PBS, pH 7.4) at 4°C. Viable cells were counted after trypan blue/eosin staining in a Goryaev chamber and used for mouse immunization.

Bacterial strains and growth conditions

The *Salmonella typhimurium* virulent strain IE 147 and *Listeria monocytogenes* virulent strain EGD were received from the collection of N.F. Gamaleya National Research Center of Epidemiology and Microbiology, the Ministry of Health of the Russian Federation (N.F. Gamaleya NRCEM, Moscow, Russia). The *S. typhimurium* strain was grown overnight in an LB broth (Amresco, USA) at 37°C; tenfold serial dilutions of the culture were then seeded on SS agar (Condalab, Spain), and the colony numbers were counted as described elsewhere. The *L. monocytogenes* strain was grown overnight in BHI broth (BD, San Jose, CA) at 37°C with stirring at 185 rpm on a thermostatic shaker (Shaker-thermostat ES 20 Biosan, Latvia). The resulting culture was diluted 1 : 100 in 200 mL of BHI broth and incubated in a thermostatic shaker at 185 rpm at 37°C until the culture reached an optical density (OD 600) equal to 1.5–1.8. Bacterial titer (CFU/mL) was measured on an ULTROSPEC 10 spectrophotometer (General Electric, USA). Freshly grown cultures of *S. typhimurium* and *L. monocytogenes* were heat-inactivated (1 hr, 60°C; and 90 min, 74°C, respectively) and used in *in vitro* studies.

Immunization

B10.D2(R101) mice were immunized i.p. with 2.0×10^7 EL4 cells/mouse. Control non-immunized mice were injected with PBS. After 60 days, mice were euthanized by cervical dislocation; spleens were isolated, and cell suspensions were prepared (see below).

Irradiation of mice

Female B10.D2(R101) and C57BL/6 mice were sublethally irradiated (4.5 Gy; Agat-R therapeutic device, Russia; a Co^{60} source with an initial power of 1.9×10^{14} Bq). Mice were sacrificed on day 10 post-irradiation, and their splenocytes were used for flow cytometry analyses and *ex vivo* functional tests.

Cell suspensions

Splenocytes were homogenized in a Potter homogenizer with a conic pestle in PBS at 4°C and pelleted (200 *g*, 5 min). Red blood cells were lyzed in a lysing buffer (BD Pharmingen, USA). Mononuclear cells were washed

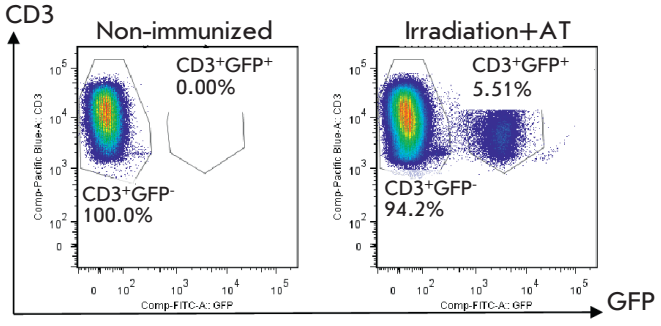


Fig. 1. The relative count of GFP^{hi} CD3$^+$ donor cells (GFP$^+$) in the spleen of C57BL/6 mice on day 10 after the sublethal irradiation and adoptive transfer. The data of one representative experiment are shown for 2.5 × 10^6 events. The data were obtained in three independent experiments, 3 mice per group

three times by centrifugation in PBS at 4°C. The cells were re-suspended in PBS for staining with monoclonal antibodies and adoptive transfer or in the complete medium for *in vitro* tests.

Adoptive transfer

Non-immunized-B10.D2(R101) mice were irradiated with 4.5 Gy. 24 h post-irradiation; mice were injected i.v. with 1.5 × 10^7 splenocytes from non-immunized or immunized syngeneic animals. Control irradiated mice received PBS as a placebo in parallel. On day 10 after the adoptive transfer, the splenocytes of the recipient mice were used as responders in *in vitro* tests. Non-immunized-C57BL/6 mice were similarly irradiated and injected with the splenocytes of non-immunized B6.GFP mice. On day 10 after the adoptive transfer, the splenocytes of the recipient mice were used for flow cytometry analyses. On day 10 after the adoptive transfer, approximately 5% of GFP$^+$ cells were detected in the spleen of irradiated mice (*Fig. 1*).

Mixed lymphocyte reaction (MLR)

The spleen cells of FVB (K^{q}I-A^{q}I-E^{q}D^q) and C57BL/6 (K^{b}I-A^{b}D^b) mice were used as non-specific and specific stimulators, respectively. The spleen cells of B10.D2(R101) mice were used as the syngeneic control. Stimulator splenocytes were treated with mitomycin C (Kyowa Hakko Kogyo Co., Ltd., Japan) (25 μg/mL, 37°C, 30 min) and washed three times in PBS by centrifugation (200*g* , 5 min, 4°C). Responders (3.0 × 10^5 cells/well) and stimulators (5.0 × 10^5 cells/well) were plated (3 : 5) in 96-well U-bottom plates (Corning Costar, Sigma Aldrich, USA) and cultured in 200 μL of a RPMI-1640 medium (PanEco, Russia) supplemented with 10% fetal bovine serum (HyClone, GE Healthcare, USA), 0.01 mg/mL ciprofloxacin (KRKA, Slovenia), 0.01 M HEPES (PanEco), and 10 mM 2-mercaptoethanol (Merck, Germany) at 37°C with 5% CO_2, for 72 h. Cell proliferation was measured by incorporation of ^{3}H- thymidine (Isotop, Russia), added in the last 8 h of culturing. The level of cell proliferative activity was expressed as the number of counts per minute (cpm).

Ex vivo immune response to pathogens

5.0 × 10^5 spleen cells of irradiated B10.D2(R101) mice and irradiated B10.D2(R101) mice 10 days after the adoptive transfer of syngeneic splenocytes from non-immunized mice were plated in 96-well U-bottom plates (Corning Costar, Sigma Aldrich) with 10^6–10^7 CFU of heat-inactivated *L. monocytogenes* (strain EGD) or 10^5 CFU of heat-inactivated *S. typhimurium* (strain IE 147), prepared as described above. The cells were cultured in 200 μL of a RPMI-1640 medium (PanEco, Russia) supplemented as described above at 37°C, with 5% CO_2, for 72 h. Cells cultured without pathogens were used to assess background proliferation. Cell proliferation was determined as described above. The index of pathogen-induced proliferation was calculated as the ratio between the levels of cell proliferation in response to bacteria and background proliferation.

Evaluation of EL-4 tumor growth and rejection *in vivo*

Sublethally irradiated B10.D2(R101) mice (with or without adoptive transfer of syngeneic splenocytes) were subcutaneously injected with 0.25 mL of a EL-4 lymphoma cell suspension (8.0 × 10^7 cells/mL). Tumor nodes were measured on days 7, 14, and 21 post-transplantation. EL-4 lymphoma was considered totally rejected when no subcutaneous tumor nodes were detected at palpation.

Antibodies

In this work, the following antibodies were used: anti-CD8α – Percp-Cy5.5 (clone 53–6.7, BD Bioscience, USA), anti-CD62L – APC-Cy7 (clone MEL-14, eBioscience, USA), anti-CD44 – APC (clone IM7, eBioscience), anti-CD3 – PE-Cy7 (clone 145-2C11, eBioscience), anti-CD122 – PE (clone TM-β 1, BD Bioscience), anti-CD5- BV421 (clone 53-7.3, BD Biosciences), anti-CXCR3 – BV421 (clone CXCR3-173, BD Biosciences), and anti-CD49d – PE (clone R1-2, BD Biosciences).

Flow cytometry

Cell samples (3.0 × 10^6) were pre-incubated with Fc block (clone 2.4G2, BD Pharmingen, USA) (10 min, 4°C) and then stained with fluorescent antibodies (40 min,

4°C). The analysis was performed on a BD FACSCanto II flow cytometer (BD Bioscience) using the FACSDiva 6.0 software (BD Bioscience). Dead cells were excluded from the analysis based on the parameters of forward and side scatter and staining with propidium iodide (BD Bioscience) or 7-AAD (BioLegend, USA). At least 10^6 events/samples were collected to characterize the peripheral T-lymphocyte populations. Data were processed using the Flow Jo 7.6 software (TreeStar Inc., USA).

Statistical analysis

Data are presented as mean ± SEM. All statistical analyses were performed using the unpaired Student's t-test. *P*-values < 0.05 were considered significant.

RESULTS

Adoptive transfer of syngeneic splenocytes suppresses the immune response in sublethally irradiated mice

In order to assess the effects of the adoptive transfer of syngeneic splenocytes on the functional status of the immune system in sublethally irradiated mice, we used non-immunized or immunized mice as donors of splenocytes (*Fig. 2A,B*). Irradiation of immunized mice resulted in insignificant (1.6-fold) suppression of the specific immune response compared to the control group of immunized non-irradiated animals, whereas the level of the non-specific immune response remained unchanged (*Fig. 2A*). Dramatic suppression of both specific and non-specific *ex vivo* allogeneic immune responses was observed in irradiated mice with adoptively transferred spleen cells of non-immunized or immunized mice (*Fig. 2B*). Accordingly, irradiated mice after the adoptive transfer exhibited prolonged dynamics of EL-4 lymphoma rejection *in vivo* compared to all control groups (*Fig. 3*).

Moreover, our data showed a significant inhibition of the immune response to both *L. monocytogenes* and *S. typhimurium* in sublethally irradiated mice with the adoptive transfer compared to the control group of irradiated animals (*Fig. 2C*). Notably, the *ex vivo* proliferative response of the splenocytes of irradiated mice without the adoptive transfer remained unchanged compared to the non-irradiated animals (*Fig. 2A,C*).

Phenotype characteristics of donor and recipient $CD3^+CD8^+$ T cells in sublethally irradiated mice after an adoptive transfer

We assumed that the immune response in sublethally irradiated mice after the adoptive transfer of syngeneic splenocytes could be inhibited due to the decrease in the absolute cell count and the relative number of $CD3^+$ T cells in the spleen of these mice. To prove this hypothesis true, we performed an adoptive transfer of the spleen cells of B6.GFP mice to sublethally irradiated C57BL/6 mice and individually analyzed populations of the recipient (GFP^-) and donor (GFP^+) T cells. Some 5% of GFP^+ donor cells were detected in the spleen of the irradiated recipients (*Fig. 1*).

The absolute cell counts in the spleen of the irradiated mice were 4.9-fold reduced compared to that in the non-irradiated animals (*Fig. 4A*). The adoptive transfer of syngeneic splenocytes resulted in a 1.5-fold increase in spleen cell counts compared to that in the control irradiated mice ($p \leq 0.01$; *Fig. 4A*).

Sublethal irradiation reduced the relative count of $CD3^+$ cells in the spleen of the mice compared to that in the non-irradiated controls (*Fig. 4B*). On day 10 after the adoptive transfer, the relative count of GFP^- $CD3^+$ cells in the spleen of the irradiated mice was approximately equal to the $CD3^+$ cell count in the spleen of the non-irradiated mice (*Fig. 4B*). The relative count of $CD3^+$ donor cells (GFP^+) was 2.0-fold higher compared to the relative count of GFP^- recipient T cells in the spleen of the irradiated mice after the adoptive transfer (*Fig. 4B*).

The population of $CD8^+$ T cells remained unchanged in the spleen of the irradiated mice and the subset of the recipient (GFP^-) T cells of irradiated mice after the adoptive transfer compared to non-irradiated mice (*Fig. 4C*). However, $CD8^+$ cells comprised 70% of the donor (GFP^+) T lymphocytes in the spleen of the irradiated mice after the adoptive transfer, equal to 1.8 times the relative count of recipient $CD3^+CD8^+$ cells (*Fig. 4C*). We assumed that donor $CD8^+$ T cells preferentially survive after the adoptive transfer and undergo homeostatic proliferation in the irradiated host. These data correlate with recent studies indicating that $CD8^+$ cells require fewer stimuli for homeostatic proliferation compared to $CD4^+$ T lymphocytes [19].

Sublethal irradiation resulted in a decrease in the relative count of naive cells and a 1.8- and 2.3-fold increase in the relative count of central memory cells and effector memory cells, respectively, within the recipient (GFP^-) $CD8^+$ T cells as compared to the non-irradiated mice (*Fig. 4D*). A total of 60% of the donor (GFP^+) $CD8^+$ T cells in the spleen of the irradiated mice after the adoptive transfer had the phenotype of memory cells (*Fig. 4D*).

Several studies have revealed $CD8^+CD122^+$ T cells with suppressive functions [15]. We evaluated the expression of CD122 on the recipient (GFP^-) and the donor (GFP^+) $CD8^+$ T cells in the spleen of the irradiated mice after the adoptive transfer (*Fig. 4E,F*). Over 97% of the donor (GFP^+) $CD8^+$ T cells acquired the phenotype $CD8^+CD122^+$ (*Fig. 4E*), whereas the relative count of $CD8^+CD122^+$ T cells within the population of the

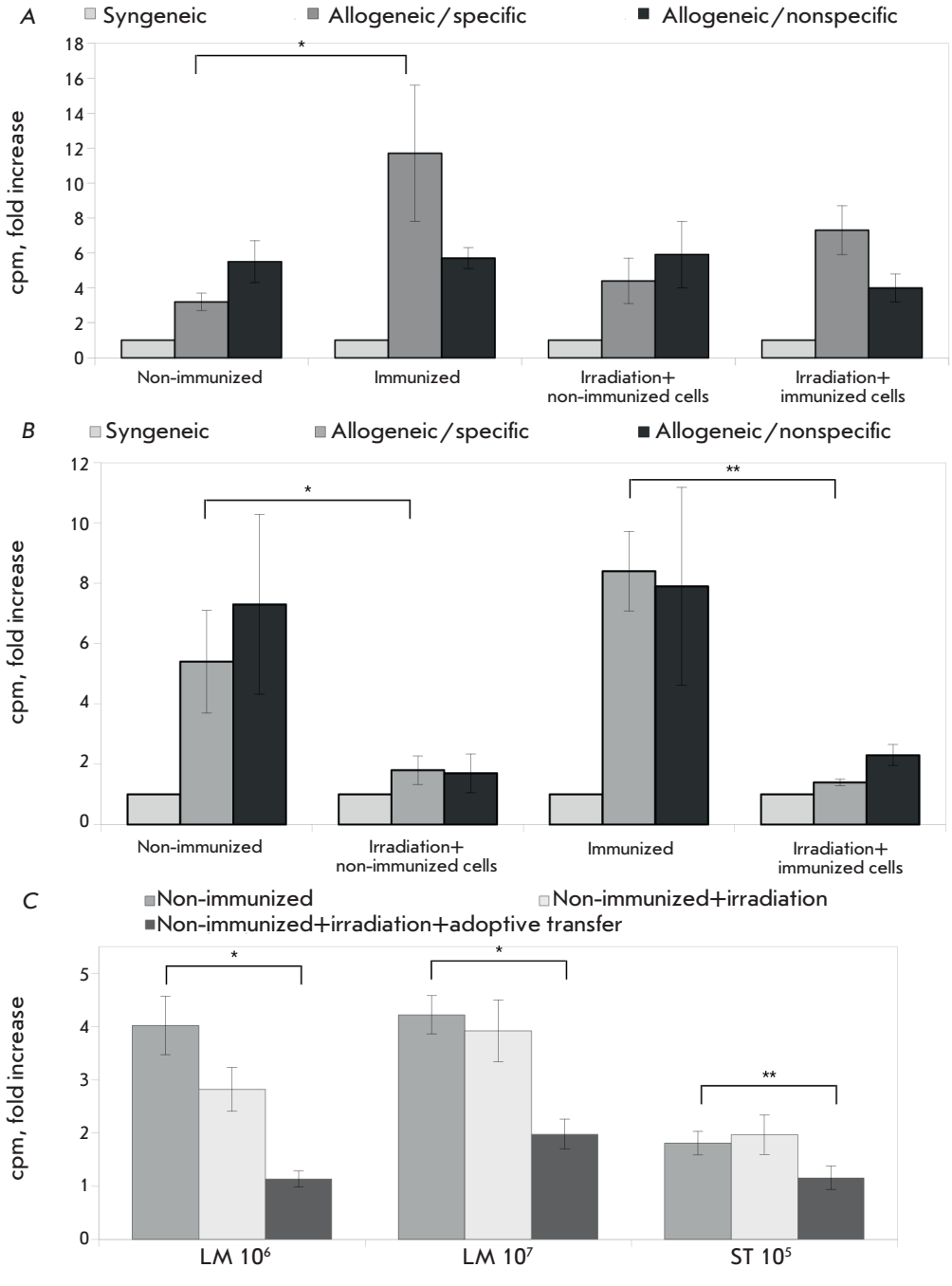


Fig. 2. Analyses of the *ex vivo* functional activity of splenocytes in the lymphopenic mice. (*A*) – The relative level of proliferation of the mixed lymphocyte culture of splenocytes of the sublethally irradiated mice in the allogeneic response. The spleen cells of sublethally irradiated mice were used as responders. Mitomycin C-treated splenocytes of syngeneic (B10.D2(R101), allogeneic/specific (C57BL/10), and allogeneic/nonspecific (FVB) mice were used as stimulators. The relative proliferation level was evaluated as a ratio between the allogeneic and syngeneic responses. The data were obtained in three independent experiments, 3 mice per group. (*B*) – The relative level of proliferation of the mixed lymphocyte culture of the splenocytes of sublethally irradiated mice after the adoptive transfer in the allogeneic response. The spleen cells of sublethally irradiated mice on day 10 after the adoptive transfer were used as responders. Mitomycin C-treated splenocytes of syngeneic (B10.D2(R101), allogeneic/specific (C57BL/10), and allogeneic/nonspecific (FVB) mice were used as stimulators. The relative proliferation level was evaluated as a ratio between the allogeneic and syn-

geneic response. The data were obtained in three independent experiments, 3 mice per group ($^{*}p \le 0.05$; $^{**}p \le 0.01$). (*C*) – The relative level of *in vitro* proliferation of the splenocytes of sublethally irradiated mice after the adoptive transfer in response to pathogens. The splenocytes of sublethally irradiated mice after the adoptive transfer were cultured with 10^6 CFU/mL (LM 10^6) or 10^7 CFU/mL (LM 10^7) of heat-inactivated *Listeria monocytogenes* or with 10^5 CFU/mL (ST 10^5) of heat-inactivated *Salmonella typhimurium*. To assess the background proliferation, the splenocytes were similarly cultured without the pathogen. The relative proliferation levels were assessed as the ratio between the pathogen-induced proliferation and the background proliferation. The data were obtained in three independent experiments, 3–6 mice per group ($^{*}p \le 0.05$, $^{**}p \le 0.01$).

recipient lymphocytes remained unchanged compared to the irradiated and non-irradiated mice (*Fig. 4E*). The level of CD122 expression in the subsets of memory cells ($CD44^+CD62L^+$) and effectors ($CD44^+CD62L^-$) within the donor (GFP^+) T cells was significantly increased compared to the respective subpopulations of the recipient (GFP^-) lymphocytes (*Fig. 4F*).

To evaluate potentially autoreactive T cells within the donor T lymphocytes, we analyzed the expression of CD5 in the population of $CD8^+CD44^+$ T cells (*Fig. 5A,B,C*). Virtually all $GFP^+CD8^+CD44^+$ T cells expressed CD5 (*Fig. 5A*), while the $CD5^+/CD5^-$ ratio in the population of the recipient (GFP^-) $CD8^+CD44^+$ cells remained unchanged compared to the control irradiated and non-irradiated mice (*Fig. 5A,B*). The expression level of CD5 in the $CD44^+CD62L^+$ cells was comparable in all experimental groups (*Fig. 5C*).

Some studies have shown suppressive functions for $CD8^+CD122^+CD49d^{low}$ T cells [18]. We evaluated the expression of the CD49d marker in the population of $CD8^+CD44^+$ T cells of the recipient (GFP^-) and donor (GFP^+) lymphocytes in the spleen of the irradiated mice after the adoptive transfer (*Fig. 5D,E,F*). Nearly 100% of the donor $CD8^+CD44^+$ T cells acquired the $CD49d^{hi}$ phenotype (*Fig. 5 D,E*), whereas the $CD49d^{low}$/$CD49d^{hi}$ ratio in the $CD8^+CD44^+$ T-cell population of the recipient (GFP^-) cells was similar to that in the irradiated and non-irradiated mice (*Fig. 5D,E*). We observed a significant increase in the CD49d expression level within the $CD44^+CD62L^+$ subset of donor GFP^+ $CD8^+$ T cells (*Fig. 5F*).

Furthermore, over 85% of the donor $CD8^+CD44^+$ T cells expressed a $CXCR3^+$ phenotype (*Fig. 5G,H*). The expression level of CXCR3 in the subpopulation of $CD44^+CD62L^+$ was comparable in all experimental groups; in the subpopulation of donor $CD44^+CD62L^-$ T cells, it was in correlation with the level of non-irradiated animals (*Fig. 5I*).

Therefore, the adoptive transfer of syngeneic splenocytes to the lymphopenic host resulted in preferential homeostatic proliferation of $CD8^+$ donor T cells that predominantly acquire the phenotype of the central memory cells $CD44^+CD62L^+$, and most donor $CD44^+$ T cells carry the $CD122^+CD5^+CD49d^{hi}CXCR3^+$ phenotype.

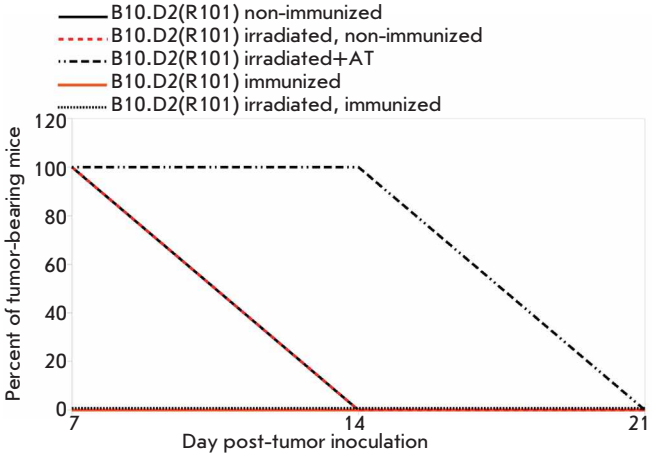


Fig. 3. The dynamics of lymphoma EL4 rejection in sublethally irradiated B10.D2(R101) mice after the adoptive transfer. The data of one representative experiment are presented, 3 mice per group

DISCUSSION

Recent data indicate that there is no strict correlation between the surface phenotype and functional characteristics of a memory T cell (long-term self-maintenance, resistance to apoptosis, simplified activation conditions, enhanced proliferation and acquisition of effector functions in response to the specific antigen). The population of $CD8^+CD44^+CD62L^+CD122^+$ cells was shown to exhibit immunosuppressive activity [13, 16–18, 20]. Commonly, this population expresses high levels of the chemokine receptor CXCR3 [17] and low levels of CD49d ($CD8^+CD122^+CD49d^{low}$) [18]. Similar populations of such suppressive $CD8^+$ T cells were detected both in mice and in humans [21].

We have shown that the adoptive transfer of syngeneic lymphocytes to irradiated mice results in the suppression of the allogeneic immune response and the immune responses to pathogens in such mice. This could be explained by the preferential home-

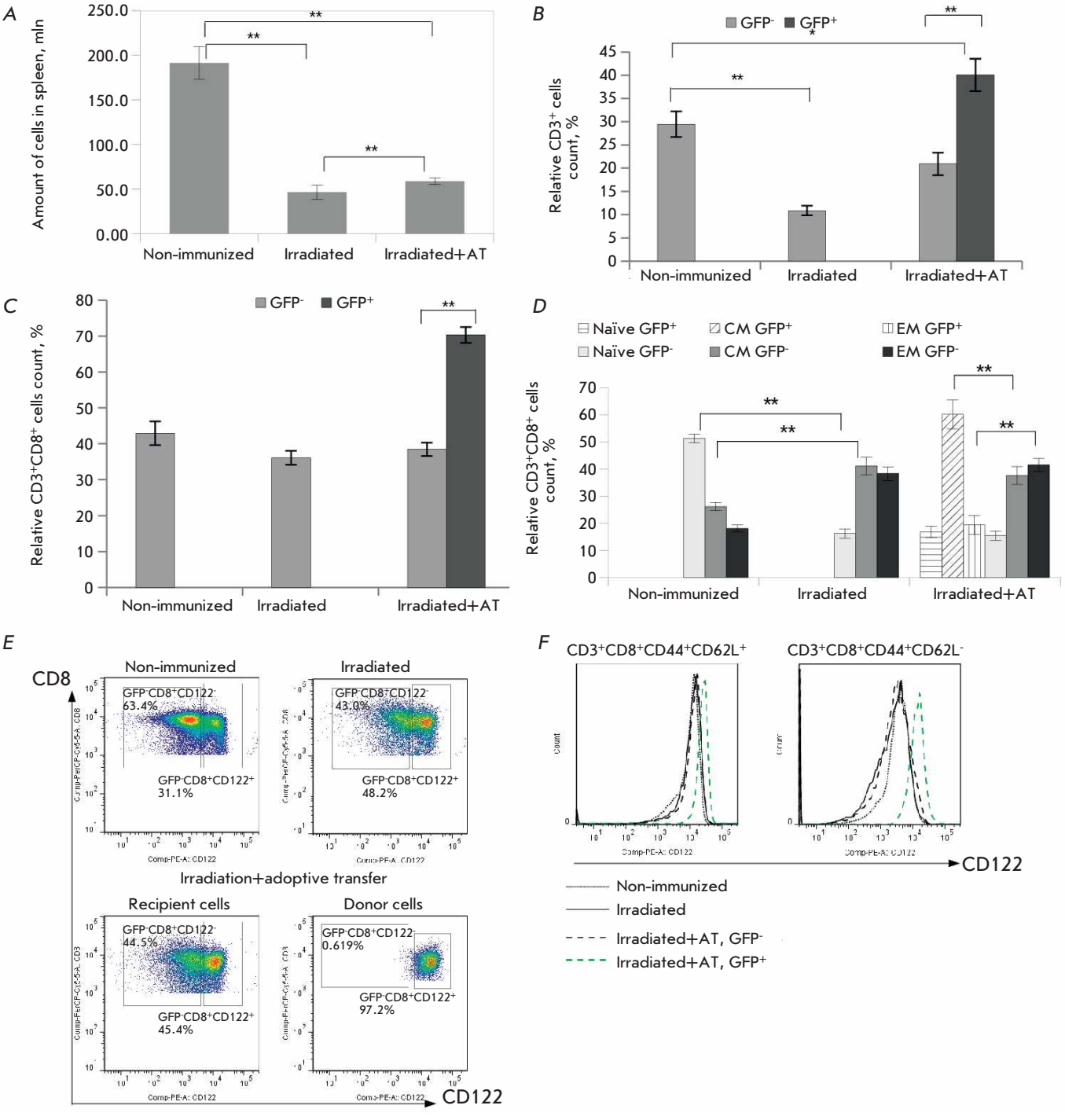


Fig. 4. The absolute cell count and the expression profile of activation markers in the population of CD8+ T cells of the donor (GFP+) and the recipient (GFP-) sublethally irradiated mice on day 10 after the adoptive transfer. (*A*) — The absolute cell count in the spleen of irradiated mice. The data were obtained in three independent experiments, 6–9 mice per group ($^{**}p \leq 0.01$). (*B*) — The relative count of CD3+ cells in the spleen of the irradiated mice. The data were obtained in three independent experiments, 4–6 mice per group ($^{*}p \leq 0.05$, $^{**}p \leq 0.01$). (*C*) — The relative count of CD3+CD8+ cells in the spleen of the irradiated mice. The data were obtained in three independent experiments, 4–6 mice per group

($^{**}p \leq 0.01$). (*D*) – The relative count of $CD8^+$ T cells with the phenotype of naive, effectors, and central memory cells. The data were obtained in three independent experiments, 6 mice per group ($^{**}p \leq 0.01$). (*E*) – The relative count of $CD8^+CD122^+$ splenocytes in the population of the donor (GFP^+) and the recipient (GFP^-) cells of mice on day 10 after the irradiation and adoptive transfer. The data were obtained in three independent experiments, 6 mice per group. The data of one representative experiment are presented. (*F*) – The expression profile of CD122 in the population of $CD8^+CD44^+CD62L^+$ and $CD8^+CD44^+CD62L^-$ T cells in the spleen of the mice on day 10 after the irradiation and adoptive transfer. The expression profiles for the donor (GFP^+) and the recipient (GFP^-) cells. The data were obtained in three independent experiments, 6 mice per group. The data of one representative experiment are presented

ostatic proliferation of T-cell clones that differ from the clonotypes involved in these immune responses. Accordingly, a decreased alloantigen-induced *ex vivo* response was observed for the T cells of irradiated mice regardless of an adoptive transfer of the spleen cells of non-immunized or immunized mice (*Fig. 2B*).

Lymphopenia could drive the homeostatic proliferation of potentially autoreactive clones. Of particular note, virtually all donor $CD8^+CD44^+$ T cells in our study expressed CD5. Several studies have indicated that the level of CD5 expression could correlate with the avidity of a T-cell receptor (TCR) to self-MHC-peptide complexes [22–24]. Interaction with self-MHC is required for T cells to proliferate under lymphopenic conditions [25, 26], and T lymphocytes with the highest level of CD5 expression (i.e., naive T cells) have the greatest homeostatic proliferation potential [3]. Accordingly, naive T cells could be the main source of virtual memory cells in the lymphopenic host [15, 26]. Consistent with these findings, we observed a 1.5-fold increase in the relative cell count of donor $CD8^+CD44^+CD62L^+CD5^+$ T lymphocytes compared to all controls (*Fig. 5A,B*).

Naive T cells are very radiosensitive [28], and total body irradiation can diminish the population of these cells (*Fig. 4D*). Therefore, we assume that under lymphopenia, without competition for self-MHC-peptide complexes, adoptively transferred donor naive T cells can rapidly acknowledge tonic signals for proliferation [29, 30] and acquire the phenotype of central memory cells (*Fig. 4D*). Thus, it seems possible that, in the lymphopenic host, the memory phenotype of T cells was a consequence of the interaction between TCR and MHC/peptide complexes and homeostatic proliferation, rather than indicative of the actual antigenic experience of this T cell. We have recently shown that in mice transgenic for the β-chain TCR, T cells expressing transgenic TCRβ predominantly show the phenotype of naive cells because of the significant competition for self-MHC-peptide complexes; T cells with endogenous TCRβ express the phenotype of effectors and memory cells as a consequence of the excessive amount of ligands available for recognition [31].

Intriguingly, in the lymphopenic host, donor $CD8^+$ T cells acquire a phenotype strikingly different from that of recipient $CD8^+$ T cells. $CD8^+$ cells comprise 70% of donor $CD3^+$ lymphocytes and predominantly carry the phenotype of the central memory cells $CD44^+CD62L^+$. Furthermore, virtually all donor $CD8^+$ lymphocytes have the $CD49d^{hi}$ phenotype and express CD122; the expression level of these markers in the subset of donor $CD44^+CD62L^+$ cells is significantly higher compared to that for the respective subpopulation of recipient $CD8^+$ T cells.

Thus, we have demonstrated that the population of donor $CD8^+$ T cells formed under homeostatic proliferation in the irradiated host acquires the $CD44^+CD62L^+CD122^+CD49d^{hi}$ phenotype, combining some phenotypic characteristics of true memory cells ($CD44^+CD62L^+CD49d^{hi}$) and those of suppressive $CD8^+$ T cells ($CD44^+CD62L^+CD122^+$) [18]. Furthermore, these donor T cells express CXCR3, another marker of suppressive $CD8^+CD122^+$ cells [17]. Considering these findings, we speculate that the adoptive transfer of syngeneic lymphocytes to an irradiated host can lead to the formation of a unique $CD8^+$ T-cell subset of donor cells exhibiting suppressive activity.

CONCLUSIONS

Consistent with previous studies, our experimental data further prove that expression of CD44 on T cells does not always indicate the actual antigenic experience of a T cell and does not necessarily lead to the acquisition of the functional properties of true memory T cells. This means that identification of $CD8^+$ memory T cells based solely on their surface phenotype is incorrect and requires confirmation through functional tests. In this study, $CD8^+$ T lymphocytes adoptively transferred to the irradiated lymphopenic host were converted to T_{ML} cells that shared the phenotypic features of true memory cells and suppressive $CD8^+$ T lymphocytes. This was accompanied through a significant deterioration of the functional state of

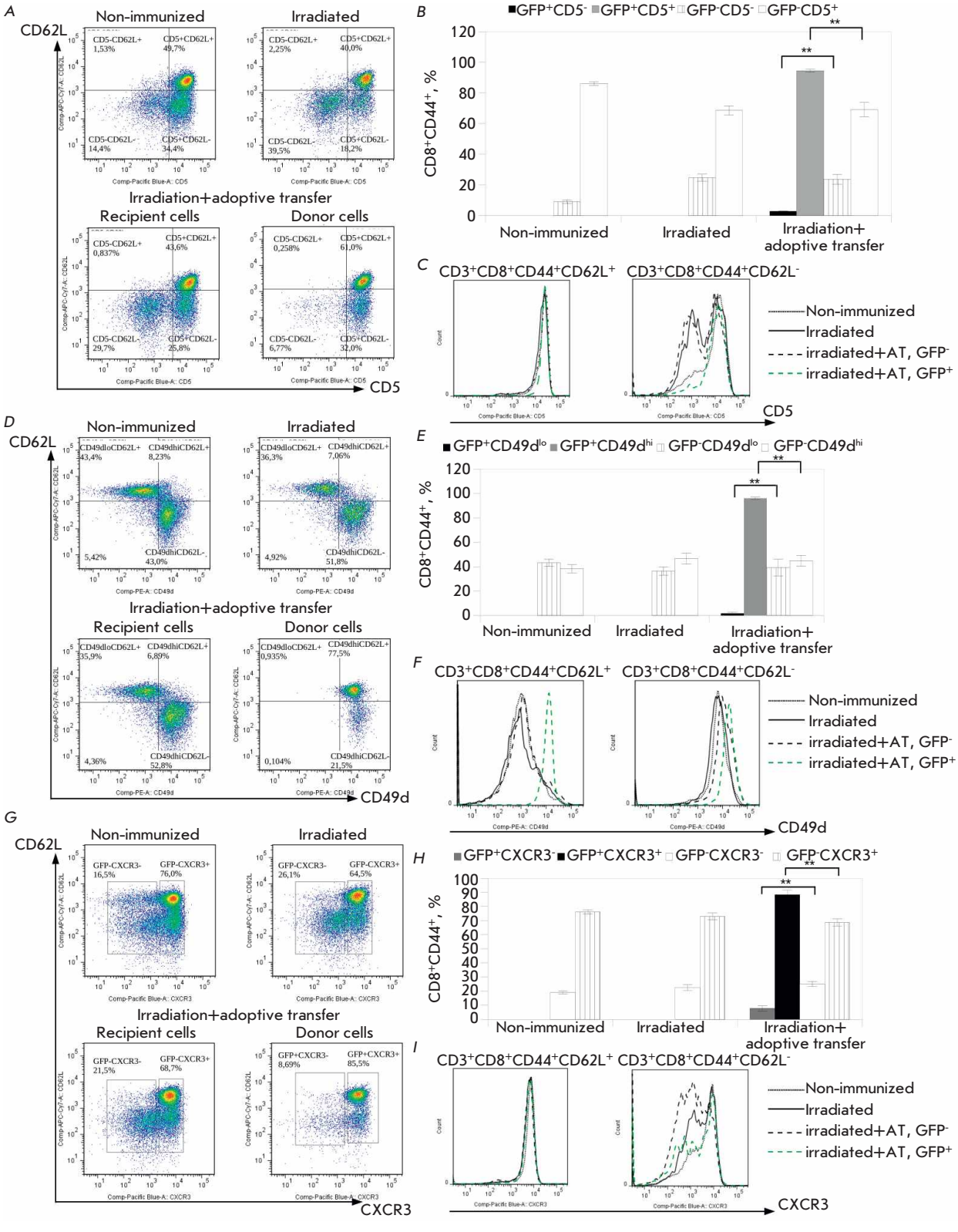


Fig. 5. The relative cell count and the expression profile of CD49d, CD5, and CXCR3 in the population $CD8^+CD44^+$ T cells in the spleen of mice on day 10 after the irradiation and adoptive transfer. Expression of CD5 (*A*), CD49d (*D*), and CXCR3 (*G*) on $CD8^+CD44^+$ T cells of the donor (GFP^+) and the recipient (GFP^-) in the spleen of irradiated mice. The data were obtained in three independent experiments, 4–6 mice per group. The data of one representative experi-

ment are presented. The relative count of $CD8^+CD44^+$ T cells with the phenotypes $CD5^{low}$ and $CD5^{hi}$ (*B*); $CD49d^-$ and $CD49d^+$ (*E*); $CXCR3^-$ and $CXCR3^+$ (*H*) in the population of the donor (GFP^+) and the recipient (GFP-) splenocytes in the irradiated mice. The data were obtained in three independent experiments, 4–6 mice per group. The data of one representative experiment are presented ($^{**}p \leq 0.01$). The expression profiles of CD5 (*C*), CD49d (*F*), and CXCR3 (*I*) on $CD8^+CD44^+CD62L^+$ and $CD8^+CD44^+CD62L^-$ T cells in the spleen of the irradiated mice. The expression profiles for the donor (GFP^+) and the recipient (GFP^-) cells are presented. The data were obtained in three independent experiments, 4–6 mice per group. The data of one representative experiment are presented

the recipient's immune system, whose T cells poorly responded to the alloantigens and bacterial antigens. Memory-like $CD8^+$ T cells [32] and suppressive $CD8^+CD44^+CXCR3^+$ T cells [17] are likely to exist in the human organism. Thus, the adoptive transfer aimed at restoring the count of immune cells in peripheral organs can lead to clinically unfavorable outcomes: i.e., a weaker response to antigens and, hence, increased predisposition or vulnerability to infectious diseases. ●

This study was supported by the Megagrant (Agreement No. 14.W03.31.0020 between the Ministry of Science and Education of the Russian Federation and Institute of Gene Biology, Russian Academy of Sciences).

This work was performed using the equipment of IGB RAS facilities supported by the Ministry of Science and Higher Education of the Russian Federation.

We thank L.M. Nesterenko and S.A. Ermolaeva ("N. F. Gamaleya National Research Center of Epidemiology and Microbiology", Ministry of Health of the Russian Federation, Moscow, Russia) for providing virulent strains of salmonella and listeria to us.

REFERENCES

1. Zhang Q., Lakkis F.G. // Front. Immunol. 2015. V. 5. P. 504–509.
2. Haluszczak C., Akue A.D., Hamilton S.E., Johnson L.D.S., Pujanauski L., Teodorovic L., Jameson S.C., Kedl R.M. // J. Exp. Med. 2009. V. 206. № 2. P. 435–448.
3. White J.T., Cross E.W., Kedl R.M. // Nat. Rev. Immunol. 2017. V. 17. № 6. P. 391–400.
4. Cho B.K., Rao V.P., Ge Q., Eisen H.N., Chen J. // J. Exp. Med. 2000. V. 192. P. 549–556.
5. Goldrath A.W., Bogatzki L.Y., Bevan M.J. // J. Exp. Med. 2000. V. 192. P. 557–564.
6. Murali-Krishna K., Ahmed R. // J. Immunol. Baltim. Md. 1950. 2000. V. 165. P. 1733–1737.
7. Jameson S.C., Lee Y.J., Hogquist K.A. // Adv. Immunol. 2015. P. 173–213.
8. Ge Q., Hu H., Eisen H.N., Chen J. // Microbes Infect. 2002. V. 4. P. 555–558.
9. Tanchot C., Le Campion A., Martin B., Léaument S., Dautigny N., Lucas B. // J. Immunol. Baltim. Md. 1950. 2002. V. 168. P. 5042–5046.
10. Moxham V.F., Karegli J., Phillips R.E., Brown K.L., Tapmeier T.T., Hangartner R., Sacks S.H., Wong W. // J. Immunol. Baltim. Md. 1950. 2008. V. 180. P. 3910–3918.
11. Oghumu S., Terrazas C.A., Varikuti S., Kimble J., Vadia S., Yu L., Seveau S., Satoskar A.R. // FASEB J. 2015. V. 29. P. 1019–1028.
12. Cheung K.P., Yang E., Goldrath A.W. // J. Immunol. 2009. V. 183. P. 3364–3372.
13. Wang L.-X., Li Y., Yang G., Pang P., Haley D., Walker E.B., Urba W.J., Hu H.-M. // Eur. J. Immunol. 2010. V. 40. P. 1375–1385.
14. Le Campion A., Gagnerault M.-C., Auffray C., Becourt C., Poitrasson-Riviere M., Lallemand E., Bienvenu B., Martin B., Lepault F., Lucas B. // Blood. 2009. V. 114. P. 1784–1793.
15. White J.T., Cross E.W., Burchill M.A., Danhorn T., McCarter M.D., Rosen H.R., O'Connor B., Kedl R.M. // Nat. Commun. 2016. V. 7. P. 11291.
16. Suzuki H., Shi Z., Okuno Y., Isobe K. // Hum. Immunol. 2008. V. 69. P. 751–754.
17. Liu J., Chen D., Nie G.D., Dai Z. // Front. Immunol. 2015. V. 6. P. 494.
18. Akanea K., Kojimab S., Mak T.W., Shikud H., Suzuki H. // Proc. Natl. Acad. Sci. USA. 2016. V. 113. № 9. P. 2460–2465.
19. Hickman SP, Turka LA. // Philos. Trans. R Soc. Lond. B Biol. Sci. 2005. V. 360. № 1461. P. 1713–1721.
20. Wan N., Dai H., Wang T., Moore Y., Zheng X.X., Dai Z. // J. Immunol. 2008. V. 180. № 1. P. 113–121.
21. Shi Z., Okuno Y., Rifa'i M., Endharti A.T., Akane K., Isobe K., Suzuki H. // Eur. J. Immunol. 2009. V. 39. № 8. P. 2106–2119.
22. Mandl J.N., Monteiro J.P., Vrisekoop N., Germain R.N. // Immunity. 2013. V. 38. P. 263–274.
23. Persaud S.P., Parker C.R., Lo W.L., Weber K.S., Allen P.M. // Nat. Immunol. 2014. V. 15. P. 266–274.
24. Fulton R.B., Hamilton S.E., Xing Y., Best J.A., Goldrath A.W., Hogquist K.A., Jameson S.C. // Nat. Immunol. 2015. V. 16. № 1. P. 107–117.
25. Goldrath A.W., Bevan M.J. // Immunity. 1999. V. 11. P. 183–190.
26. Kieper W.C., Jameson S.C. // Proc. Natl. Acad. Sci. USA. 1999. V. 96. P. 13306–13311.
27. Sosinowski T., White J.T., Cross E.W., Haluszczak C., Marrack P., Gapin L., Kedl R.M. // J. Immunol. 2013. V. 190. P. 1936–1947.
28. Yao Z., Jones J., Kohrt H., Strober S. // J. Immunol. 2011.

V. 187. P. 4100–4108.
29. Cho J.H., Kim H.O., Surh C.D., Sprent J. // Immunity. 2010. V. 32. P. 214–226.
30. Takada K., Jameson S.C. // J. Exp. Med. 2009. V. 206. P. 2253–2269.
31. Silaeva Yu.Yu., Kalinina A.A., Vagida M.S, Khromykh L.M., Deikin A.V., Ermolkevich T.G., Sadchikova E.R., Goldman I.L., Kazansky D.B.// Biochemistry (Moscow). 2013. V. 78. № 5. P. 549–559.
32. van Kaer L. // Eur. J. Immunol. 2015. V. 45. P. 1916–1920.

Peculiarities of the Presentation of the Encephalitogenic MBP Peptide by HLA-DR Complexes Providing Protection and Predisposition to Multiple Sclerosis

A. E. Mamedov[1*], I. N. Filimonova[1], I. V. Smirnov[1,2], A. A. Belogurov[1,3]

[1]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry Russian Academy of Sciences, Moscow, 117997 Russia

[2]Institute of Fundamental Medicine and Biology, Kazan (Volga) Federal University, Kazan, 420008 Russia

[3]Lomonosov Moscow State University, Moscow, 119991 Russia

*E-mail: bioaz12@gmail.com

Received May 18, 2020; in final form, August 5, 2020

DOI: 10.32607/actanaturae.11008

ABSTRACT **Predisposition to multiple sclerosis (MS), a chronic autoimmune disease of the central nervous system, is due to various factors. The genetic component is considered one of the most important factors. HLA class II genes contribute the most to the development of MS. The HLA-*DRB1**15 allele group is considered one of the main genetic risk factors predisposing to MS. The group of HLA-*DRB1**01 alleles was shown to have a protective effect against this disease in the Russian population. In this work, we compared the binding of the encephalitogenic fragment of the myelin basic protein (MBP) to two HLA-DR complexes that provide protection against and predisposition to MS: HLA-DR1 (HLA-*DRB1**0101) and HLA-DR15 (HLA-*DRB1**1501), respectively. We found that the myelin peptide MBP_{88-100} binds to HLA-DR1 at a rate almost an order of magnitude lower than the viral peptide of hemagglutinin (HA). The same was true for the binding of MBP_{85-97} to HLA-DR15 in comparison with viral pp65. The structure of the C-terminal part of the peptide plays a key role in the binding to HLA-DR1 for equally high-affinity N-terminal regions of the peptides. The IC_{50} of the myelin peptide MBP_{88-100} competing with viral HA for binding to HLA-DR1 is almost an order of magnitude higher than that of HA. As for HA, the same was also true for the binding of MBP_{85-97} to HLA-DR15 in comparison with viral pp65. Thus, autoantigenic MBP cannot compete with the viral peptide for binding to protective HLA-DR1. However, it is more competitive than viral peptide for HLA-DR15.**

KEYWORDS **multiple sclerosis, HLA-DR, protective allele, risk allele, MBP peptide, viral peptide.**

ABBREVIATIONS **APC – antigen-presenting cell; ELISA – enzyme-linked immunosorbent assay; HLA – human leukocyte antigen; MBP – basic myelin protein; MS – multiple sclerosis; PBS – phosphate-buffered saline; TRX – thioredoxin.**

INTRODUCTION

The human leukocyte antigen (HLA) genes encode proteins that can bind to and present antigenic peptides. Therefore, they play a critical role in the immune response to pathogens and autoimmunity [1]. Binding of antigenic peptides to HLA class II molecules leads to the formation of binary peptide–HLA complexes. These complexes are presented on the surface of antigen-presenting cells (APCs) and recognized by CD4 T cell receptors [2]. Newly synthesized HLA proteins are protected against aggregation by the invariant chain [3]. In the endosomal compartment, the invariant chain is partially degraded, thus leaving the CLIP peptide in the binding groove [4, 5]. CLIP can be further exchanged for antigenic peptides, which form as a result of antigen processing in endosomes. The exchange process is catalyzed by the HLA-DM protein [6]. The peptide–HLA complex is transported next to the APC surface for recognition by CD4 T cells. The mechanisms of peptide presentation by HLA class II molecules are well known [7]. However, it remains unclear how the formation and presentation of autoantigen–HLA complexes lead to autoimmune reactions, and there is substantial interest in the topic. Thus, the identification of the autopeptide–

HLA complexes associated with autoimmune responses may provide a clue to our understanding of the pathogenesis of autoimmune diseases [8–10].

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system which is characterized by inflammation, demyelination, and neurodegeneration [11]. The nature of the genetic predisposition to MS is complex and depends on a combination of multiple genetic and epigenetic factors, not to mention environmental factors [12]. The genes in the HLA region are considered to contribute substantially to the risk of MS [13]. Certain alleles of the highly polymorphic HLA class II gene *DRB1* appear to be a significant genetic determinant in the pathology of MS and can affect both predisposition and resistance to the disease [14]. The HLA-*DRB1**1501 allele and the haplotype associated with it (*DQA1**0102, *DQB1**0602, *DRB1**1501, and *DRB5**0101) have been known as universal risk factors for MS since the 1970s. An analysis of the association of HLA with MS in Northern European populations revealed the groups of HLA-*DRB1* alleles (*DRB1**03, *01, *10, *11, *14, *08) in positive or negative correlation with the risk of the disease [15]. Furthermore, the autoantigenic peptides presented by the risk alleles have been identified. HLA-DRB1*1501 binds a fragment of myelin basic protein (MBP), the encephalitogenic peptide MBP_{85-99} [8], while HLA-DRB5*0101 presents the MBP_{86-105} peptide [10]. The CD4 T cell clones that recognize these peptide–HLA complexes associated with the disease have been identified as well [16–18].

It was previously shown in a representative cohort of ethnic Russian patients with MS and conditionally healthy individuals that the group of HLA-*DRB1**01 alleles is associated with MS resistance, while HLA-*DRB1**15 alleles are positively associated with the disease. An analysis of the interaction of proteins encoded by the HLA-*DRB1**1501 risk allele and the protective allele HLA-*DRB1**0101 with the MBP library demonstrated that both proteins can bind myelin peptide MBP_{81-104} with similar affinity [19]. However, it is unclear how binding of the same myelin fragment provides protection in the case of one allele and predisposition to the disease in the case of the other allele. Therefore, the aim of this study was to analyze the kinetic characteristics of the interaction of peptide MBP_{81-104} with the protective HLA-DR1 and predisposing HLA-DR15 in patients with MS, as well as to compare them to their viral antigenic determinants.

EXPERIMENTAL

Expression and purification of proteins

Recombinant proteins HLA-DR1 (the product of the HLA-*DRA1**0101 and HLA-*DRB1**0101 genes), HLA-DR15 (the product of the HLA-*DRA1**0101 and HLA-*DRB1**1501 genes), and HLA-DM were obtained using the method described previously [20]. The CLIP peptide (PVSKMRMATPLLMQA) was covalently bound to the N-terminus of the β-chains of HLA-DR1 and HLA-DR15 via a linker with a thrombin cleavage site, at which the peptide was cleaved for further experiments (1 h, 20 U/mg, 25°C). Proteins were concentrated in PBS and stored at 4°C.

Peptides fused to thioredoxin were designed and obtained using a previously constructed MBP epitope library [21]. Genetic constructs coding for HA, pp65, myelin peptides (MBP_{88-100} and MBP_{85-97}), MBP with point mutations (V86A, V87A, F89A, and F90A), and chimeric peptides (HA-MBP, MBP-HA, and pp65-MBP) were obtained by PCR using the MBP epitope library as a template. The protein constructs were presented by peptides fused to the C-terminus of bacterial thioredoxin through a flexible linker $(SGGGG)_3S$ carrying His-tags for purification. The construct carrying only thioredoxin with a linker (TRX) was used as a negative control. All thioredoxin-fused peptides were obtained using the method described previously [21]. The peptides were chemically biotinylated with EZ-Link Sulfo-NHS-LC-biotin (Thermo Fisher Scientific) at a molar ratio of 1 : 20 for 30 min at 25°C. Proteins were concentrated in PBS and stored at –20°C.

ELISA for analyzing HLA-DR peptide binding

Biotinylated peptide MBP_{81-104} and its variants with point mutations (V86A, V87A, F89A, and F90A) (750 nM) were incubated in 50 μL of PBS with CLIP-bound HLA-DR (HLA-DR1 or HLA-DR15) (150 nM) at 37°C for 18 h (*Fig. 1A*). Thioredoxin with a linker (TRX) was used as a negative control. DR–peptide complexes were then added to anti-HLA-DR antibodies (L243) immobilized on the plate and blocked by PBS containing a 2% skim milk powder. The biotinylated peptide bound to HLA-DR was quantified using horseradish peroxidase-conjugated streptavidin.

In a competitive assay, biotinylated HA and pp65 peptides (150 nM) were incubated with the corresponding HLA-DR (HLA-DR1 or HLA-DR15) (150 nM) in the presence of either non-biotinylated HA, pp65, myelin (MBP_{88-100} and MBP_{85-97}) or chimeric (HA-MBP, MBP-HA, and pp65-MBP) peptides at concentrations of 1,000; 500; 250; 125; 62.5; 31.2; 15.6; and 7.8 nM in 50 μL of PBS at 37°C for 18 h. Experiments were carried out in triplicate.

ELISA for analyzing the kinetics of peptide loading onto HLA-DR

The corresponding HLA-DR bound to CLIP (HLA-DR1 or HLA-DR15) (150 nM) was incubated in the presence

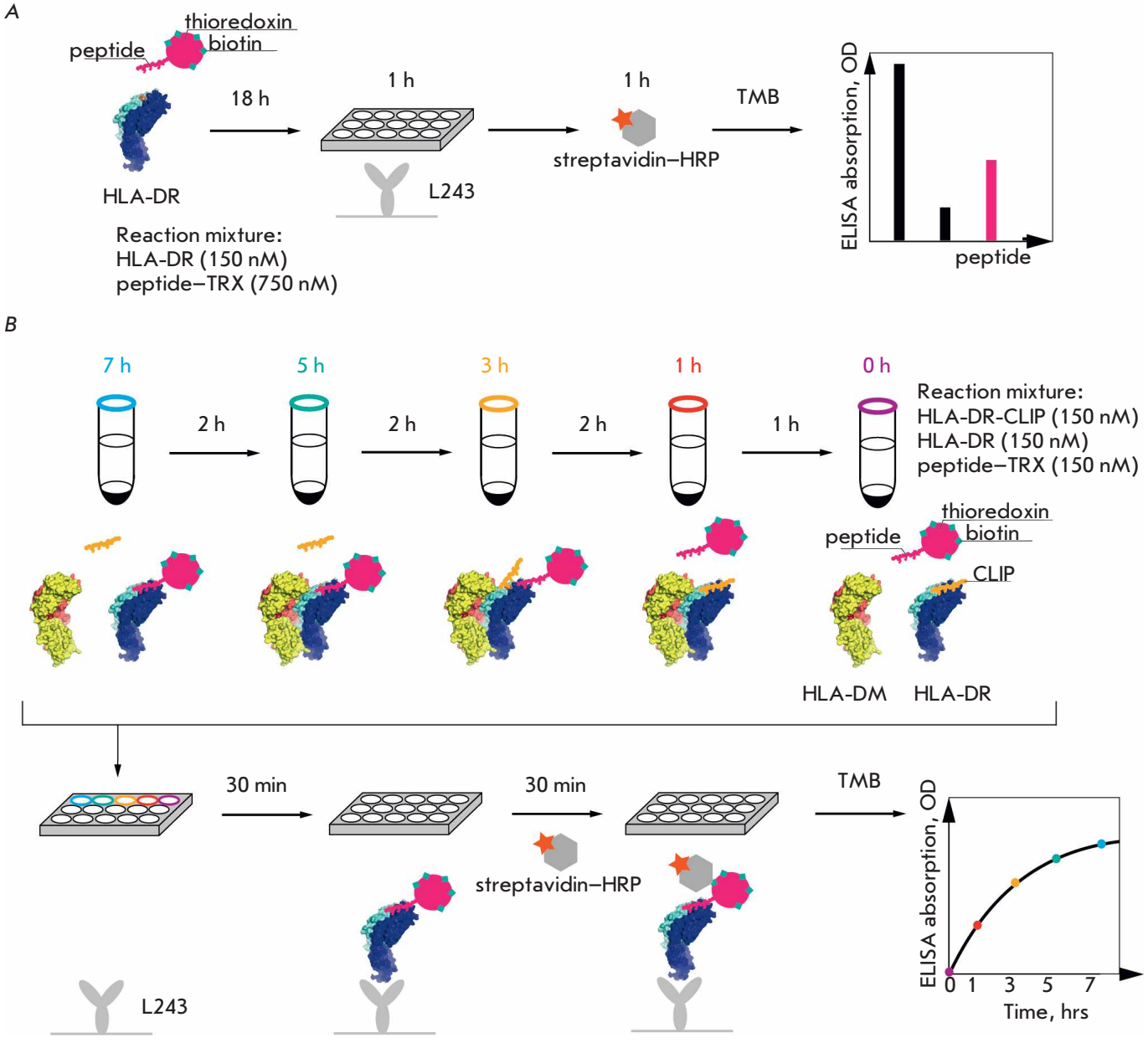


Fig. 1. Schematic representation of ELISA for the binding of HLA-DR peptides (*A*) and kinetics of peptide loading onto HLA-DR (*B*). Each time point is marked with color. L243 – immobilized monoclonal antibodies to HLA-DR

of HLA-DM (150 nM) in 50 μL of citrate buffer (50 mM sodium citrate, 150 mM NaCl, pH 5.3) with either biotinylated HA, pp65, myelin (MBP_{88-100} and MBP_{85-97}) or chimeric (HA-MBP, MBP-HA, and pp65-MBP) peptides (150 nM) at 37°C for 7, 5, 3, 1, and 0 h (*Fig. 1B*). For each time point, the experimental system was mixed separately every 2 h starting from the longest incubation time (7 or 5 h), after which all time points were simultaneously added to the plate. ELISA was performed as described above, with the only difference being that the time of incubation of the reaction mixtures in the plate with streptavidin was reduced to 30 min. Experiments were carried out in triplicate. The kinetic curves were analyzed using the Enzyme Kinetics module of the SigmaPlot software (Sigma-Aldrich). The binding curves were fitted using a nonlinear least-squares fit to the Langmuir binding model describing a 1 : 1 binding stoichiometry.

RESULTS AND DISCUSSION

Determining the MBP_{81-104} epitopes recognized by HLA-DR1 and HLA-DR15

We have compared the kinetic characteristics of the interaction between the encephalitogenic fragment of the myelin basic protein MBP_{81-104} and human major histocompatibility complex class II proteins, namely MS-protective HLA-DR1 and MS-predisposing HLA-DR15 [19], as well as their antigenic determinants of viral origin. In order to conduct our analysis, it was first necessary to determine the binding epitope within MBP_{81-104} recognized by HLA-DR1. Alanine scanning (substitution of hydrophobic and aromatic residues with alanine starting from the N-terminus of the peptide (*Fig. 2A*)) of MBP_{81-104} revealed a Phe90 residue acting as a hydrophobic anchor at P1 (*Fig. 2B*). This led us to suggest that the pockets P6/P7 and P9 in HLA-DR1 that are bound to MBP_{81-104} are occupied by Thr95, Pro96, and Thr98 residues. Identification of the MBP_{81-104} epitope responsible for binding to HLA-DR15, in which Val87 and Phe90 are located at positions P1 and P4, respectively [8], was confirmed using the corresponding mutant forms of MBP_{81-104} (*Fig. 2C*).

Comparison of the kinetics of MBP peptide loading onto HLA-DR1 and HLA-DR15

At the next stage, we studied the kinetics of the binding of HLA-DR1 to the peptides $HA_{306-318}$, MBP_{88-100}, and their chimeric constructs MBP-HA and HA-MBP in the presence of HLA-DM, which accelerates the rate of CLIP exchange for the peptide under study (*Fig. 3B*). HA is a fragment of the influenza virus hemagglutinin, a classic viral antigenic determinant for HLA-DR1 [22]. For comparison with the HLA-*DRB1**1501 risk allele, binding curves for the interaction of HLA-DR15 with peptide $pp65_{109-123}$ (which is a fragment of a cytomegalovirus protein), a HLA-DR15 viral determinant [23], myelin peptide MBP_{85-97}, and chimeric construct pp65-MBP were also obtained (*Fig. 3C*). It is important to note that, in chimeric peptides, the boundary between the N- and C-terminal regions of the constituent peptides lay between the amino acid residues at positions P4 and P5 (*Fig. 3A*).

Viral peptide HA is known to possess a high affinity to the peptide-binding groove of HLA-DR1 [24]. Therefore, the kinetic curve for the interaction between peptide HA and HLA-DR1 reaches a plateau after 8 h (*Fig. 3B*). At the same time, the myelin peptide MBP_{88-100} binds to HLA-DR1 at a rate almost an order of magnitude lower than that of the viral peptide. Thus, we can assume that protective HLA-DR1 kinetically distinguishes between the exogenous viral and endogenous myelin antigens. The chimeric peptide HA-MBP, which contains the N-terminal region of HA (306–311) and the C-terminal part of MBP (94–100), binds to HLA-DR1 at a low rate. This rate is similar to the kinetics of interaction with MBP_{88-100}. However, in the case of the chimeric MBP-HA peptide composed of the N-terminal region of MBP (88–93) and the C-terminal region of HA (312–318), the binding rate is very high. The same is true in the case of binding of natural viral HA. Based on these findings, we can conclude that the kinetic parameters of binding of chimeric peptides to HLA-DR1 indicate the importance of the C-terminal region for an efficient interaction with HLA-DR1 with equally high-affinity N-terminal peptide regions. The N-terminal parts of the fragments under study contain

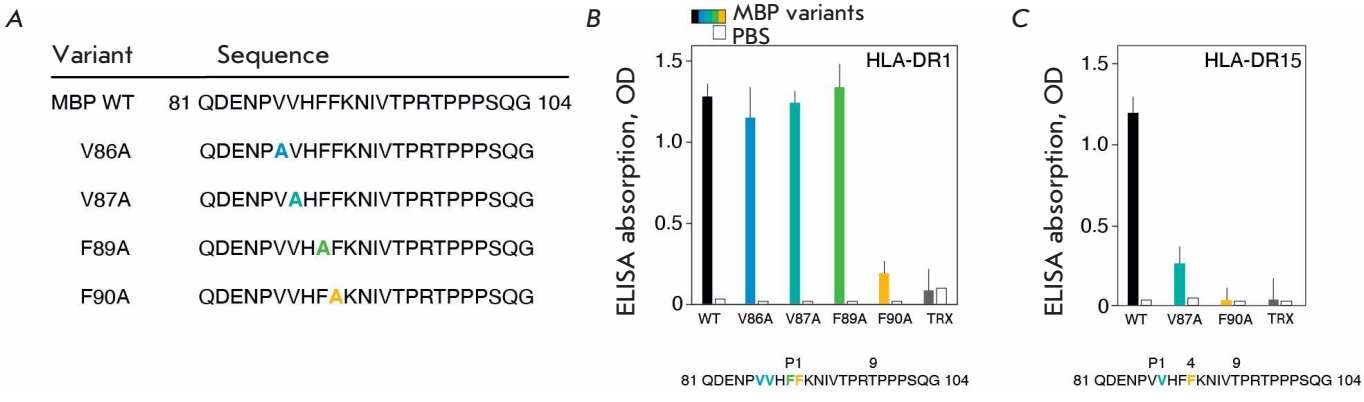


Fig. 2. (*A*) Sequences of peptide MBP_{81-104} and its variants with point amino acid substitutions to alanine. Point substitutions are indicated with different colors. (*B*, *C*) Binding of peptide MBP_{81-104} and its variants (750 nM) carrying point amino acid substitutions to alanine with CLIP-bound HLA-DR1 (*B*) and HLA-DR15 (*C*) (150 nM). Column colors correspond to the colors of point substitutions. White bars represent the background signal (PBS). Thioredoxin (TRX) with a linker was used as a negative control. Standard deviation is presented

the main anchor in the P1 binding pocket: the aromatic residues Tyr308 and Phe90 in the case of HA and MBP, respectively. One can assume that it is the presence of Pro96 in the C-terminal region of MBP_{88-100} and chimeric HA-MBP peptides that changes the peptide position in the binding groove. This happens due to the inherent conformational rigidity of proline and impairs any interaction of the peptide with the P7 binding pocket. In the HA and MBP-HA peptides, the P7 pocket contains a hydrophobic Leu314 within the C-terminal region, which favors binding.

An analysis of the allele responsible for the risk of MS demonstrated that viral, myelin, and chimeric peptides bind to HLA-DR15 at similar rates (*Fig. 3C*). In contrast to HLA-DR1, the key element for peptide binding in HLA-DR15 is the P4 pocket, where aromatic amino acid residues fit ideally. The hydrophobic P1 pocket is second in significance. Therefore, the efficiency of binding of the viral, myelin, and chimeric peptides can be due to the amino acid residues favoring an interaction of peptides with the peptide-binding groove in HLA-DR15. These amino acids are located in the pockets P1 and P4, which are important for peptide loading: Ile111 and Tyr114 in the case of pp65, as well as Val87 and Phe90 in the case of MBP. Despite the fact that the viral peptide pp65 also contains proline in the P7 pocket at the C-terminal region, it does not decrease the efficiency of peptide interaction with the peptide-binding groove. This happens because the pockets P6/P7/P9 play a lesser role than P4 in the case of HLA-DR15. A discrepancy in the rate of interaction of HA with HLA-DR1 and pp65 with HLA-DR15 (about fivefold) can be attributed to differences in the structure of the pockets of these HLA-DR complexes and the presence of anchor residues in the corresponding peptides (*Fig. 3A*).

Differing rates of loading of various peptides of exogenous and endogenous nature onto MS-protective HLA-DR1 and MS-predisposing HLA-DR15 may be an indication that the kinetic component (rather than the thermodynamic one) plays a greater role in the interaction between the MHC II complex and the antigens.

Comparison of the competitiveness of MBP for binding to HLA-DR1 and HLA-DR15

Taking into account the fact that myelin peptide binds to both HLA-DR1 and HLA-DR15, albeit at different rates, the question of whether it can compete for binding with high-affinity viral antigens remained open. To clarify this issue, we conducted some experiments to study the competitive ability of HA, myelin peptide MBP_{88-100}, as well as the chimeric peptides HA-MBP and MBP-HA to bind HLA-DR1 in the presence of viral HAbio (*Fig. 3D*). An analysis was also performed for the interaction of pp65, myelin peptide MBP_{88-100}, and chimeric peptide pp65-MBP with HLA-DR15 in the presence of viral pp65bio (*Fig. 3E*). The kinetic data indicate that HA and chimeric peptide MBP-HA can effectively compete with viral HAbio for HLA-DR1. Moreover, addition of these peptides significantly decreases the ELISA signal starting from a concentration of 30 nM. On the contrary, addition of myelin peptide MBP_{88-100} and chimeric peptide HA-MBP insignificantly reduces the ELISA signal, which is observed only at high concentrations (starting from 300 nM) (*Fig. 3D*). The IC_{50} values of these peptide pairs differ by almost an order of magnitude, which indicates that myelin peptide MBP_{88-100} cannot effectively compete with viral HA for binding to HLA-DR1 (*Fig. 3A*). In the case of HLA-DR15, the decline in the ELISA signal in the competitive reactions with MBP_{85-97} and pp65 starts at a pp65bio concentration of 30 nM (*Fig. 3E*). This is similar to the interaction between the HA peptide and HLA-DR1. At the same time, the IC_{50} of peptide MBP_{85-97} is approximately threefold lower than that of pp65. Thus, unlike MBP_{88-100}, MBP_{85-97} is even more competitive than viral peptide (*Fig. 3A*).

CONCLUSIONS

According to our findings, it is fair to assume that, in contrast to HLA-DR15, it is unlikely that a fragment of the myelin basic protein is presented as its complex with HLA-DR1 on the surface of antigen-presenting cells at the density required for the activation of the T cell response. Apparently, the protective properties of the HLA-*DRB1**0101 allele are associated with the ability of its protein product HLA-DR1 to distinguish kinetically between myelin and exogenous peptides. Meanwhile, HLA-DR15, which is associated with the risk of MS, can efficiently present the MBP fragment even when competing with exogenous peptides such as viral pp65. Our data suggest that the same encephalitogenic myelin fragment can be presented at a completely different rate depending on the HLA-DR allele. In other words, the immunogenicity of myelin components in patients with MS may be largely determined by genetic predisposition due to carriage of a specific HLA-DR allele rather than by their accessibility to immune cells. ●

This study was supported by the Russian Science Foundation (grant No. 17-74-30019) and the Russian Foundation for Basic Research with European Molecular Biology Laboratory (grant No. 18-54-74006).

A

Peptide	Sequence	$V_{obs} \times 10^{-12}$, M/s	IC_{50}, nM
	P1 45 9	+ HLA-DR1	+HAbio
MBP	88 HFFKNIVTPRTPP 100	1.3 ± 0.3	>1000
MBP-HA	HFFKNINTLKLAT	7.1 ± 1.4	200 ± 40
		$p>0.05$	$p>0.05$
HA-MBP	PKYVKQVTPRTPP	1.6 ± 0.3	>1000
	P1 45 9		
HA	PKYVKQNTLKLAT	10.6 ± 2.1	300 ± 50
	P1 45 9	+ HLA-DR15	+pp65bio
MBP	85 PVVHFFKNIVTPR 97	60 ± 9	120 ± 20
pp65-MBP	MSIYVYKNIVTPR	55 ± 7	250 ± 40
		$p>0.05$	$p<0.05$
	P1 45 9		
pp65	MSIYVYALPLKMLNI	50 ± 6	350 ± 60

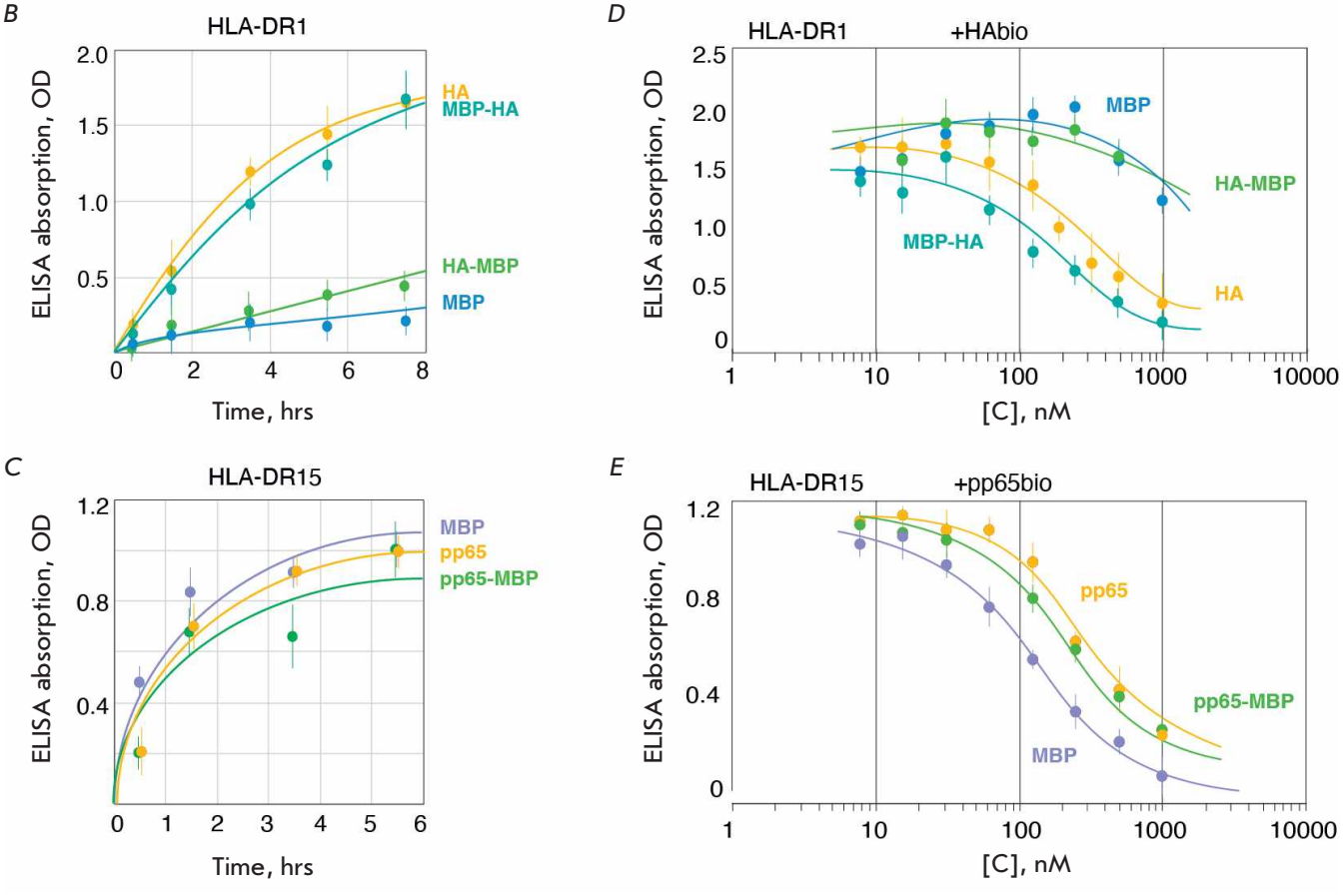


Fig. 3. (*A*) Sequences of the peptides MBP_{88-100}, MBP-HA, HA-MBP, HA, MBP_{85-97}, pp65-MBP, and pp65. Various parts of the chimeric peptides, as well as the positions of the amino acid residues P1/4/5/9, are indicated with different colors. For each of the peptides, the initial rates of interaction with the corresponding CLIP-bound HLA-DR1 or HLA-DR15, as well as the IC_{50} values in competition with the HAbio and pp65bio peptides, are shown. (*B*, *C*) Kinetics of binding of the biotinylated peptides MBP_{88-100}, MBP-HA, HA-MBP, and HA (150 nM) to CLIP-bound HLA-DR1 (150 nM) (*B*), as well as of the biotinylated peptides MBP_{85-97}, pp65-MBP, and pp65 (150 nM) to CLIP-bound HLA-DR15 (150 nM) (*C*) in the presence of HLA-DM (150 nM). (*D*, *E*) Competitive interaction of HLA-DR1 (150 nM) and HLA-DR15 (150 nM) with the biotinylated peptides HAbio (150 nM) (*D*) and pp65bio (150 nM) (*E*), respectively, in the presence of increasing concentrations (7.8 nM – 1 μM) of non-biotinylated peptides MBP_{88-100}, MBP-HA, HA-MBP, and HA (*D*), as well as MBP_{85-97}, pp65-MBP, and pp65 (*E*) in the presence of HLA-DM (150 nM). Standard deviation and p-values are presented

REFERENCES

1. Todd J.A., Acha-Orbea H., Bell J.I., Chao N., Fronek Z., Jacob C.O., McDermott M., Sinha A.A., Timmerman L., Steinman L. // Science. 1988. V. 240. P. 1003–1009.
2. Huppa J.B., Davis M.M. // Nat. Rev. Immunol. 2003. V. 3. № 12. P. 973–983.
3. Cresswell P. // Annu. Rev. Immunol. 1994. V. 12. P. 259–293.
4. Ghosh P., Amaya M., Mellins E., Wiley D.C. Nature. 1995. V. 378. P. 457–462.
5. Jasanoff A., Park S.J., Wiley D.C. // Proc. Natl. Acad. Sci. USA. 1995. V. 92. № 21. P. 9900–9904.
6. Sherman M.A., Weber D.A., Jensen P.E. // Immunity. 1995. V. 3. № 2. P. 197–205.
7. Roche P.A., Furuta K. // Nat. Rev. Immunol. 2015. V. 15. № 4. P. 203–216.
8. Smith K.J., Pyrdol J., Gauthier L., Wiley D.C., Wucherpfennig K.W. // J. Exp. Med. 1998. V. 188. № 8. P. 1511–1520.
9. Yin Y., Li Y., Mariuzza R.A. // Immunol. Rev. 2012. V. 250. № 1. P. 32–48.
10. Li Y., Li H., Martin R., Mariuzza R.A. // J. Mol. Biol. 2000. V. 304. № 2. P. 177–188.
11. Hemmer B., Kerschensteiner M., Korn T. // Lancet Neurol. 2015. V. 14. № 4. P. 406–419.h
12. Sawcer S., Franklin R.J.M., Ban M. // Lancet Neurol. 2014. V. 13. № 7. P. 700–709.
13. Canto E., Oksenberg J.R. // Mult. Scler. 2018. V. 24. № 1. P. 75–79.
14. Zhang Q., Lin C.Y., Dong Q., Wang J., Wang W. // Autoimmun. Rev. 2011. V. 10. № 8. P. 474–481.
15. Ramagopalan S.V., Ebers G.C. // Genome Med. 2009. V. 1. № 11. P. 1–5.
16. Yin Y., Li Y., Kerzic M.C., Martin R., Mariuzza R.A. // EMBO J. 2011. V. 30. № 6. P. 1137–1148.
17. Hahn M., Nicholson M.J., Pyrdol J., Wucherpfennig K.W. // Nat. Immunol. 2005. V. 6. № 5. P. 490–496.
18. Li Y., Huang Y., Lue J., Quandt J.A., Martin R., Mariuzza R.A. // EMBO J. 2005. V. 24. P. 2968–2979.
19. Mamedov A.E., Zakharova M.Y., Favorova O.O., Kulakova O.G., Boyko A.N., Knorre V.D., Vorobieva N.A., Khurs E.N., Kiselev I.S., Baulina N.M., et al. // Dokl. Biochem. Biophys. 2019. V. 485. № 1. P. 115–118.
20. Mamedov A.E., Ponomarenko N.A., Belogurov A.A., Gabibov A.G. // Bull. Exp. Biol. Med. 2016. V. 161. № 3. P. 442–446.
21. Belogurov A.A., Kurkova I.N., Friboulet A., Thomas D., Misikov V.K., Zakharova M.Y., Suchkov S.V., Kotov S.V., Alehin A.I., Avalle B., et al. // J. Immunol. 2008. V. 180. № 2. P. 1258–1267.
22. Stern L.J., Brown J.H., Jardetzky T.S., Gorga J.C., Urban R.G., Strominger J.L., Wiley D.C. // Nature. 1994. V. 368. № 6468. P. 215–221.
23. Muixí L., Carrascal M., Alvarez I., Daura X., Martí M., Armengol M.P., Pinilla C., Abian J., Pujol-Borrell R., Jaraquemada D. // J. Immunol. 2008. V. 181. № 1. P. 795–807.
24. Yin L., Maben Z.J., Becerra A., Stern L.J. // J. Immunol. 2015. V. 195. № 2. P. 706–716.

Voltage-Sensing Domain of the Third Repeat of Human Skeletal Muscle $Na_V1.4$ Channel As a New Target for Spider Gating Modifier Toxins

M. Yu. Myshkin[1], A. S. Paramonov[1], D. S. Kulbatskii[1], E. A. Surkova[1], A. A. Berkut[1], A. A. Vassilevski[1], E. N. Lyukmanova[1,2], M. P. Kirpichnikov[1,2], Z. O. Shenkarev[1*]

[1]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997 Russia

[2]Biological Faculty, Lomonosov Moscow State University, Moscow, 119234 Russia

*E-mail: zakhar-shenkarev@yandex.ru

Received November 30, 2020; in final form, January 19, 2021

DOI: 10.32607/actanaturae.11279

ABSTRACT **Voltage-gated sodium channels (Na_V) have a modular architecture and contain five membrane domains. The central pore domain is responsible for ion conduction and contains a selectivity filter, while the four peripheral voltage-sensing domains (VSD-I/IV) are responsible for activation and rapid inactivation of the channel. "Gating modifier" toxins from arthropod venoms interact with VSDs, influencing the activation and/or inactivation of the channel, and may serve as prototypes of new drugs for the treatment of various channelopathies and pain syndromes. The toxin-binding sites located on VSD-I, II and IV of mammalian Na_V channels have been previously described. In this work, using the example of the Hm-3 toxin from the crab spider *Heriaeus melloteei*, we showed the presence of a toxin-binding site on VSD-III of the human skeletal muscle $Na_V1.4$ channel. A developed cell-free protein synthesis system provided milligram quantities of isolated (separated from the channel) VSD-III and its ^{15}N-labeled analogue. The interactions between VSD-III and Hm-3 were studied by NMR spectroscopy in the membrane-like environment of DPC/LDAO (1 : 1) micelles. Hm-3 has a relatively high affinity to VSD-III (dissociation constant of the complex Kd ~6 μM), comparable to the affinity to VSD-I and exceeding the affinity to VSD-II. Within the complex, the positively charged Lys25 and Lys28 residues of the toxin probably interact with the S1–S2 extracellular loop of VSD-III. The Hm-3 molecule also contacts the lipid bilayer surrounding the channel.**

KEYWORDS **cell-free protein synthesis, ligand–receptor interaction, NMR spectroscopy, sodium channels, gating modifier toxins.**

ABBREVIATIONS: DPC – dodecylphosphocholine, LDAO – n-dodecyl-N,N-dimethylamine-N-oxide; Na_V – voltage-gated sodium channel; VSD – voltage-sensing domain; VSD-III – voltage-sensing domain from the third repeat of α-subunit of human $Na_V1.4$ channel; TM – transmembrane; RM – reaction mixture.

INTRODUCTION

Voltage-gated Na^+-channels (Na_V) are transmembrane (TM) proteins responsible for the ascending phase of the action potential in excitable cells. These channels consist of a pore-forming α-subunit with which regulatory β-subunits are associated (*Fig. 1A*). The α-subunit includes four homologous repeats (I–IV), each of those containing a voltage-sensing domain (VSD, TM segments S1–S4) and S5–S6 segments that form the pore of the channel [1]. The β-subunits have one TM segment and an extracellular immunoglobulin domain [2]. The human genome contains 10 genes encoding the α-subunits of Na_V and four genes encoding β-subunits. The $Na_V1.4$ channel is expressed in skeletal muscle, and mutations in its α-subunit gene (*SCN4A*) lead to a number of congenital disorders of the musculoskeletal system, such as myotonia, paramyotonia, hyperkalemic and hypokalemic periodic paralysis, myasthenia gravis, and myopathy [3].

Na_Vs are targets for many neurotoxins from different organisms. At least eight receptor-binding sites for toxins have been identified in the VSD and the pore of the channel [4]. In the extracellular loops of VSDs of repeats II and IV, two canonical binding sites for spider

and scorpion toxins were identified (*Fig. 1A*) [5]. Toxins acting on VSD-IV (site 3) inhibit channel inactivation, and toxins acting on VSD-II (site 4) (e.g., "gating modifier" toxins of spiders) affect channel activation [5]. Although the extracellular interfaces of VSD-I and III can be partially closed by the immunoglobulin domains of the β-subunits [6, 7], these domains, involved in channel activation, may also contain toxin-binding sites available in some pathophysiological conditions. Thus, it was shown [8] that some toxins inhibit the activation of the chimeric $K_V2.1$ channel containing S3–S4 loops from VSD-I or III of the $Na_V1.2$ channel and do not inhibit the original $K_V2.1$ channel. The search for the binding site of neurotoxins in eukaryotic channels by site-directed mutagenesis is difficult, since the α-subunit of Na_V contains four VSDs, each of which can take part in the formation of a response to the toxin's action.

Earlier, we showed that the extracellular loop S3–S4 of VSD-I of the human $Na_V1.4$ channel is the main binding site for the Hm-3 toxin from the venom of the spider *Heriaeus melloteei* [9]. In addition, Hm-3 interacts with the S1–S2 extracellular loop of VSD-II, but with a much lower affinity [10]. The Hm-3 toxin consists of 35 amino acid residues and has a charge of +4 at neutral pH. The secondary structure of Hm-3 includes several β-turns and a β-hairpin formed by Cys23–Cys34 residues. The spatial structure of Hm-3 is stabilized by three disulfide bonds, which form the so-called "cystine knot" [11]. Several aromatic residues form a hydrophobic cluster on the surface of Hm-3; therefore, like other "gating modifier" spider toxins, Hm-3 has an affinity for membranes [11] and, apparently, attacks the VSDs from the membrane-bound state. Toxins that belong to this family are interesting not only as tools for the structural and functional study of Na_V, but they can also serve as prototypes for new drugs. For example, Hm-3 can block aberrant leakage currents (ω-currents) arising in the $Na_V1.4$ channel with mutations in VSD-I and II, leading to the development of periodic paralysis [9, 10].

In this work, using the Hm-3 toxin as an example, we have shown for the first time that a toxin-binding site is present in VSD-III of the human $Na_V1.4$ channel. In

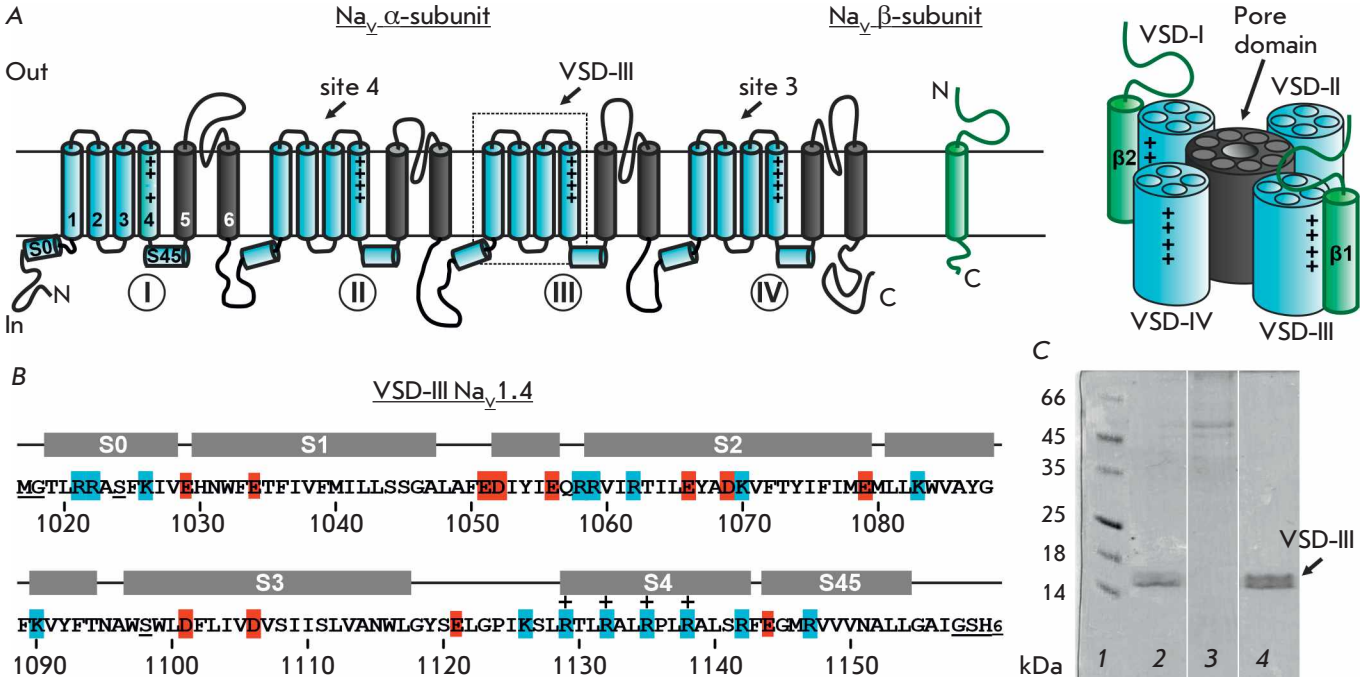


Fig. 1. (*A*) Spatial organization of eukaryotic Na_V channels in the membrane. (*B*) Amino acid sequence of VSD-III used in the study. Artificially introduced residues are underlined. The secondary structure is presented according to the known spatial structure of the channel [6] (PDB ID: 6AGF). The negatively and positively charged residues are highlighted in red and blue, respectively. Positively charged residues in the voltage sensor S4 helix are marked by a "+" sign. (*C*) Purification of VSD-III by Ni^{2+}-affinity chromatography. Lanes: *1* – molecular weight marker; *2* – solubilized RM pellet; *3* – column wash; and *4* – elution with 500 mM imidazole. Molecular weight of VSD-III: 16.3 kDa

order to do this, we used an alternative approach based on the production of a recombinant isolated (separated from the channel) VSD and an analysis of the binding sites by NMR spectroscopy. Several works have demonstrated that it is possible to perform structural NMR studies of isolated VSDs [12] and their complexes with toxins [9, 10].

MATERIALS AND METHODS

Isolated VSD-III (residues 1019–1157, *Fig. 1B*) was obtained using a dialysis-type conjugated cell-free synthesis system based on the S30 extract from *Escherichia coli* using protocols developed for other VSDs [9, 10]. The genetic construct for synthesizing VSD-III with the C-terminal His6-tag was cloned into the pIVEX2.3d plasmid vector, which provides a high efficiency in cell-free synthesis. The VSD-III sequence contains two Cys residues that are not involved in the formation of disulfide bonds. To reduce the tendency towards aggregation of the recombinant VSD-III, these residues were replaced by Ser (*Fig. 1B*, underlined). Cell-free synthesis was performed without adding membrane-mimicking components to the reaction mixture (RM). In this case, the synthesized VSD-III accumulated in the form of a precipitate with a purity of more than 90% (*Fig. 1C*). A ^{15}N-labeled analogue of VSD-III was synthesized using a ^{15}N isotope-enriched mixture of 16 amino acids (Cortecnet, Les Ulis, France) obtained from algae and the individual ^{15}N-labeled amino acids Asn, Gln, and Trp. Cysteine was not added to the synthesis reaction, since the VSD-III variant used in this work did not contain that amino acid. The yields of unlabeled and ^{15}N-labeled VSD-III samples were 0.5 and 0.35 mg per 1 mL of RM, respectively. For NMR studies, the precipitate containing the synthesized VSD-III was dissolved in a 10% dodecylphosphocholine (DPC) solution, purified by Ni^{2+} affinity chromatography in the presence of 0.5% DPC (*Fig. 1C*), and transferred to the target buffer (20 mM Tris-Ac, pH 5.5), and the N,N-dimethyldodecylamine-N-oxide (LDAO) detergent was added to a 1: 1 molar ratio with DPC. Previously, mixed DPC/LDAO micelles were used as a membrane-mimicking medium to study complexes of VSD-I and II with the Hm-3 toxin [9, 10]. Detergent concentrations were monitored by 1D ^{1}H NMR spectra. The NMR spectra were recorded on an AVANCE III 800 spectrometer (Bruker).

RESULTS AND DISCUSSION

The general appearance of the 2D ^{1}H,^{15}N correlation NMR spectrum of VSD-III (*Fig. 2A*) corresponded to the spectra of VSD-I and II obtained earlier [9, 10]. The observed small dispersion of $^{1}H^{N}$ signals is characteristic of helical TM proteins. However, the spectrum contained no more than 90 signals of backbone HN groups out of the 130–140 expected signals. In the corresponding spectral regions, six of the eight HN signals of the Gly residues and four of the five $HN^{\varepsilon 1}$ signals of the side chains of the Trp residues are presented. The absence of some signals in the spectrum, as well as the inhomogeneous intensity and half-width of the observed signals, is indicative of a conformational exchange in the μs–ms range. These processes are probably associated with the plasticity of the VSD-III structure and the dynamics of contacts between TM helices. The observed signal-broadening did not allow us to obtain assignment of the VSD-III NMR signals; therefore, the interaction with Hm-3 was studied qualitatively, without mapping of the binding site in the VSD.

Samples of unlabeled and ^{15}N-labeled Hm-3 were obtained using recombinant production in *E. coli* cells [9, 11]. To study the interaction of VSD-III with Hm-3, unlabeled Hm-3 was added stepwise to a sample of ^{15}N-labeled VSD in DPC/LDAO micelles to a 1 : 4 molar ratio of VSD/toxin. Detergent concentration was kept constant to prevent changes in the toxin distribution between the water phase and the micelles. According to the previously obtained data on the interaction of Hm-3 with DPC/LDAO micelles [9], ~97% of toxin molecules bound to micelles under the experimental conditions. After addition of the toxin, changes in the chemical shifts and amplitudes of some signals were observed in the spectrum of VSD-III (*Fig. 2B*). These changes were an indication that the VSD–toxin interaction was specific. The reversible process of formation–dissociation of the VSD/Hm-3 complex has a characteristic time in the μs–ms range, and for different VSD signals this exchange process is either fast or intermediate (on the NMR time scale). The dissociation constant of the complex was determined by approximating the dependence of the chemical shift of the VSD-III signals on the Hm-3 concentration (*Fig. 2B*), taking into account the contribution of the Hm-3/micelle interaction [9]. The obtained value (5.8 ± 3.8 μM) corresponded to the dissociation constant of the VSD-I/Hm-3 complex (6.2 ± 0.6 μM) [9] and was lower than the value for the complex with VSD-II (~11 μM) [10], which indicates stronger interaction of the toxin with VSD-I and VSD-III.

Back titration, when unlabeled VSD-III was added to a sample of ^{15}N-labeled Hm-3, showed that the positively charged residues Lys25 and Lys28 located in the β-hairpin of the toxin, as well as the Phe12 residue buried in the hydrophobic region of the micelle, are involved in the formation of a complex with VSD-III (*Fig. 3*). This binding site coincides with the sites responsible for the interaction of Hm-3 with VSD-I and II [9, 10]. In the course of these earlier studies, it was

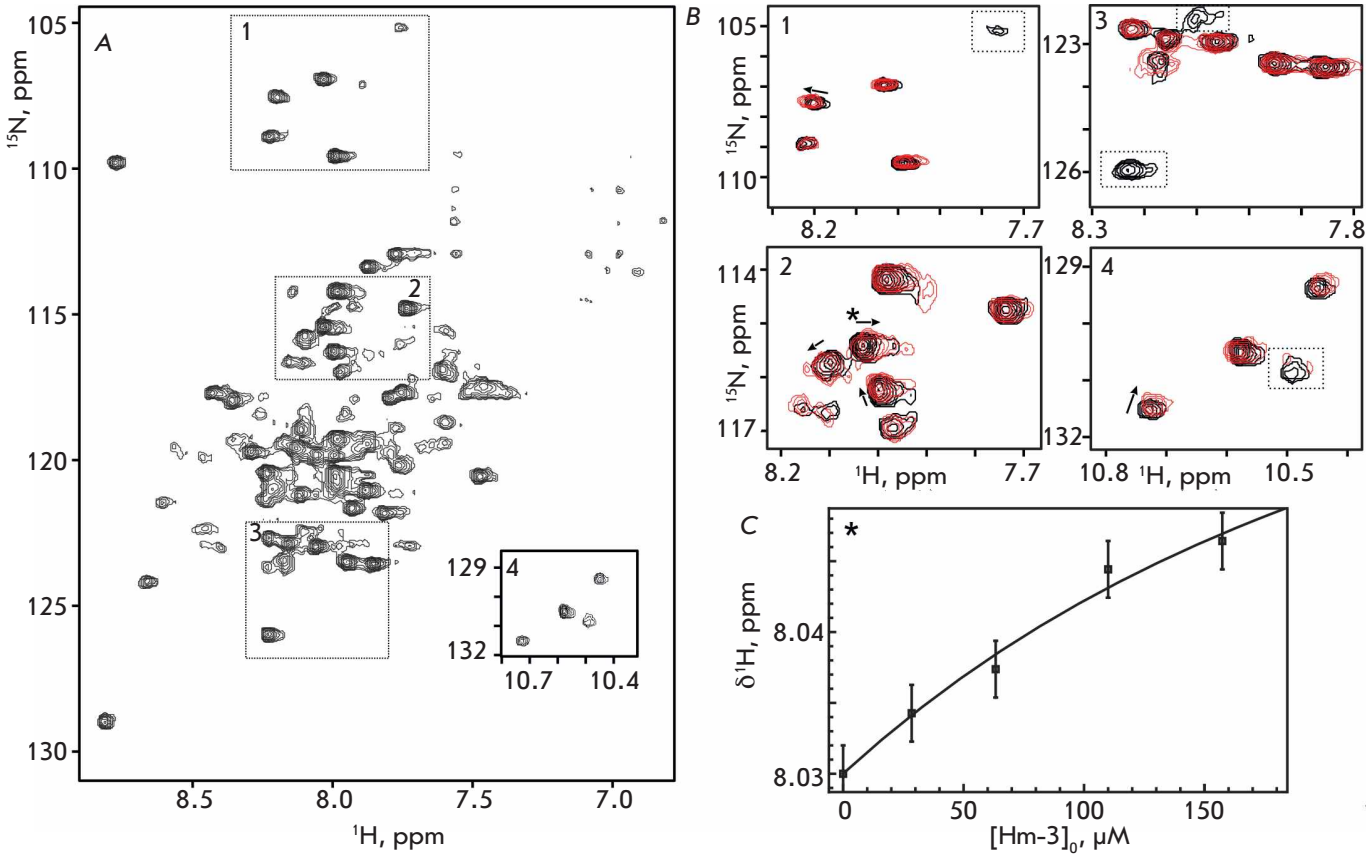


Fig. 2. The NMR study of ^{15}N-labelled VSD-III interaction with non-labelled Hm-3. (*A*) The 2D ^{1}H,^{15}N-TROSY spectrum of 43 μM VSD-III in DPC/LDAO micelles (45/45 mM, 800 MHz, pH 5.5, 45°C). (*B*) Overlay of the VSD-III spectral fragments, before (black) and after addition of 160 μM Hm-3 (red). Arrows indicate the direction of the changes in signal position. Dashed lines indicate the signals that disappear after the addition of Hm-3. (*C*) Changes in ^{1}H chemical shift of the HN signal at 8.03/115.4 ppm (marked by an asterisk on panel *B2*) approximated by the equation describing binding in the presence of an excess of detergent micelles

shown that the pair of charged Hm-3 residues (Lys25 and Lys28) can specifically interact with helical motifs containing two negatively charged residues (Asp or Glu) separated by two or three uncharged residues. In the VSD-III sequence, such motifs are found only in the S1–S2 extracellular loop and in the TM portion of the S2 helix (*Fig. 1B*). However, according to the well-known spatial structure of the human $Na_V1.4$ channel [6], the Glu1066 and Asp1069 residues are located deep in the TM portion of the S2 helix and their side chains are turned inside the VSD-III molecule. Taking into account the amphipathic properties of Hm-3 [11], we assume that the toxin cannot penetrate deep into the membrane and interact with these residues. Meanwhile, the side chains of the Glu1051, Asp1052, and Glu1056 residues located in the S1–S2 loop region are accessible to the solvent and can interact with the Hm-3 molecule bound to the membrane surface.

On the contrary, another extracellular loop of VSD-III, S3–S4, contains a single negatively charged residue Glu1121 and probably cannot act as a binding site for the Hm-3 toxin. This is consistent with the results of a previous study of the $K_V2.1$ chimeric channel containing the loop S3–S4 transplanted from the VSD-III channel $Na_V1.4$, during which no significant interaction with the Hm-3 toxin was revealed [9]. Thus, the data obtained indicate that the extracellular loop S1–S2 of the VSD-III of the human $Na_V1.4$ channel contains a site capable of interacting with "gating modifier" spider toxins. It should be noted that the study of toxin-binding sites located at the S1–S2 region of the VSD of Na_V channels using chimeric channels is

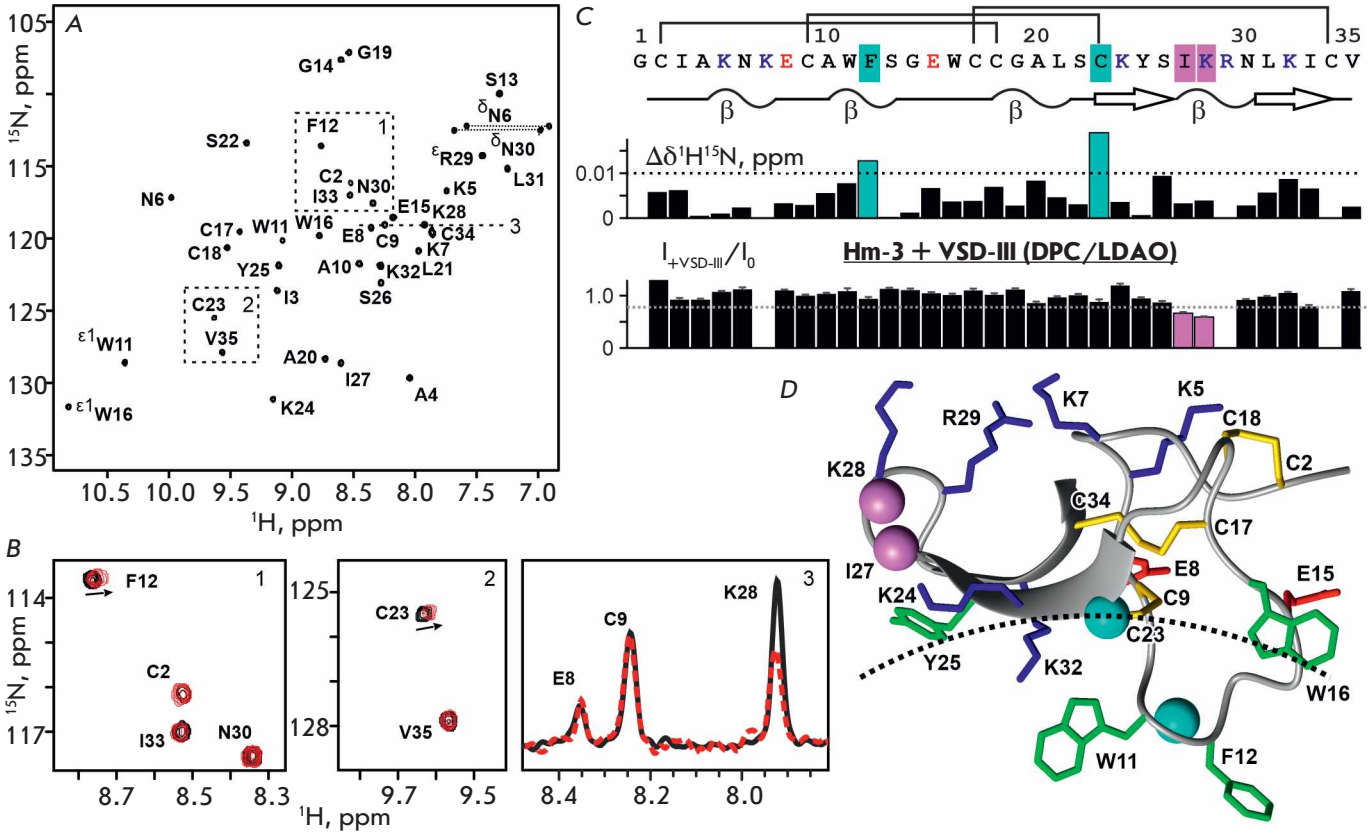


Fig. 3. The NMR study of interaction between ^{15}N-labelled Hm-3 and non-labelled VSD-III. (*A*) The 2D ^{1}H,^{15}N-HSQC spectrum of 30 μM Hm-3 in DPC/LDAO micelles (45/45 mM, 800 MHz, pH 5.5, 45°C). (*B*) Overlay of the Hm-3 spectra fragments, before (black) and after addition of 40 μM VSD-III (red, the final VSD-III/Hm-3 ratio is 3 : 2). Arrows indicate the direction of changes in the signal position. Panel *B3* depicts the selective decrease in Lys28 signal intensity after addition of VSD-III. (*C*, *D*) The primary, secondary, and spatial structures of Hm-3 toxin [11] (PDB ID: 2MQU), and the changes in the chemical shifts and intensities of Hm-3 signals after addition of VSD-III. The positively and negatively charged residues and disulphide bridges are shown. Residues whose signals undergo significant changes are highlighted (*C*) and marked by balls (*D*). The dashed line represents the detergent micelle surface [9]

apparently impossible. Attempts to transplant S1–S2 loops from various channels into K_V2.1 resulted in non-functional chimeras [13].

The system of cell-free synthesis of VSD-III developed in this work will make it possible to further investigate the interaction between the domain and other toxins and can also be used for screening drug prototypes that selectively interact with VSD-III. The proposed method for NMR study of Na_V pharmacology has an advantage over methods based on the study of chimeric channels, since it allows one to map the toxin residues important for interaction with voltage-sensing domains and study toxin-binding sites that are located not only in the S3–S4, but also in the S1–S2 loop. ●

This work was supported by the Russian Science Foundation (grant No. 16-14-10338).

REFERENCES

1. Catterall W.A. // J. Physiol. 2012. V. 590. P. 2577–2589.
2. O'Malley H.A., Isom L.L. // Annu. Rev. Physiol. 2015. V. 77. P. 481–504.
3. Cannon S.C. // Handbook Exp. Pharmacol. 2018. V. 246. P. 309–330.
4. Xu L., Ding X., Wang T., Mou S., Sun H., Hou T. // Drug. Discov. Today. 2019. V. 24. P. 1389–1397.
5. Stevens M., Peigneur S., Tytgat J. // Front. Pharmacol. 2011. V. 2. P. 71.
6. Pan X., Li Z., Zhou Q., Shen H., Wu K., Huang X., Chen J., Zhang J., Zhu X., Lei J., et al. // Science. 2018. V. 362.

P. eaau2486.
7. Shen H., Liu D., Wu K., Lei J., Yan N. // Science. 2019. V. 363. P. 1303–1308.
8. Bosmans F., Martin-Eauclaire M.F., Swartz K.J. // Nature. 2008. V. 456. P. 202–208.
9. Männikkö R., Shenkarev Z.O., Thor M.G., Berkut A.A., Myshkin M.Y., Paramonov A.S., Kulbatskii D.S., Kuzmin D.A., Sampedro Castañeda M., King L., et al. // Proc. Natl. Acad. Sci. USA. 2018. V. 115. P. 4495–4500.
10. Myshkin M.Y., Männikkö R., Krumkacheva O.A., Kulbatskii D.S., Chugunov A.O., Berkut A.A., Paramonov A.S., Shulepko M.A., Fedin M.V., Hanna M.G., et al. // Front. Pharmacol. 2019. V. 10. P. 953.
11. Berkut A.A., Peigneur S., Myshkin M.Y., Paramonov A.S., Lyukmanova E.N., Arseniev A.S., Grishin E.V., Tytgat J., Shenkarev Z.O., Vassilevski A.A. // J. Biol. Chem. 2015. V. 290. P. 492–504.
12. Shenkarev Z.O., Paramonov A.S., Lyukmanova E.N., Shingarova L.N., Yakimov S.A., Dubinnyi M.A., Chupin V.V., Kirpichnikov M.P., Blommers M.J., Arseniev A.S. // J. Am. Chem. Soc. 2010. V. 132. P. 5630–5637.
13. Alabi A., Bahamonde M., Jung H., Kim J.I., Swartz K.J. // Nature. 2007. V. 450. P. 370–375.

A Novel Modulator of STIM2-Dependent Store-Operated Ca^{2+} Channel Activity

A. Y. Skopin[1], A. D. Grigoryev[1], L. N. Glushankova[1], A. V. Shalygin[1], G. Wang[2], V. G. Kartzev[3], E. V. Kaznacheyeva[1*]

[1]Institute of Cytology of Russian Academy of Sciences, St. Petersburg, 194064 Russia
[2]College of Pharmaceutical Sciences, Soochow University, Suzhou, Jiangsu, 215123 China
[3]InterBioscreen Ltd., Chernogolovka, 142432 Russia
*E-mail: evkazn@incras.ru

Received November 25, 2020; in final form, December 24, 2020
DOI: 10.32607/actanaturae.11269

ABSTRACT **Store-operated Ca^{2+} entry is one of the main pathways of calcium influx into non-excitable cells, which entails the initiation of many intracellular processes. The endoplasmic reticulum Ca^{2+} sensors STIM1 and STIM2 are the key components of store-operated Ca^{2+} entry in mammalian cells. Under physiological conditions, STIM proteins are responsible for store-operated Ca^{2+} entry activation. The STIM1 and STIM2 proteins differ in their potency for activating different store-operated channels. At the moment, there are no selective modulators of the STIM protein activity. We screened a library of small molecules and found the 4-MPTC compound, which selectively inhibited STIM2-dependent store-operated Ca^{2+} entry (IC_{50} = 1 μM) and had almost no effect on the STIM1-dependent activation of store-operated channels.**

KEYWORDS **calcium, store-operated Ca^{2+} entry, STIM1, STIM2, 2-APB, Orai.**

ABBREVIATIONS **2-APB – 2-aminoethoxydiphenyl borate; 4-MPTC – 4-methyl-2-(2-propylpyridin-4-yl)-N-(pyridin-2-yl)thiazole-5-carboxamide; CC1 – coiled-coil 1 domain; DMSO – dimethyl sulfoxide; STIM – stromal-interacting molecule; SOAR – STIM-ORAI activating region; Tg – thapsigargin; ER – endoplasmic reticulum.**

INTRODUCTION

An increase in the concentration of cytoplasmic Ca^{2+} ions is one of the common cellular responses to extracellular stimulation of membrane receptors by physiologically active substances that trigger a wide range of intracellular cascades. Under physiological conditions, the intracellular Ca^{2+} response to an agonist includes not only entry of extracellular Ca^{2+} into the cell, but also depletion of the intracellular Ca^{2+} stores located in the endoplasmic reticulum (ER) [1]. Plasma membrane channel-mediated Ca^{2+} entry into the cell in response to the depletion of intracellular Ca^{2+} stores or store-operated Ca^{2+} entry [2] provides a significant part of the Ca^{2+} ion influx into the cell. The entry is induced by STIM proteins (STIM1 and STIM2), which are Ca^{2+} sensors in the ER lumen. The STIM1 protein, which is the main activator of store-operated Ca^{2+} entry, was the first to be characterized [3, 4]. The STIM1 and STIM2 proteins differ in their affinity for Ca^{2+} ions and ability to interact with plasma membrane channels [5]. STIM2 is more sensitive to small changes in the concentration of stored Ca^{2+} and is a weaker activator of store-operated Ca^{2+} entry than STIM1. STIM1 is most likely responsible for the cellular Ca^{2+} response to an extracellular signal, while STIM2 regulates the basal levels of cytosolic and stored Ca^{2+} [6]. In addition, STIM2 facilitates STIM1 transition to the active state [7]. Under physiological conditions, STIM1 and STIM2 activate various store-operated channels in the cell [8], which are formed by proteins belonging to the Orai [9, 10] and TRP [11–13] families. STIM proteins are involved in a wide range of pathologies. For instance, a long-term increase in the neuronal Ca^{2+} concentration, which is caused by an enhanced activity of STIM proteins and leads to cell death, is observed in Huntington's disease [14, 15], Alzheimer's disease [16, 17], cerebral ischemia [18], and traumatic brain injury [19, 20]. Changes in STIM expression levels are typical for several breast cancers [21] and colon carcinoma [22]. Thus, changes in the activity of STIM proteins, in particular decreased

STIM2 activity, may possess a potential therapeutic effect. In basic research, a STIM2 activity modulator would be an essential tool to be used to distinguish between STIM1- and STIM2-mediated signaling pathways, because such pharmacological agents are currently unavailable.

Researchers have actively used a wide range of store-operated Ca^{2+} entry inhibitors. Most of these inhibitors modulate the activity of store-operated Ca^{2+} channels. However, these compounds are often poorly characterized and have more than one target. One of the most commonly used compounds, 2-aminoethoxydiphenyl borate (2-APB), was first characterized as a blocker of IP_3-induced Ca^{2+} release [23]. It is now widely used as a store-operated Ca^{2+} entry inhibitor at concentrations exceeding 50 μM. In addition, 2-APB, at a concentration of 5 μM, can potentiate store-operated entry [24]. The mechanism of 2-APB action is not fully understood; this compound is known to have several targets and, in particular, to exert a modulatory effect on the activity of various channels; e.g., TRPV [25, 26] and Orai3 [27] channels. The 2-APB compound also enhances non-specific Ca^{2+} leak from the ER lumen [28].

When ER Ca^{2+} stores are filled, STIM proteins are in an inactive conformation stabilized by the interaction between the CC1 (Coiled-Coil 1) and SOAR (STIM-Orai Activating Region) domains. Following Ca^{2+} store depletion, STIM proteins undergo multimerization, change their conformation, and expose the SOAR domain for interaction with plasma membrane channels [29]. The 2-APB compound, at concentrations of about 10 μM, is known to induce store-operated Ca^{2+} entry by transforming STIM2 into its active conformation [30]. On the contrary, 2-APB at a higher concentration (50 μM) stabilizes an inactive STIM1 conformation by enhancing the interaction between the CC1 and SOAR domains. Thus, it inhibits the interaction of the SOAR domain with Orai1 channels and the activation of the channels. Interestingly, increased Orai1 expression partially reverses this action [31].

Thus, 2-APB directly interacts with STIM proteins and provides a good basis for the search for a more selective modulator of store-operated Ca^{2+} entry. In this work, we have tested a library of 250 chemical compounds received from InterBioScreen Ltd. possessing a chemical structure similar to that of 2-APB, in order to identify a selective modulator of STIM2 activity. A 4-MPTC compound was found to inhibit STIM2-dependent Ca^{2+} entry (IC_{50} = 1 μM) but had almost no effect on the STIM1-mediated mechanism of store-operated channel activation. The other 249 compounds from the library had a divergent, and non-selective, effect.

EXPERIMENTAL

Cell lines

The following HEK293-derived cell lines, kindly provided by Jonathan Soboloff and Mohamed Trebak, were used in the study: STIM1Orai3 (a cell line expressing exogenous STIM1-YFP and Orai3-CFP proteins), STIM2Orai3 (a cell line expressing exogenous STIM2-YFP and Orai3-CFP proteins) [32], STIM1 KO (a CRISPR/Cas9-mediated STIM1 gene knockout cell line), STIM2 KO (a CRISPR/Cas9-mediated STIM2 gene knockout cell line), and Orai3 KO (a CRISPR/Cas9-mediated Orai3 knockout cell line) [30]. The cell lines were cultured in a DMEM medium (Biolot, Russia) supplemented with 10% fetal bovine serum, as well as the antibiotics penicillin (100 U/ml) and streptomycin (0.1 mg/ml) at 37°C and 5% CO_2.

Fluorescence analysis

Changes in the intracellular Ca^{2+} concentration were measured using a Fluo-4 AM calcium indicator (Thermo Fisher Scientific, USA). The cells were plated into 96-well culture plates 48 h prior to the analysis. The cells were first incubated in a HBSS solution (2 mM $CaCl_2$, 130 mM NaCl, 25 mM KCl, 1.2 mM $MgCl_2$, 10 mM HEPES, and 10 mM glucose) containing 5 μM Fluo-4 AM for 1 h and then in a HBSS solution supplemented with either 4-MPTC (InterBioScreen Ltd., Russia) or 1% DMSO (Sigma-Aldrich, USA) for 30 min. Measurements were performed in the presence of 2 mM calcium in the extracellular solution using a FLUOstar Omega microplate reader (BMG Labtech, Germany). Data are presented as Fluo-4 fluorescence intensity values normalized to the basal fluorescence value.

Electrophoresis and immunoblotting

The cells were grown in 60-mm Petri dishes and then lysed by adding a protease inhibitor cocktail. Proteins were separated by 8% denaturing PAGE. The proteins were transferred to a nitrocellulose membrane using a semi-dry transfer unit (Hoefer Pharmacia Biotech., Germany). Primary antibodies to STIM1 (Cell Signaling #4917, USA), STIM2 (Cell Signaling #5668, USA), and α-tubulin (Sigma-Aldrich #T6074, USA) were diluted at a ratio of 1 : 1000. Next, secondary anti-mouse IgG antibodies (Sigma-Aldrich #A0168, USA) against α-tubulin and anti-rabbit IgG antibodies (Sigma-Aldrich #A0545, USA) against STIM1 and STIM2 were used. Blots were visualized on a BioRad Cell Imaging System (Bio-Rad Laboratories, Inc., USA).

Low-molecular-weight compounds for screening, including 4-MPTC, were kindly provided by InterBioScreen Ltd. (ibscreen.com) in dry form. The compounds

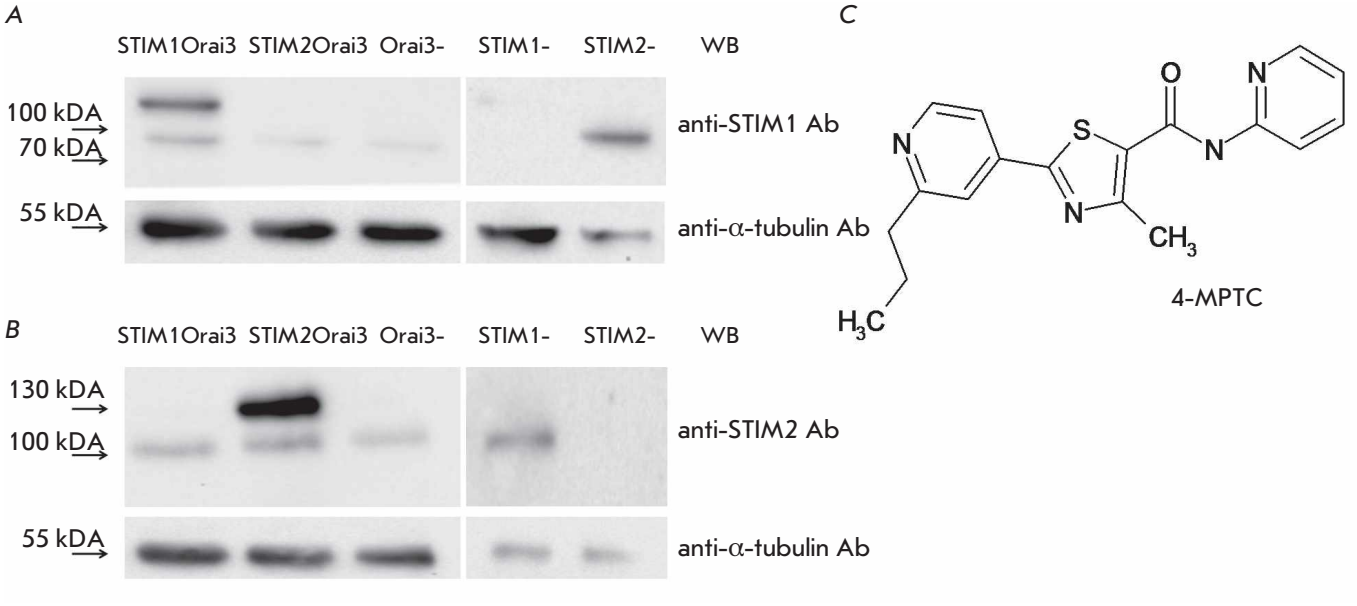


Fig. 1. Expression levels of STIM proteins in the STIM1Orai3, STIM2Orai3, Orai3 KO, STIM1 KO, and STIM2 KO cell lines. (*A*) Western blot using anti-STIM1 antibodies. (*B*) Western blot using anti-STIM2 antibodies. Anti-α-tubulin antibodies were used as a control to assess the uniformity of sample loading. (*C*) Structural formula of 4-MPTC

were dissolved in DMSO to a final concentration of 10 mM.

Statistical analysis

Statistical analysis was performed using the Origin 8 software. The results of fluorescence measurements were checked for normality using the Fisher's test. Data groups were compared using the Bonferroni test. Statistically significant differences are denoted in figures as follows: "*" – the confidence interval of $p < 0.05$, "**" – $p < 0.01$, "***" – $p < 0.001$; "n.s." – not statistically significant differences.

RESULTS AND DISCUSSION

In order to search for low-molecular-weight compounds that modulate the activity of STIM2 proteins, we used a model cell line derived from HEK293 cells stably expressing exogenous STIM2 and Orai3 proteins (STIM2Orai3 cell line) (*Fig. 1A*). The effect of the test compounds on the amplitude of a cellular Ca^{2+} signal in response to the depletion of intracellular Ca^{2+} stores was recorded using the Fluo-4 AM calcium indicator. Intracellular Ca^{2+} stores were depleted by adding 1 μM thapsigargin (Tg), a selective inhibitor of the ER Ca^{2+} pump, to the extracellular solution. At the first stage, the effect of the library of 2-APB analogs on the Tg-induced Ca^{2+} response was tested. For this purpose, the cells were incubated in HBSS solutions containing one of the 250 test compounds (at a concentration of 100 μM) for 30 min prior to starting the experiments. Next, the amplitude of the Ca^{2+} response to the addition of 1 μM Tg was assessed. As a result of library screening, we selected 4-MPTC (*Fig. 1C*), the compound that most strongly affected the Tg-induced Ca^{2+} response in STIM2Orai3 cells: the Ca^{2+} response was inhibited by 39 ± 3% compared to that in the cells incubated in a solution supplemented with 1% DMSO (*Fig. 2A*). Since 4-MPTC significantly inhibits the Tg-induced Ca^{2+} response in cells with increased STIM2 and Orai3 levels, we may suggest that 4-MPTC modulates the activity of these proteins. The direct action of 4-MPTC on Orai3 is supported by the fact that 2-APB can activate the Orai3 channel [27]. To test the effect of 4-MPTC on Orai3 channels, HEK293 cells with Orai3 knockout (the Orai3 KO cell line) were used. Incubation of Orai3 KO cells with 4-MPTC changed the shape of the Tg-induced Ca^{2+} response and decreased its amplitude by 12 ± 3% (*Fig. 2B*). Furthermore, incubation of HEK293 cells expressing exogenous STIM1 and Orai3 proteins (the STIM1Orai3 cell line) with 4-MPTC did not inhibit the amplitude of the Tg-induced Ca^{2+} response (*Fig. 2B*) and, therefore, did not decrease the activity of the Orai3 channels. Hence, the Orai3 protein is not a selective target for 4-MPTC.

The activity of store-operated channels in a cell is known to be modulated by both the STIM1 and STIM2 proteins [8]. The predominant pathway of

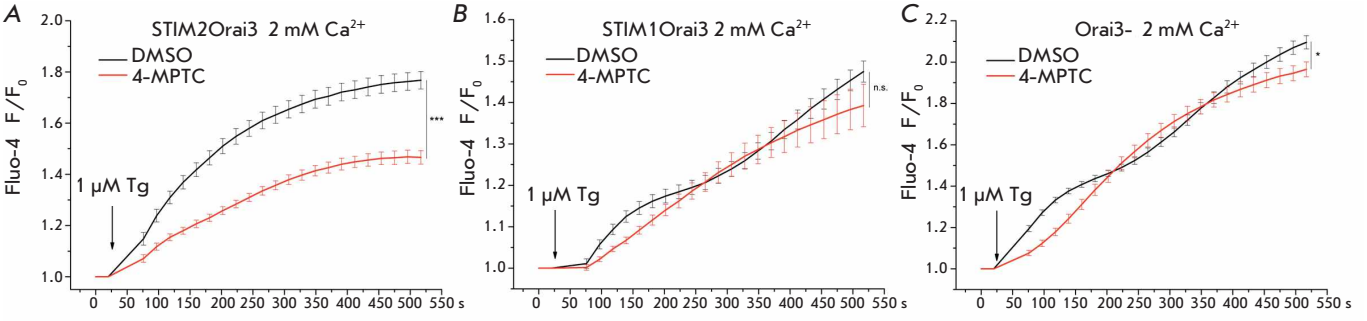


Fig. 2. Effect of the 4-MPTC compound on the Tg-induced calcium response. Measurements were performed in cell lines with (*A*) exogenous STIM2 and Orai3 protein expression, (*B*) exogenous STIM1 and Orai3 protein expression, and (*C*) Orai3 protein knockout. The dependence of Fluo-4 fluorescence, normalized to the basal fluorescence level, on time is presented. Prior to starting the experiment, the cells were incubated in HBSS supplemented with 100 μM 4-MPTC for 30 min. Control cells were incubated in HBSS containing 1% DMSO for 30 min. Data are presented as means ± s.e.m. (n = 12)

store-operated entry activation can be modulated through either the STIM1 protein or the STIM2 protein by changing their expression levels. HEK293 cells expressing exogenous STIM1 and Orai3 proteins were used to test the effect of 4-MPTC on STIM1. As mentioned above, incubation of STIM1Orai3 cells with 4-MPTC changes the shape of the Tg-induced Ca^{2+} response without decreasing its amplitude (*Fig. 2B*). Since 4-MPTC significantly reduced the Ca^{2+} response amplitude but did not alter the curve's shape in STIM2Orai3 cells (*Fig. 2A*), we may suggest that this compound affects the pathway of store-operated calcium entry activation through STIM2, but not through the STIM1 protein. A change in the curve's shape for the Orai3 KO and STIM1Orai3 cell lines is quite typical and reflects a decrease in the rate of the Ca^{2+} response. Since the endogenous STIM2 protein is present in Orai3 KO and STIM1Orai3 cells (*Fig. 1B*), 4-MPTC can reduce its activity and, thereby, change the dynamics of both the release of Ca^{2+} from the store into the cytoplasm and the entry of extracellular Ca^{2+} ions. Knockout of STIM2 using short interfering RNAs results in a similar effect on the Ca^{2+} response; it decreases Ca^{2+} release from the store [33] and subsequent Ca^{2+} entry [4, 34]. Cell lines overexpressing STIM proteins (STIM1Orai3 and STIM2Orai3) contain endogenous STIM1 and STIM2 (*Fig. 1A,B*), which complicates data interpretation. Therefore, we further used STIM1 (the STIM1 KO cell line) and STIM2 knockout cells (the STIM2 KO cell line), which are devoid of this drawback (*Fig. 1A,B*).

When STIM1 expression is completely suppressed, the STIM2 protein becomes the key and only activator of store-operated Ca^{2+} entry [4]. Pre-incubation of STIM1 KO cells with 4-MPTC decreased the Tg-induced Ca^{2+} response by 57 ± 8% compared to that in control cells (incubation with 1% DMSO) (*Fig. 3A*). It should be noted that 4-MPTC more effectively inhibits store-operated Ca^{2+} entry under these conditions. For example, the Tg-induced Ca^{2+} response was inhibited by 57% in STIM1 KO cells and by only 39% in STIM2Orai3 cells. A significant change in the shape of the Tg-induced response is observed after incubation of STIM2-knockout cells in which the STIM1 protein is the only activator of store-operated Ca^{2+} entry with 4-MPTC. The calcium concentration increases more slowly in these cells than in the control cells, with the maximum Ca^{2+} response amplitude being 61 ± 5% higher compared to that in the control (*Fig. 3B*). 4-MPTC was experimentally demonstrated to act divergently in STIM1 KO and STIM2 KO cell lines: it inhibits the Ca^{2+} response through the STIM2-dependent pathways and enhances it through the STIM1 pathways. Thus, the selected compound, 4-MPTC, enables differentiation between the pathways activating store-operated Ca^{2+} entry through different STIM proteins; however, the mechanism of action of this compound requires further clarification.

The 4-MPTC compound has a typical concentration–effect curve (*Fig. 4*). We analyzed the effect of 4-MPTC at a concentration of 0.001, 0.1, 1, 10, and 100 μM on the Tg-induced Ca^{2+} response in STIM2Orai3 cells. The half-maximal inhibitory concentration (IC_{50}), calculated by curve-fitting, equals 1 μM.

Thus, given our findings, we may conclude that the use of 4-MPTC in cell lines expressing predominantly the STIM2 protein (STIM1 KO and STIM2Orai3) significantly inhibits the amplitude of the Tg-induced Ca^{2+} response, while the use of 4-MPTC in cell lines producing predominantly the STIM1 protein (STIM2

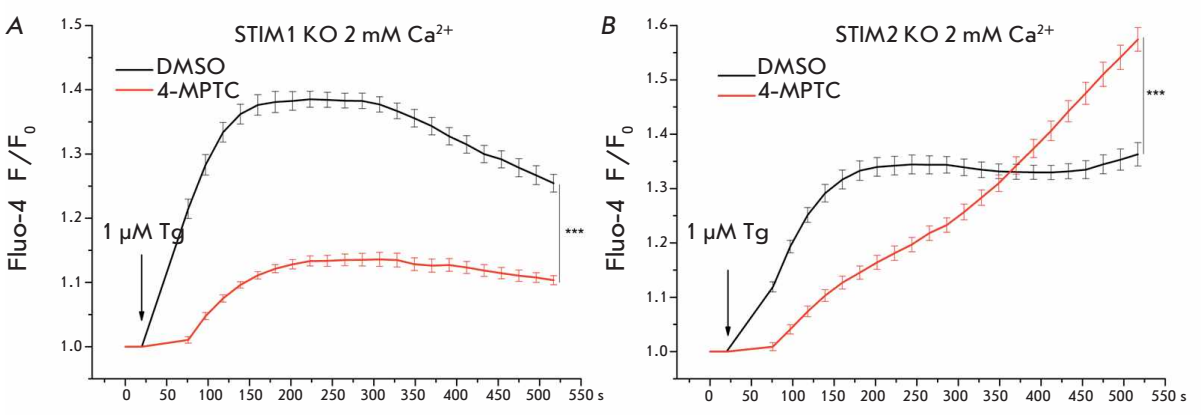


Fig. 3. Effect of the 4-MPTC compound on the Tg-induced calcium response. Measurements were performed in (*A*) STIM1 knockout and (*B*) STIM2 knockout cells. The dependence of the Fluo-4 fluorescence, normalized to the basal fluorescence level, on time is presented. Prior to starting the experiment, the cells were incubated in HBSS supplemented with 100 μM 4-MPTC for 30 min. Control cells were incubated in HBSS containing 1% DMSO for 30 min. Data are presented as means ± s.e.m. (n = 12)

KO, STIM1Orai3), on the contrary, changes the shape of the Ca^{2+} response curve, without decreasing its amplitude. Thus, 4-MPTC selectively inhibits store-operated Ca^{2+} entry via the STIM2-mediated pathway, but not the STIM1-mediated pathway.

Despite the fact that 2-APB is widely used as a store-operated Ca^{2+} entry inhibitor, it does not appear to selectively inhibit store-operated Ca^{2+} entry and also has a divergent concentration-dependent effect. 2-APB derivatives have been investigated in the search for an inhibitor lacking these disadvantages [35–41]. Most of the identified compounds inhibit store-operated Ca^{2+} entry at lower concentrations than 2-APB but are at the same time unable to activate Ca^{2+} entry at certain concentrations; in other words, they have better inhibitory properties than the parent compound. More attention in the search for new inhibitors of store-operated Ca^{2+} entry has been paid to the STIM1-dependent pathway of activation, while the STIM2-mediated pathway often has remained unexplored. For example, MDA-MB-231 cells, in which STIM1 and Orai1 proteins play a key role in store-operated Ca^{2+} entry, as well as HEK293 cells expressing STIM1 and Orai-family proteins, have been used as model cell lines in experiments [42, 43]. A study of the compounds DPB163-AE and DPB162-AE demonstrated that they interact differently with STIM1 and STIM2 but eventually inhibit store-operated Ca^{2+} entry through both proteins [37]. The 4-MPTC compound, identified in our study, has an inhibitory effect on the STIM2-mediated pathway and does not inhibit Ca^{2+} entry through the STIM1-dependent pathway.

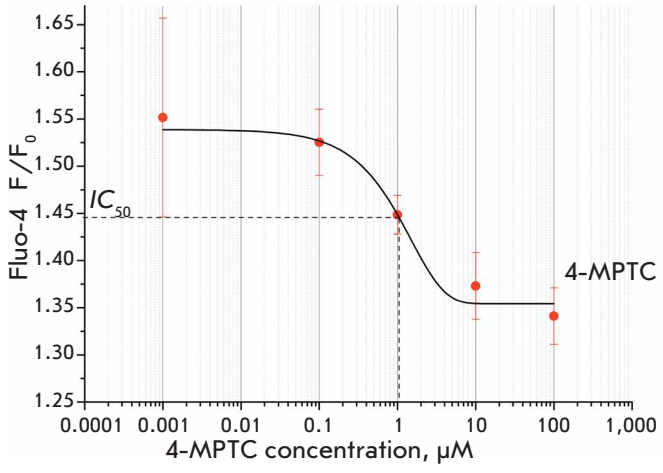


Fig. 4. Dependence of the Tg-induced calcium response amplitude on the 4-MPTC concentration. Measurements were performed in a cell line expressing exogenous STIM2 and Orai3 proteins. The Fluo-4 fluorescence intensity at minute 9 of the experiment, normalized to the basal fluorescence level, is presented. Prior to the experiment, the cells were incubated at various concentrations of 4-MPTC (0.001, 0.1, 1, 10, and 100 μM) for 30 min. Data are presented as means ± s.e.m. (n = 6). The half-maximum inhibitory level and half-maximum inhibitory concentration (IC_{50} = 1 μM) are denoted by dotted lines

CONCLUSION

Screening of a library of structure 2-APB analogs has yielded an 4-MPTC compound that has an inhibitory effect on the Tg-induced Ca^{2+} response through the STIM2-dependent pathway of Ca^{2+} influx but does not inhibit Ca^{2+} entry through the STIM1-dependent pathway. The mechanism of action of this compound on the STIM2 protein requires further investigation. ●

This study was supported by the Russian Foundation for Basic Research (grant No. 17-54-80006; AS, AG, LG, EK) and Russian Science Foundation (grant No. 19-14-00114; AS, ASh, EK).

REFERENCES

1. Putney J.W. // Cell Calcium 1986. V. 7. № 1. P. 1–12.
2. Yao Y., Tsien R.Y. // J. Gen. Physiol. 1997. V. 109. № 6. P. 703–715.
3. Zhang S.L., Yu Y., Roos J., Kozak J.A., Deerinck T.J., Ellisman M.H., Stauderman K.A., Cahalan. M.D. // Nature. 2005. V. 437. № 7060. P. 902–905.
4. Liou J., Kim M.L., Heo W.D., Jones J.T., Myers J.W., Ferrell J.E., Meyer T. // Curr. Biol. 2005. V. 15. № 13. P. 1235–1241.
5. Stathopulos P.B., Zheng L., Li G., Plevin M.J., Ikura M. // Cell. 2008. V. 135. № 1. P. 110–122.
6. Brandman O., Liou J., Park W.S., Meyer T. // Cell. 2007. V. 131. № 7. P. 1327–1339.
7. Ong H.L., De Souza L.B., Zheng C., Cheng K.T., Liu X., Goldsmith C.M., Feske S., Ambudkar I.S. // Sci. Signaling. 2015. V. 8. № 359. P. 3.
8. Shalygin A., Skopin A., Kalinina V., Zimina O., Glushankova L., Mozhayeva G.N., Kaznacheyeva E. // J. Biol. Chem. 2015. V. 290. № 8. P. 4717–4727.
9. Prakriya M. // Curr. Topics Membranes. 2013. V. 71. P. 1–32.
10. Prakriya M., Feske S., Gwack Y., Srikanth S., Rao A., Hogan P.G. // Nature. 2006. V. 443. № 7108. P. 230–233.
11. Skopin A., Shalygin A., Vigont V., Zimina O., Glushankova L., Mozhayeva G.N., Kaznacheyeva E. // Biochimie. 2013. V. 95. № 2. P. 347–353.
12. Kaznacheyeva E., Glushankova L., Bugaj V., Zimina O., Skopin A., Alexeenko V., Tsiokas L., Bezprozvanny I., Mozhayeva G.N. // J. Biol. Chem. 2007. V. 282. № 32. P. 23655–23662.
13. Cheng K.T., Ong H.L., Liu X., Ambudkar I.S. // Adv. Exp. Med. Biol. 2011. V. 704. P. 435–449.
14. Vigont V., Kolobkova Y., Skopin A., Zimina O., Zenin V., Glushankova L., Kaznacheyeva E. // Front. Physiol. 2015. V. 6. P. 337.
15. Nekrasov E.D., Vigont V.A., Klyushnikov S.A., Lebedeva O.S., Vassina E.M., Bogomazova A.N., Chestkov I.V., Semashko T.A., Kiseleva E., Suldina L.A., et al. // Mol. Neurodegen. 2016. V. 11. P. 27.
16. Ryazantseva M., Skobeleva K., Glushankova L., Kaznacheyeva E. // J. Neurochem. 2016. V. 136. № 5. P. 1085–1095.
17. Ryazantseva M., Skobeleva K., Kaznacheyeva E. // Biochimie. 2013. V. 95. № 7. P. 1506–1509.
18. Berna-Erro A., Braun A., Kraft R., Kleinschnitz C., Schuhmann M.K., Stegner D., Wultsch T., Eilers J., Meuth S.G., Stoll G., Nieswandt B. // Sci. Signaling. 2009. V. 2. № 93. P. ra67.
19. Rao W., Zhang L., Peng C., Hui H., Wang K., Su N., Wang L., Dai S.H., Yang Y.F., Chen T., Luo P., Fei Z. // Biochim. Biophys. Acta – Mol. Basis Dis. 2015. V. 1852. № 11. P. 2402–2413.
20. Hou P.F., Liu Z.H., Li N., Cheng W.J., Guo S.W. // Cell. Mol. Neurobiol. 2015. V. 35. № 2. P. 283–292.
21. Miao Y., Shen Q., Zhang S., Huang H., Meng X., Zheng X., Yao Z., He Z., Lu S., Cai C., Zou F. // Breast Cancer Res. 2019. V. 21. № 1. P. 99.
22. Sobradillo D., Hernández-Morales M., Ubierna D., Moyer M.P., Núñez L., Villalobos C. // J. Biol. Chem. 2014. V. 289. № 42. P. 28765–28782.
23. Maruyama T., Kanaji T., Nakade S., Kanno T., Mikoshiba K. // J. Biochem. 1997. V. 122. № 3. P. 498–505.
24. Prakriya M., Lewis R.S. // J. Physiol. 2001. V. 536. № 1. P. 3–19.
25. Hu H.Z., Gu Q., Wang C., Colton C.K., Tang J., Kinoshita-Kawada M., Lee L.Y., Wood J.D., Zhu M.X. // J. Biol. Chem. 2004. V. 279. № 34. P. 35741–35748.
26. Singh A.K., Saotome K., McGoldrick L.L., Sobolevsky A.I. // Nat. Commun. 2018. V. 9. № 1. P. 2465.
27. Schindl R., Bergsmann J., Frischauf I., Derler I., Fahrner M., Muik M., Fritsch R., Groschner K., Romanin C. // J. Biol. Chem. 2008. V. 283. № 29. P. 20261–20267.
28. Missiaen L., Callewaert G., De Smedt, H. Parys J.B. // Cell Calcium. 2001. V. 29. № 2. P. 111–116.
29. Soboloff J., Rothberg B.S., Madesh M., Gill D.L. // Nat. Rev. Mol. Cell Biol. 2012. V. 13. № 9. P. 549–565.
30. Emrich S.M., Yoast R.E., Xin P., Zhang X., Pathak T., Nwokonko R., Gueguinou M.F., Subedi K.P., Zhou Y., Ambudkar I.S., et al. // J. Biol. Chem. 2019. V. 294. № 16. P. 6318–6332.
31. Wei M., Zhou Y., Sun A., Ma G., He L., Zhou L., Zhang S., Liu J., Zhang S.L., Gill D.L. Wang Y. // Pflugers Arch. Eur. J. Physiol. 2016. V. 468. № 11–12. P. 2061–2074.
32. Parvez S., Beck A., Peinelt C., Soboloff J., Lis A., Monteilh-Zoller M., Gill D.L., Fleig A., Penner R. // FASEB J. 2008. V. 22. № 3. P. 752–761.
33. Thiel M., Lis A., Penner R. // J. Physiol. 2013. V. 591. № 6. P. 1433–1445.
34. Bird S., Hwang S.Y., Smyth J.T., Fukushima M., Boyles R.R., Putney J.W. // Curr. Biol. 2009. V. 19. № 20. P. 1724–1729.
35. Djillani A., Doignon I., Luyten T., Lamkhioued B., Gangloff S. C., Parys J. B., Nüße O., Chomienne C., Dellis O. // Cell Calcium. 2015. V. 58. № 2. P. 171–185.
36. Djillani A., Nüße O., Dellis O. // Biochim. Biophys. Acta – Mol. Cell Res. 2014. V. 1843. № 10. P. 2341–2347.
37. Goto J.I., Suzuki A.Z., Ozaki S., Matsumoto N., Nakamura T., Ebisui E., Fleig A., Penner R., Mikoshiba K. // Cell Calcium. 2010. V. 47. № 1. P. 1–10.
38. Hendron E., Wang X., Zhou Y., Cai X., Goto J.I., Mikoshiba K., Baba Y., Kurosaki T., Wang Y., Gill D.L. // Cell Calcium. 2014. V. 56. № 6. P. 482–492.
39. Zhou H., Iwasaki H., Nakamura T., Nakamura K., Maruyama T., Hamano S., Ozaki S., Mizutani A., Mikoshiba

K. // Biochem. Biophys. Res. Commun. 2007. V. 352. № 2. P. 277–282.
40. Hofer A., Kovacs G., Zappatini A., Leuenberger M., Hediger M.A., Lochner M. // Bioorg. Med. Chem. 2013. V. 21. № 11. P. 3202–3213.
41. Dellis O., Mercier P., Chomienne C. // BMC Pharmacol. 2011. V. 11. P. 1.
42. Schild A., Bhardwaj R., Wenger N., Tscherrig D., Kandasamy P., Dernič J., Baur R., Peinelt C., Hediger M.A., Lochner M. // Internat. J. Mol. Sci. 2020. V. 21. № 16. P. 1–28.
43. Wang X., Wang Y., Zhou Y., Hendron E., Mancarella S., Andrake M.D., Rothberg B.S., Soboloff J., Gill D.L. // Nat. Commun. 2014. V. 5. P. 3183.

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 *Acta Naturae* is on the list of the leading periodicals of the Higher Attestation Commission of the Russian Ministry of Education and Science. The journal Acta Naturae is indexed in PubMed, Web of Science, Scopus and RCSI databases.

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 50,000 characters with spaces (approximately 30 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 (approximately 15 pages in A4 format, including tables and references). They should contain no more than ten figures.

A short report must include the study's rationale, experimental material, and conclusions. A short report should not exceed 12,000 symbols (5–6 pages in A4 format including no more than 12 references). It should contain no more than three figures.

The manuscript and all necessary files should be uploaded to www.actanaturae.ru:

1) text in Word 2003 for Windows format;
2) the figures in TIFF format;
3) the text of the article and figures in one pdf file;
4) the article's title, the names and initials of the authors, the full name of the organizations, the abstract, keywords, abbreviations, figure captions, and Russian references should be translated to English;
5) the cover letter stating that the submitted manuscript has not been published elsewhere and is not under consideration for publication;
6) the license agreement (the agreement form can be downloaded from the website www.actanaturae.ru).

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 600 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 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.

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 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: actanaturae@gmail.com, tel.: (495) 727-38-60.