ActaNaturae

At the Interface of Three Nucleic Acids: The Role of RNA-Binding Proteins and Poly(ADP-ribose) in DNA Repair

Modulation of DNA repair protein activity

PARylation of YB-1. Propagation of DNA damage signal

Translation regulation

YB-1

3

YB-1

PARP1

6

YB-1

4

PARG

2

PARG inhibition

pre-mRNA

YB-1

YB-1

A

Stimulation of PARP1 activity

1

RNA polymerase II

YB-1 as a trans-cription factor

YB-1

Relocation to the nucleus

Genotoxic stress

5

YB-1

mRNA

YB-1

YB-1

Processing of mRNA

mRNA

YB-1 as a major component of RNA granules
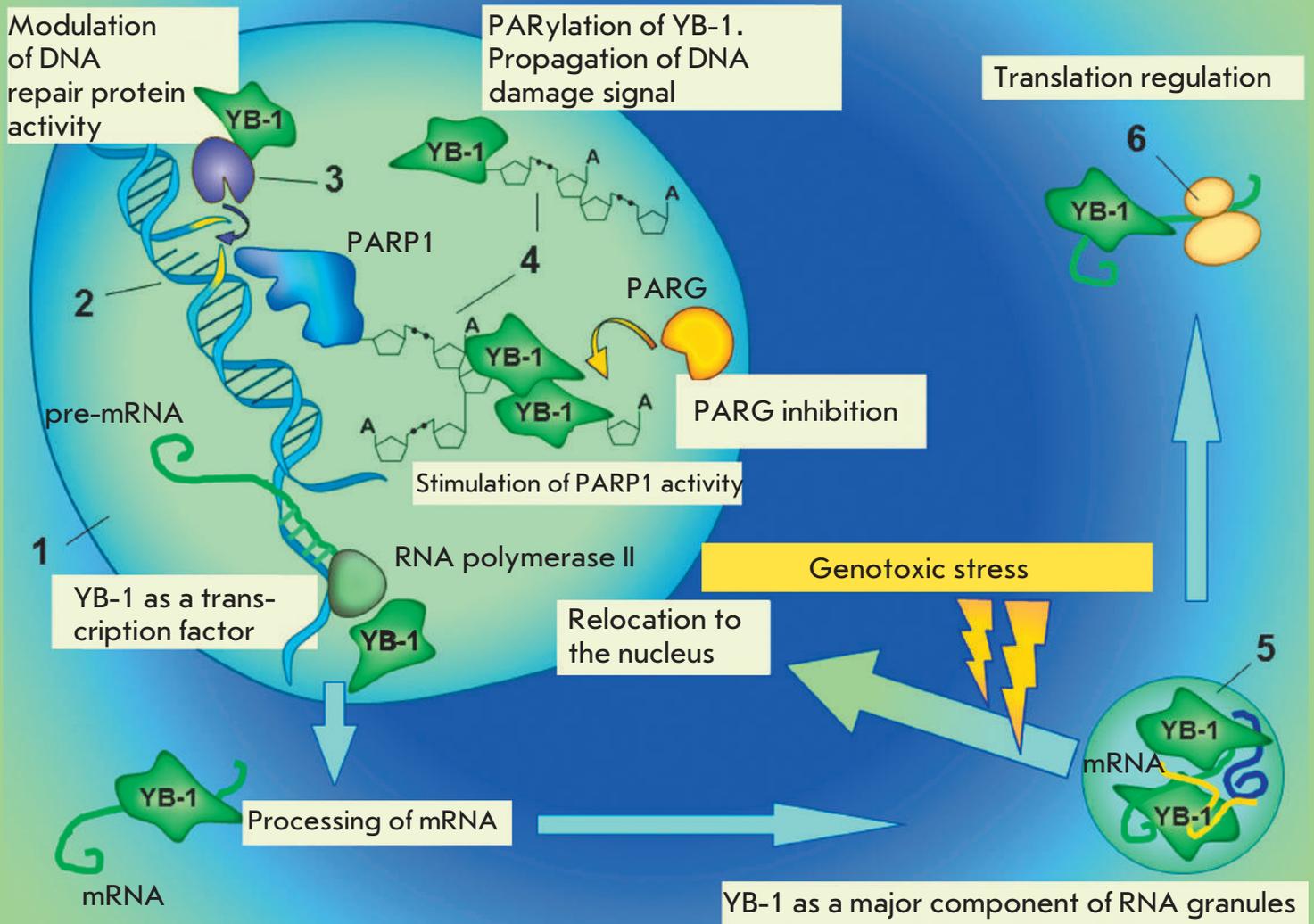
C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors

A. A. Fedotova, A. N. Bonchuk, V. A. Mogila, P. G. Georgiev

This review attempts to inventory the available data on the abundant but still poorly understood family of proteins with clusters of the C2H2 zinc finger domains. The distinctive features of C2H2 zinc finger proteins include strong and specific binding to a long and unique DNA recognition target sequence and rapid expansion within various animal taxa during evolution.
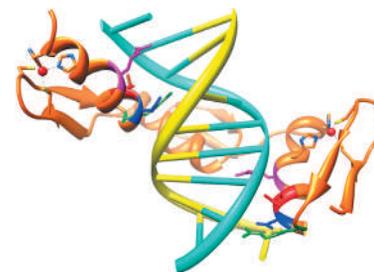
A model of the site-specific DNA recognition by C2H2 zinc finger domains
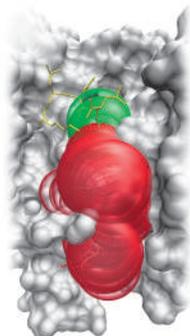
Structure Modeling of Human Tyrosyl-DNA Phosphodiesterase 1 and Screening for Its Inhibitors

I.V. Gushchina, D.K. Nilov, A.L. Zakharenko, O.I. Lavrik, V.K. Švedas

DNA repair enzyme tyrosyl-DNA phosphodiesterase 1 (Tdp1) represents a potential molecular target for anticancer therapy. A human Tdp1 model has been constructed using the methods of quantum and molecular mechanics, taking into account the ionization states of the amino acid residues in the active site and their interactions with the substrate and competitive inhibitors. The developed molecular model allowed us to uncover new Tdp1 inhibitors.

The substrate-binding groove in the human Tdp1 model

Comparing New-Generation Candidate Vaccines against Human Orthopoxvirus Infections

R. A. Maksyutov, S. N. Yakubitskyi, I. V. Kolosova, S. N. Shchelkunov

The lack of immunity to the variola virus in the population, increasingly more frequent cases of human orthopoxvirus infection, and increased risk of the use of the variola virus as a bioterrorism agent calls for the development of modern, safe vaccines against orthopoxvirus infections. The proposed immunization protocols can be used to develop safe vaccination strategies against smallpox and other human orthopoxvirus infections.

Time-course of mortality after double immunization of mice with the preparations under study (DNA vaccine, VACΔ6 and LIVP VACV strains), followed by challenge with ECTV
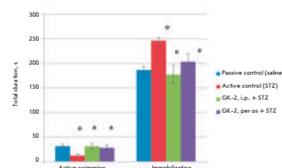
Low-Molecular-Weight NGF Mimetic Corrects the Cognitive Deficit and Depression-like Behavior in Experimental Diabetes

R. U. Ostrovskaya, S. S. Yagubova, T. A. Gudasheva, S. B. Seredenin

The effect of the mimetic of dimeric dipeptide NGF loop 4, GK-2, on a model of streptozotocin-induced type 2 diabetes in C57Bl/6 mice was studied. The study revealed the ability of GK-2 to ameliorate hyperglycemia induced by streptozotocine in C57Bl/6 mice, to restore learning ability in the Morris Water Maze test, and to overcome depression after both intraperitoneal and peroral long-term administration. The presence of the listed properties and their preservation in the case of peroral treatment determines the prospects of research.

Indicators of a depressive-like status in C57Bl/6 mice

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

CONTENTS

REVIEWS

RESEARCH ARTICLES

CONTENTS

IMAGE ON THE COVER PAGE
(see the article by Alemasova *et al.*)

At the Interface of Three Nucleic Acids: The Role of RNA-Binding Proteins and Poly(ADP-ribose) in DNA Repair

E. E. Alemasova[1], O. I. Lavrik[1,2]*
[1]Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, 630090, Russia
[2]Novosibirsk State University, Novosibirsk, 630090, Russia
*E-mail: lavrik@niboch.nsc.ru
Received: September 14, 2016; in final form January 19, 2017

ABSTRACT RNA-binding proteins (RBPs) regulate RNA metabolism, from synthesis to decay. When bound to RNA, RBPs act as guardians of the genome integrity at different levels, from DNA damage prevention to the post-transcriptional regulation of gene expression. Recently, RBPs have been shown to participate in DNA repair. This fact is of special interest as DNA repair pathways do not generally involve RNA. DNA damage in higher organisms triggers the formation of the RNA-like polymer − poly(ADP-ribose) (PAR). Nucleic acid-like properties allow PAR to recruit DNA- and RNA-binding proteins to the site of DNA damage. It is suggested that poly(ADP-ribose) and RBPs not only modulate the activities of DNA repair factors, but that they also play an important role in the formation of transient repairosome complexes in the nucleus. Cytoplasmic biomolecules are subjected to similar sorting during the formation of RNA assemblages by functionally related mRNAs and promiscuous RBPs. The Y-box-binding protein 1 (YB-1) is the major component of cytoplasmic RNA granules. Although YB-1 is a classic RNA-binding protein, it is now regarded as a non-canonical factor of DNA repair.
KEYWORDS DNA repair, intrinsically disordered proteins, poly(ADP-ribose), RNA-binding proteins, Y-box-binding protein 1.
ABBREVIATIONS BER − base excision repair; IDP − intrinsically disordered protein; IDPR − intrinsically disordered protein region; LCD − low complexity domain; PAR − poly(ADP-ribose); PTM − posttranslational modification; RBD − RNA-binding domain; RBP − RNA-binding protein; RNP − ribonucleoprotein.

INTRODUCTION

DNA, RNA and poly(ADP-ribose) (PAR) are the three essential cellular nucleic acids whose functions are tightly interlinked and effected by specific mediator proteins. Some of DNA-, RNA-, and PAR-binding proteins can also interact with other types of nucleic acids distinct from their classic targets. These proteins contain a broad range of disordered regions in their structure that can accommodate any ligand upon binding. In this review, we attempt to summarize recent research findings pertaining to the interactions between the three essential nucleic acids driven by multifunctional cellular proteins. As an example, Y-box-binding protein 1 (YB-1) is discussed.

INTERFERENCE OF DNA REPAIR AND TRANSCRIPTION

Base excision repair (BER) provides a clear picture of DNA repair and RNA metabolism coupling, since numerous molecules of this pathway, including APE1, SMUG1 and PARP1, are involved in RNA metabolism [1]. Obviously, transcription factors can mediate DNA repair by regulating the expression of repair enzymes [2]. However, the reverse is also possible: a few DNA repair enzymes may serve as transcriptional coactivators [3]. For example, thymine DNA glycosylase (TDG), which is involved in BER, is capable of activating gene transcription by recruiting coactivators [4]. The enzyme performs dynamic demethylation at promoters of silent and developmentally poised genes, as well as active gene enhancers for a rapid transcriptional response [5, 6].

DNA repair and transcription do not tend to occur simultaneously. At least, this is true for constitutively expressed housekeeping genes. Some bulky DNA damage stall RNA-polymerase II progression and trigger nucleotide excision repair (NER) (this subpathway of NER is called transcription-coupled NER (TC-NER) [7]. The mutagenic potential of other DNA lesions is minimized by inhibiting transcription at the site of a lesion; for instance, gene expression is down-

regulated during BER-assisted repair of oxidatively damaged DNA [8].

Signal-dependent and developmentally poised genes, on the contrary, require scheduled DNA damage to the promoter in order to trigger transcriptional activation [3]. An important regulatory mechanism for the expression of such genes is the promoter-proximal pausing of RNA polymerase II [9]. Transcription is activated, while elongation is suppressed at early time-points [10]. The escape of paused RNA polymerase II into productive elongation is mediated by DNA repair enzymes and chromatin remodeling factors. For example, the estrogen receptor activates lysine-specific histone demethylase 1 (LSD1), which demethylates histone H3. The oxidation process is accompanied by the release of a hydrogen peroxide byproduct, which converts adjacent guanines to 8-oxoguanine (8-oxoG) [11]. The repair of 8-oxoG by DNA glycosylases induces single-strand breaks that serve as entry points to DNA endonucleases, including topoisomerase IIβ [12]. When long genes are expressed, TopoIIβ creates DNA breaks not only in the promoters, but also in the reading frames, thus maintaining transcription elongation [13]. Recent findings have demonstrated that inhibition of topoisomerases suppresses the expression of long genes in yeasts [14, 15]. There is a view that the ensuing double-stranded DNA breaks relax DNA and recruit DNA damage response proteins and repair enzymes, such as PARP1 and DNA protein kinases, which leads to licensing of chromatin for transcription [12]. In human cells, DNA breaks and respective DNA repair signals are involved in the release of paused Pol II into productive synthesis and elongation of the genes that are activated following exposure to external stressors [16]. Poly(ADP-ribose) (PAR) polymerase 1 (PARP1) has been identified among the chromatin remodeling factors that control Pol II pausing. PARP1 is believed to play a role in transcription elongation due to PAR-coupled nucleosome disassembly [17]. However, poly(ADP-ribosyl)ation induced by DNA damage in the proximity of gene promoters also seems to attract the RNA-binding proteins important for Pol II docking.

Interestingly, RNA transcripts arising from a DNA lesion may trigger repair activation. It has been shown that spontaneous double-stranded DNA breaks induce ectopic transcription to give rise to short non-coding RNAs (DSB-induced small RNAs, diRNAs) 21 nucleotides long [18]. Francia *et al.* showed that diRNAs recruit enzymes to repair double-stranded breaks at the site of origin [18]. Talhaoui *et al.* have recently discovered a role for PARP1 and PARP2 in poly(ADP-ribosyl)ation of DNA strand break termini [19]. It is possible that this mechanism can contribute both to chromatin remodeling and DNA repair [19].

Some transcription factors have been shown to directly participate in DNA repair [20]. These transcription factors are thought to trigger local chromatin remodeling, thus activating DNA repair in the target sequences [21].

Collectively, transcription factors provide an extra layer of protection to the genome. Every tissue undergoes DNA damage from different sources: very high rates of oxygen metabolism in neurons lead to elevated levels of oxidative DNA lesions, whereas skin cells cope with increased UV-induced DNA damage [20]. Since transcription factors are regulated by extracellular signals and stress-activated pathways, they can confer protection to cells of a certain type [20]. Due to heterogeneous DNA repair along the genome (there is a gradient of DNA repair, with the rate decreasing towards the 3'-end of the gene), transcription factors ensure the genomic stability of the key promoter and enhancer regions of the genes being transcriptionally regulated [22].

EUKARYOTIC "RNA OPERONS"

F. Jacob and J. Monod were the first to propose the term "operon" in 1961. According to the theory, a cluster of genes is located sequentially within an operon. The genes in the operon are together transcribed into one polycistronic mRNA, which is further translated to yield the final components of a functional complex in close proximity to each other to ensure rapid assembly. Later studies into the ribosomal profile of Escherichia coli gene expression supported this theory and demonstrated that proteins are synthesized precisely to meet the stoichiometry of the multiprotein complex [23].

DNA operons are rare in the genome of eukaryotes, and mRNAs are mainly monocistronic. The loss of DNA operons in higher organisms could be attributed to the polar effect of nonsense mutations and the complicated regulatory network of synthesis of multifunctional proteins, which are abundant in the eukaryotic cell [24]. For this reason, the eukaryotic expression is partly regulated at the post-transcriptional level, with mRNAs that encode functionally related proteins assembling into RNA operons (*Fig. 1*), thus acquiring a common fate [25]. The principal structural and functional unit of this process is the numerous RNA-binding proteins (RBPs) that bind to RNA motifs to form ribonucleoprotein (RNP) complexes [26]. RNP complexes structurally represent the RNA operon, which allows functionally related proteins arising from different mRNAs to be jointly translated at a single cytoplasmic location [27]. The potential of RNP complexes to act dynamically and independently of the cellular environment is attributed to the mechanism called liquid demixing [28−37] that is triggered by intrinsically disordered RNA-binding proteins.
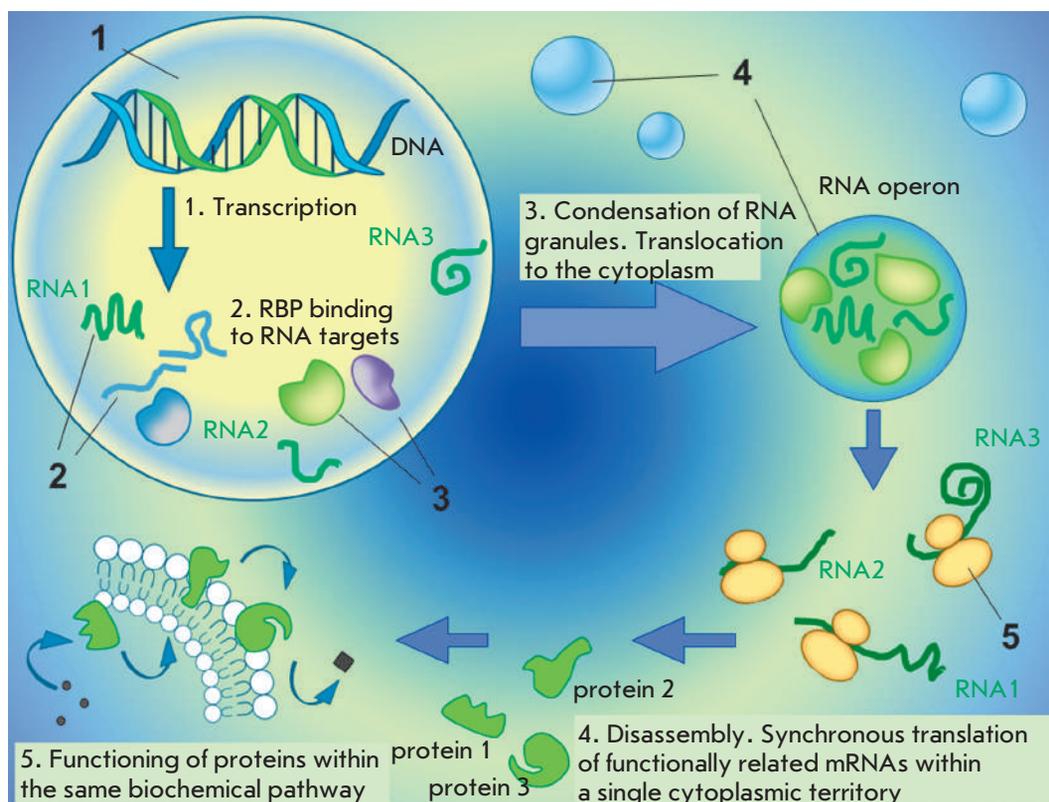
Fig. 1. Eukaryotic RNA operons (schematic representation) 1 – nucleus; 2 – pre-mRNA; 3 – RNA-binding proteins; 4 – cytoplasmic RNA granules (RNA operons); 5 – ribosome. The figure schematically shows the formation and functioning of cytoplasmic RNA assemblages. These complexes of functionally related mRNAs and RBPs act as RNA operons that facilitate the synchronous translation of proteins involved in the same biochemical pathway

THE NEW FUNCTIONS OF RNA-BINDING PROTEINS IN RESPONSE TO DNA DAMAGE

The recent progress achieved in research has highlighted the role that RNA-binding proteins play as guardians of genome stability [38].

DNA damage induces down-regulation of gene expression at different levels. The first step involves the suppression of transcription and pre-mRNA 3'-end processing [39, 40]. The biosynthesis of functional proteins decreases following a switch in alternative splicing from in-frame variants to variants prone to nonsense-mediated decay [41, 42]. Finally, DNA lesions affect the stability of many mRNAs [43] and inhibit translation [44, 45].

However, although the overall expression levels drop, the DNA repair machinery possesses specific mechanisms that allow it to enable the synthesis of the proteins engaged in the repair process. Suppressed translation may not affect the mRNAs that encode repair enzymes [46]. According to the model of RNA operon, mRNAs coding for functionally related proteins are together regulated at the post-transcriptional level. Overall, a single RBP such as HuR can control the expression of a broad range of genes that are involved in DNA repair [47–49].

RNA-binding proteins mediate transcription and chromatin remodeling, and they can directly participate in DNA repair [50, 51]. RBPs migrate to the sites of DNA damage [52–54], which can be explained by their ability to bind to the short non-coding mRNAs (ncRNA) that are formed at the site of a break [18, 50, 55], or by an RNA-independent mechanism.

Gene transcription at a high biosynthesis rate or with long transcripts sometimes continues into the S-phase [56], with a possibility for RNA-DNA-hybrids (R-loops), which impact the transcription and threaten genome integrity [57]. R-loop formation is prevented mainly due to RNA-binding protein-coupled packing of pre-mRNA during synthesis [58, 59].

Post-translational modification (PTM) of proteins is crucial to a cellular response to DNA damage. RBP is a primary set of proteins that are phosphorylated [60, 61] and poly(ADP-ribosyl)ated [62] under the control of DNA damage. Genotoxic stressors also trigger an increase in the levels of acetylation of certain RNA-binding proteins [63].

Finally, DNA damage facilitates the bidirectional relocation of RNA-binding proteins between the nucleus and the cytoplasm [64, 65], thus contributing to the coordinated regulation of RNA metabolism and DNA repair by multifunctional RBPs.

RNA-BINDING PROTEINS: MODULE ORGANIZATION

The bulk of cellular mRNA is associated with RNA-binding proteins in the form of RNP complexes.

Disruption of RNA granule formation results in various disorders [66, 67]. Interaction with RBPs is required for the regulation of RNA metabolism at different levels, from biosynthesis to decay. RNA-binding proteins fulfill key functions in such processes as pre-mRNA splicing [68], polyadenylation [69], transport to the cytoplasm, and translation. RBPs also have a role in the processing of non-coding RNA: the so called microRNA (miR), circular RNA (circRNA), and long non-coding RNA (lncRNA) [70−72]. Over all, RNA-binding proteins constitute an important class of post-transcriptional gene regulators.

There are a total of 1,500 RBPs known to date [73, 74]. Many RNA-binding proteins have a modular structure, in which a few basic RNA-binding domains (RBD) are arranged to accommodate a broad range of RNA sequences [75]. Certain RBDs tend to bind short sequences and display poor affinity for RNA; however, the interaction interface formed by multiple modules ensures a high affinity and specificity towards an RNA target. The superposition of weak interactions facilitates the regulation of assembly and disassembly of RNP complexes that may be mediated by an RNA-like polymer of poly(ADP-ribose) [76, 77]. Owing to the modular structure of RNA-binding proteins, different RNAs may be targeted by the same RBP [75]. A beautiful example of specific target binding promoted by tandem RBDs is the proteins of the Pumilio family (Puf), in which three amino acid side chains of each of the protein's eight domains establish contacts with a different RNA base [78]. This "RNA recognition code" could be utilized to produce proteins with the desired binding specificity [79]. RBDs, for example, RNA-binding motif (RRM), in certain cases may also serve for protein-protein interaction [80].

It has been recently shown that besides regular RBDs, an essential role in RNA recognition is played by intrinsically disordered protein regions (IDPRs), which are highly enriched in RNA-binding proteins as compared to the total human proteome [81]. A total of 20% of mammalian proteins identified as RBPs are intrinsically disordered by over 80% [82]. Like regular RBDs, the regions with disordered sequences in RNA-binding proteins are arranged into modules that are repeated nonrandomly within a single amino acid sequence and, in some cases, may combine with globular domains [82]. Importantly, the emergence of disordered proteins in RBPs correlates with the complexity of the transcriptome in eukaryotes during evolution [83].

DANCING PROTEINS, CHAMELEON PROTEINS, 4D AND PROTEIN CLOUDS

The new terms [84−87] coined to describe proteins without a stable 3D structure reflect the global flexibility and dynamic landscapes of intrinsically disordered proteins (IDPs) or protein regions (intrinsically disordered protein regions, IDPRs) [88]. Since the 3D protein structure is maintained by non-covalent atomic forces such as hydrogen bonding, hydrophobic interactions, van der Waals forces, etc., the intrinsic disorder, as well as the unique structure of globular proteins, is encoded by the amino acid sequence. The combination of a high net charge and low mean hydrophobicity drives the emergence of a natively unfolded protein conformation under physiological conditions [89]. The amino acid sequence of IDPs and IDPRs is enriched in Pro, Arg, Gly, Gln, Ser, Glu, Lys, and Ala but depleted in Cys, Trp, Tyr, Phe, Ile, Leu, Val, and Asn [90].

Intrinsically disordered proteins partially adopt a certain 3D structure following a change in the environment or upon binding to a ligand [91]. Their folding may also be facilitated by an elevated temperature boosting hydrophobic interactions [92], pH changes decreasing the net charge [92], as well as the presence of ions neutralizing electrostatic repulsion between clusters of amino acid residues of the same charge [93, 94]. Inside the cell, intrinsically disordered proteins adopt a rigid secondary structure after binding to ligands: small molecules, cofactors, proteins, nucleic acids, membranes, etc. [91, 95].

The functions of most proteins, in particular IDPs, are modulated through post-transcriptional modifications (PTM). As many as 300 PTMs have been identified to occur in the cell [96]. Although DNA only encodes 20 amino acids, the diversity of amino acid residues in proteins exceeds 140, owing to PTMs [97]. Proteins are mainly targeted in the disordered regions [98, 99].

IDPs and IDPR-containing proteins seem to play a central role in interactomes [100]. About 30−40% of eukaryotic proteins carry lengthy IDPRs [101], with intrinsically disordered proteins carrying out the key functions in transcription and intracellular signaling cascades [102]. In 2005, it was first suggested that hub proteins (containing multiple protein-protein interaction links within interactomes) might be enriched in IDPR [103]. Extensive studies allowed researchers to differentiate hub proteins into static and dynamic hubs [104, 105]: the former clustering into modules, which represent functional complexes with a high degree of interplay between the components (such as the transcription initiation machine), while the latter ensure interconnection of the modules [106]. IDPRs proved to be significantly enriched in dynamic hubs [107], hence elucidating the role of intrinsic disorder in guiding cellular processes [100].

IDPRs have plenty of functions. They are responsible for the autoinhibition of enzymes. In this regard, disorder-to-order transition acts as a switch on-switch
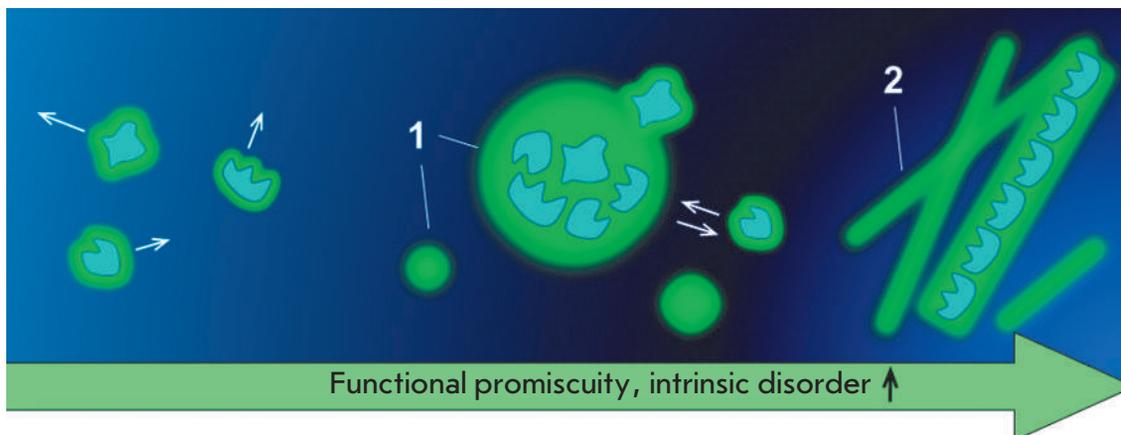
Fig. 2. Phase transitions of biomolecules 1 – functional membraneless organelles; 2 – pathological amyloid aggregates of proteins. Cellular biomolecules undergo phase transitions as water does. In the gaseous state, biomolecules are dispersed throughout the cell and do not interact with each other. A local increase in the concentration of promiscuous and intrinsically disordered proteins results in intracellular liquid demixing and induces the assembly of membraneless compartments that have liquid-like properties [30–32]. The liquid-like state is maintained by multiple weak interactions among the interaction partners. An irreversible transition into a condensed liquid state appears to lead to amyloid fibers that are associated with such disorders as Alzheimer's disease [35]

off mechanism for the target protein [108]. This mechanism is employed for the activation of PARP1 during DNA repair, resulting in DNA damage signaling [109]. Another interesting example is the role of IDPR-containing proteins in protein quality control, with chaperone disorder-to-order transition being stress-induced [110]. There is data suggesting that IDPs act as molecular shields that prevent the aggregation of intrinsically disordered proteins by steric interference under stress conditions [111]. IDPRs can also regulate tissue-specific protein interactions at the transcriptional level. Buljan *et al.* [112] and Ellis *et al.* [113] showed that the enrichment of IDPRs in proteins is due to tissue-specific spliced exons [112]. Similarly, tissue-specific exons contribute to the majority of the disordered regions targeted by PTMs and motifs binding partner molecules [112]. The proteins translated from mRNAs enriched in tissue-specific exons occupy central positions in protein interaction networks and have different interaction partners in these tissues [112].

The presence of conserved IDPRs in the structure of mammalian early DNA base excision repair enzymes is a unique feature that their homologues in lower organisms do not have [114]. The IDPRs of repair enzymes are involved in DNA damage recognition, binding to interaction partners; they provide key sites for the PTMs that modulate stability, enzyme-, and DNA-binding activity, the intracellular localization of repair proteins; and they provide higher organisms with an advantage over the protein size, reducing intracellular crowding [115–119].

Finally, IDPs and IDPRs play a crucial role in the formation of dynamic macromolecular assemblages inside the cell, including RNP granules and DNA repair complexes.

PHASE TRANSITIONS OF BIOMOLECULES

According to the recent findings reported in [29, 30, 33–35, 37], biochemical processes inside the cell are separated by phase transitions of biomolecules (*Fig. 2*). This paradigm states that the formation of membraneless compartments is similar to that of dispersed droplets upon emulsion breakdown (so called *liquid demixing*) [28–30, 120–122]. Intrinsically disordered proteins play a key role in phase transition events [31]. The structural plasticity and conformational flexibility of IDPs allow them to interact with multiple, structurally unrelated partners [32]. Many IDPs contain low-complexity domains (LCDs) that are prone to multimerization, driven by favorable changes in potential energy [33]. Liquid demixing results in the separation of proteins and their ligands within a compartment with a microenvironment distinct from that of other cellular plasm, thus increasing the local concentrations of interacting molecules and promoting biochemical processes [34].

The formation of RNP complexes is one of the important representations of membraneless compartmentalization by means of phase transitions of mRNA and corresponding IDPR-containing RNA-binding proteins [27]. The RNAs present in these complexes maintain their solubility [35, 36], which seems to facilitate downstream translation [27]. However, phase transitions
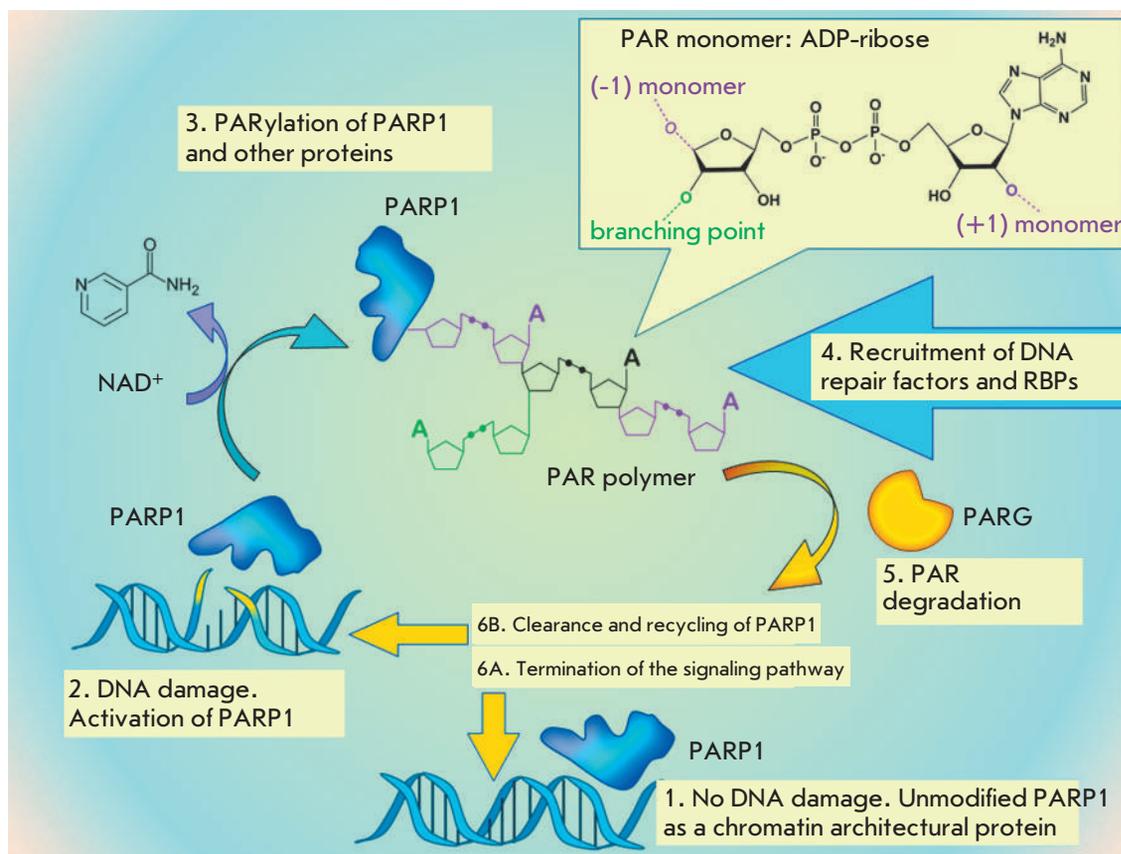
could occur independently of RNAs only in the presence of proteins, such as in the case of formation of centrosomes (microtubule nucleation sites) [123]. Altmeyer *et al*. reported that the assembly of multiprotein repair complexes at the sites of DNA damage is achieved through liquid demixing. It was also suggested that the formation of a non-membranous DNA repair compartment also has a role in the bridging of DNA ends and their protection from nucleases [124, 125].

Phase transitions of proteins and nucleic acids to give rise to dynamic ensembles is initiated by an increase in the concentration of components, followed by self-aggregation [126], or could occur in response to changes in the microenvironment, such as pH, ionic strength, or temperature [127]. In addition, certain biomolecules are able to act as nucleation centers of multiprotein complexes, followed by separation of the intracellular plasma into two liquid phases with varying properties [37].

Single-stranded RNA [27, 128] and DNA (ssDNA) [129] are the preferred options for the nucleation of phase transition. Both biomolecules display significantly more plasticity as compared to double-stranded DNA and share such properties as a negative charge and relatively low complexity due to a limited presence of unique building units. All these features are indica-

tive of intrinsic disorder [33]. Higher organisms reached the peak of intracellular plasma self-organization upon acquisition of a "third nucleic acid", poly(ADP-ribose), a polymer with no ability to store information, an extremely simple structure consisting of ADP-ribose units, and a short lifetime. It is possible that poly(ADP-ribose) is the key agent in the regulation of phase transitions in the cell.

POLY(ADP-RIBOSE) AND POLY(ADP-RIBOSYL)ATION

Poly(ADP-ribose) is a linear or branched polymer chain consisting of identical molecular units: monomers of ADP-ribose produced from NAD^+ via PARP1-catalyzed PAR synthesis (*Fig. 3*) [130]. Under physiological conditions, PAR has a dynamic multiglobular structure depending on the polymer size, which allows the polymer to fit the structure of the bound ligand [131]. Adenine residues in PAR, identically to those in nucleic acids, adopt an anti-conformation that is capable of base stacking and formation of hydrogen bonds [132]. The secondary structure of PAR as a helix, which has been confirmed *in vitro* by spectral analysis [133], can occur at high ionic strength (4 M NaCl) or upon binding to proteins under physiological conditions [132]. The PAR polymer carries two negatively charged

phosphates in each monomer (ADP-ribose unit), while RNA and ssDNA only carry one negative charge per unit [134]. In the absence of genotoxic stressors, intracellular PAR levels are very low and ADP-ribose exists in a relatively stable state of monomers and oligomers (half-life $t_{1/2}$ ~7.7 h). Extensive local biosynthesis of a very short-lived PAR polymer ($t_{1/2}$ less than 1 min) is triggered by DNA damage [135−137]. The prominent feature of poly(ADP-ribose) is its involvement in post-translational protein modifications.

By analogy with DNA and RNA, the enzymes that catalyze the synthesis of PAR are called PAR polymerases (PARPs). The human PARP family includes 17 members with similar catalytic domains [138]. Only four members are capable of catalyzing PAR synthesis: PARP1, PARP2, and two tankirases [138, 139]. PARP1 and PARP2 act as guardians of genome integrity [140]. Tankirases synthesize linear PAR chains up to 20 monomers long [141]. Their functions are exerted when the spindle apparatus begins to form [142]. Tankirases also control centrosome functions [143].

PARP1 is activated upon binding to exposed bases on the loose ends of DNA breaks [144]. Recognition of a DNA lesion induces conformational changes in the autoinhibitory domain of PARP1, which locally unfolds, thus ceasing to interfere with NAD^+ binding in the active center [109]. As a result of intermolecular rearrangement of PARP1 attracting the catalytic domain to the damage site, the automodification domain is positioned close to the active center open to modification by PAR [145]. This finding provides insight into why PARP is the preferred target for poly(ADP-ribosyl)ation [134]. The PAR acceptor amino acid residues identified in PARP1 and other poly(ADP-ribosyl)ation targets to date are multivarious: Lys, Arg, Glu, Asp, Cys, Ser, Thr, Sep (through the phosphate group) and Asn, although charged amino acid residues are typically responsible for this function [146−149]. Bearing in mind that the rate of PAR biosynthesis is limited to NAD^+ breakdown, it is tempting to suggest that the binding of ADP-ribose to a target protein in the presence of activated PARP1 occurs via any amino acid residue exposed on the protein surface [125]. Specific PAR-mediated modulation of cellular processes can be achieved through different local microenvironments of PARP1 and its ligand, rather than through specific PAR acceptor sites in the target protein [125].

PAR binds non-covalently to many proteins. Among the proteins associated with PAR and/or prone to this PTM are certain repair enzymes, chromatin remodeling proteins, RNA-binding proteins, and transcription factors [62, 150]. Numerous functions exerted by PAR in the cell are implemented via dynamic interactions between poly(ADP-ribose) and PAR-binding proteins.

Protein relocation caused by local synthesis of PAR influences cellular signaling, DNA damage response, transcription regulation, protein stability, and cell fate [151]. Several PAR-binding modules have been described; their structure varies from completely ordered domains to intrinsically disordered regions capable of forming multivalent contacts with the PAR polymer [125].

PAR can also be recognized by RNA- and DNA- binding motifs [125]. Since not only specific interactions but also dynamic changes in the concentrations of interacting molecules influence macromolecular ensembles, PAR may outcompete RNA binding of RBPs at the peaks of PARylation, resulting in RBPs relocalization to DNA damage sites [152]. The DNA-binding domains of DNA repair enzymes and transcription factors may also facilitate the recruitment of these proteins to the DNA damage sites in a PAR-dependent mechanism [153, 154].

It has been recently shown that PAR can nucleate the intracellular phase transitions of such RNA-binding proteins as FUS (TLS), EWS (EWSR1), and TAF15 at microlaser-generated sites of DNA lesions [124]. Intracellular compartmentalization initiated by PAR-dependent phase separation can underlie the mechanisms by which poly(ADP-ribose) is involved in DNA- and RNA-dependent cellular events: for example, the formation of stress-granules [155], nucleoli [156], spliceosomes [157], and transcriptosomes [158]. In the event of transcription regulation, the phase transition of FUS (TLS), EWS (EWSR1), and TAF15 at gene promoters appears to create sites for the binding of the C-terminal disordered domain of RNA-polymerase II [159]. PARylation in close proximity to promoters seems to facilitate transcription, especially if keeping in mind that DNA breaks in promoters and reading frames may be scheduled [5, 13, 17].

Long-lived PAR carries such risks as stripping RNA- and DNA binding proteins off their ligands, phase transitions of dynamic droplets into the insoluble protein aggregates found in pathological states [33], as well as the energy crisis arising from depleted NAD^+ pools [160]. That is why PARylation is subjected to tight control by the enzymes that break down PAR and remove ADP-ribose residues from modified proteins [161]. The key ADP-ribose-degrading enzyme is poly(ADP-ribose)glycohydrolase (PARG), which exhibits endo- and exo-hydrolase activities; the latter activity being dominant over the first one [162]. Since degradation occurs when the polymer is available, PAR-binding proteins can potentially counteract PARG. PARG is actually unable to cleave the proximal ADP-ribose monomer, which appears to be due to steric hindrance [163]. ADP-ribose units are removed from mono(ADP-ribosyl)ated proteins by specific enzymes [164]. Dynamic regulation of PAR levels may provide a physiological balance be-
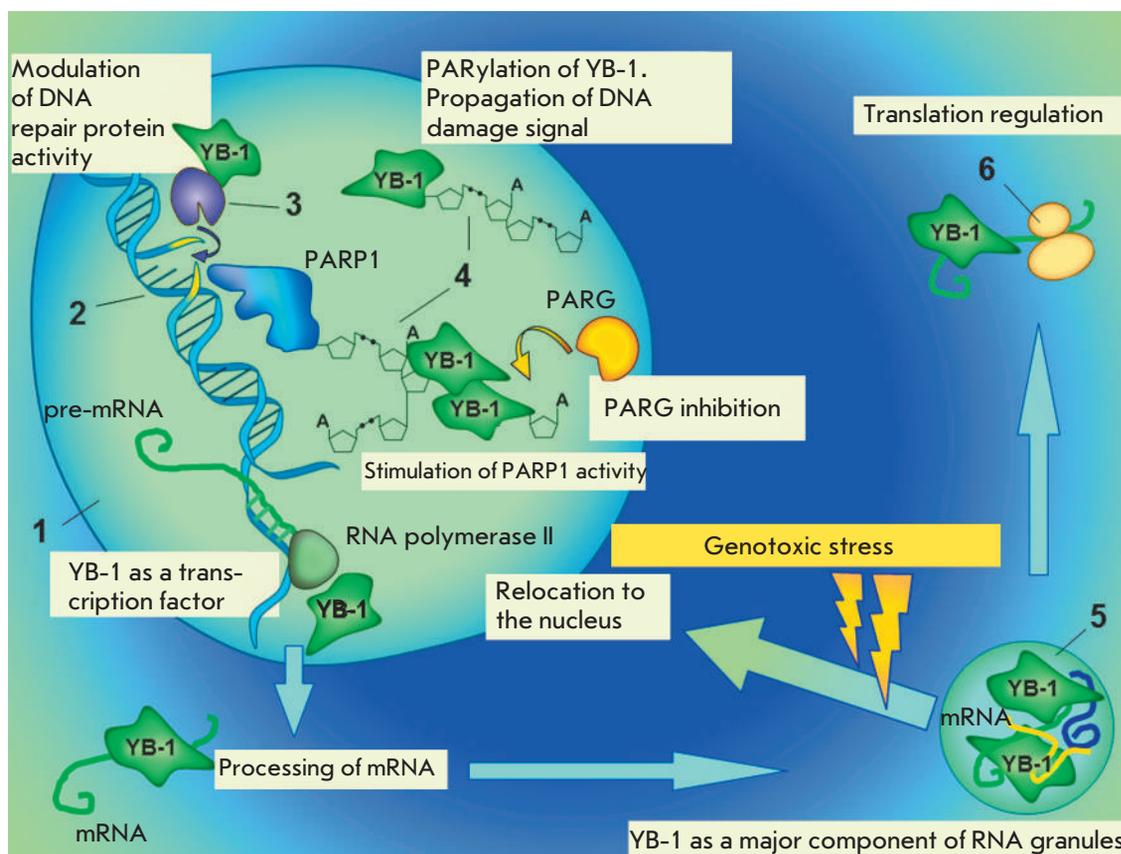
Fig. 4. Switching of YB-1 cellular functions upon genotoxic stress (schematic representation) 1 – nucleus; 2 – DNA damage; 3 – DNA repair enzyme; 4 – poly(ADP-ribose); 5 – cytoplasmic RNA granule; 6 – ribosome

tween DNA- and RNA-protein interactions in different cellular contexts.

Y-BOX-BINDING PROTEIN 1

The Y-box-binding protein 1 (YB-1) is an example of a multifunctional protein acting at the "interface of three nucleic acids." While binding to DNA [165, 166], YB-1 carries out its functions in transcription [167] and likely in DNA repair [166, 168]. YB-1, as a transcription factor, controls the expression of stress-induced genes and the genes involved in DNA repair [167, 169, 170]. As an RNA-binding protein [167, 171], YB-1 mediates pre-mRNA splicing, is one of the major proteins constituting RNP granules in the cytoplasm [172], and modulates mRNA translation [167, 173]. There is evidence that YB-1 interacts with multiple noncoding RNAs [174, 175] and exhibits strong affinity for damaged DNA and RNA [166, 168, 176], as well as PAR-binding properties [150]. Genotoxic stress induces a relocation of YB-1 from the cytoplasm to the nucleus [177−180]. Under certain conditions, this stress-induced trafficking occurs following a specific post-translational modification of YB-1 − partial proteolytic cleavage by the 20S proteasome [181].

The bulk of the YB-1 structure is natively unfolded [167], which facilitates interaction promiscuity and confers the ability to self-aggregate, allowing for multimerization in the presence of RNA and DNA [182] or the formation of amyloid fibrils at a high ionic strength [183]. YB-1 binds to a wide range of DNA repair enzymes: base excision repair enzymes (NEIL2 [177], APE1 [184], DNA polymerase β [177], DNA polymerase δ [185], PCNA [186], DNA-ligase IIIα [177], NEIL1, PARP1, and PARP2 [187]), mismatch repair enzymes (MSH2 [185]), and DNA double-stranded breaks repair enzymes (Ku80 [185]). YB-1 is required for the recognition of bulky lesions by NER factor XPC-HR23b [188] and modulates the activity of key and regulatory BER enzymes [177, 187, 189−191].

YB-1 is found in stress granules [192], is necessary in centrosome formation [193], and has a potential role in nucleolar disassembly [194]. The emergence of these membraneless compartments, as well as the formation of repair complexes at sites of DNA lesions, is orchestrated by poly(ADP-ribose) [155, 156, 195]. Recent findings have demonstrated that YB-1 is able to modulate PAR biosynthesis depending on the level of DNA damage [187] and acts as a target for poly(ADP-ribosyl)ation [187, 196]. Another feature is the fact that YB-1 protects PAR from cleavage by PARG, extending the half-life of the polymer [187]. *Figure 4* schematically depicts the role played by YB-1 in PAR and RNA metabolism.

Over all, a transcription factor and one of the key RNA-binding cytoplasmic proteins, YB-1 display a plethora of additional functions that come into play under genotoxic conditions. Besides transcriptional and post-transcriptional regulation of gene expression, the functions of YB-1 may include participation in DNA repair and regulation of repair complex formation through PAR-dependent phase transitions of intrinsically disordered proteins and DNA repair factors enriched in IDPRs. YB-1 represents a possible pathway in which RBP may act as an extra guardian of genome integrity under stress conditions.

CONCLUSIONS

It appears that, the higher the level of an organism, the higher is the organizational complexity of its regulatory pathways. At the same time, the limited size of the cell prompts proteins to assume a multifunctional role. The multifunctionality, i.e., the ability to assume different functions, is closely linked to the ability to have many interaction partners whose structure in most cases is determined by the function performed by a protein in the cell. A modular structure that provides a variable degree of specificity cannot solve this problem, because the number of possible interactions remains limited. This limitation is beautifully addressed by reducing the information volume of the primary structure of nucleic acids and proteins. V. Uversky [88, 197] conclusively demonstrated that a reduced protein sequence leads to the maximum possible structural complexity. The occurrence of natively unfolded proteins dramatically expanded the range of intracellular interactions due to the unique features of this protein kingdom [197]. The intrinsic multivalence and their small size render these proteins instrumental in a variety of cellular processes and make them central players in interactomes, thus acting as key regulators of protein networking.

Along with the emergence of new functions in the proteome during evolution, higher eukaryotes have developed a wide array of noncoding nucleic acids that regulate basic RNA- and DNA-protein interactions. The maintenance of genome integrity, particularly, depends on the "third nucleic acid," poly(ADP-ribose), generated from NAD^+ in the presence of DNA damage. PAR formation, which modulates the interactions between RNA- and DNA-binding proteins and their targets, leads to the assemblage of functional complexes. These functional assemblages are required to regulate the key processes that take place in cellular metabolism under stress conditions. ●

REFERENCES

1. Vohhodina J., Harkin D.P., Savage K.I. // Wiley Interdiscip. Rev. RNA. 2016. V. 7. № 5. P. 604−619.
2. Goodwin J.F., Schiewer M.J., Dean J.L., Schrecengost R.S., de Leeuw R., Han S., Ma T., Den R.B., Dicker A.P., Feng F.Y., Knudsen K.E. // Cancer Discov. 2013. V. 3. № 11. P. 1254−1271.
3. Fong Y.W., Cattoglio C., Tjian R. // Mol. Cell. 2013. V. 52. № 3. P. 291−302.
4. Chen D., Lucey M.J., Phoenix F., Lopez-Garcia J., Hart S.M., Losson R., Buluwela L., Coombes R.C., Chambon P., Schär P., Ali S. // J. Biol. Chem. 2003. V. 278. № 40. P. 38586−38592.
5. Shen L., Wu H., Diep D., Yamaguchi S., D'Alessio A.C., Fung H.L., Zhang K., Zhang Y. // Cell. 2013. V. 153. № 3. P. 692−706.
6. Cortellino S., Xu J., Sannai M., Moore R., Caretti E., Cigliano A., Le Coz M., Devarajan K., Wessels A., Soprano D., *et al.* // Cell. 2011. V. 146. № 1. P. 67−79.
7. Mellon I., Hanawalt P.C. // Nature. 1989. V. 342. № 6245. P. 95−98.
8. Khobta A., Epe B. // Mutat. Res. 2012. V. 736. № 1−2. P. 5−14.
9. Adelman K., Lis J.T. // Nat. Rev. Genet. 2012. V. 13. № 10. P. 720−731.
10. Jonkers I., Lis J.T. // Nat. Rev. Mol. Cell. Biol. 2015. V. 16. № 3. P. 167−177.
11. Perillo B., Ombra M.N., Bertoni A., Cuozzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., Avvedimento E.V. // Science. 2006. V. 319. № 5860. P. 202−206.
12. 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.
13. Bunch H., Lawney B.P., Lin Y.F., Asaithamby A., Murshid A., Wang Y.E., Chen B.P., Calderwood S.K. // Nat. Commun. 2015. V. 6. P. 10191. doi: 10.1038/ncomms10191
14. Joshi R.S., Piña B., Roca J. // Nucleic Acids Res. 2012. V. 40. № 16. P. 7907−7915.
15. 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. e1003128.
16. Bunch H. // FEBS Lett. 2016. V. 590. № 8. P. 1064−1075.
17. Petesch S.J., Lis J.T. // Trends Genet. 2012. V. 28. № 6. P. 285−294.
18. Francia S., Michelini F., Saxena A., Tang D., de Hoon M., Anelli V., Mione M., Carninci P., d'Adda di Fagagna F. // Nature. 2012. V. 488. № 7410. P. 231−235.
19. Talhaoui I., Lebedeva N.A., Zarkovic G., Saint-Pierre C., Kutuzov M.M., Sukhanova M.V., Matkarimov B.T., Gasparutto D., Saparbaev M.K., Lavrik O.I., *et al.* // Nucleic Acids Res. 2016. V. 44. № 19. P. 9279−9295.
20. Malewicz M., Perlmann T. // Exp. Cell Res. 2014. V. 329. № 1. P. 94−100.

21. Frit P., Kwon K., Coin F., Auriol J., Dubaele S., Salles B., Egly J.M. // Mol. Cell. 2002. V. 10. № 6. P. 1391–1401.

22. Tu Y., Tornaletti S., Pfeifer G.P. // EMBO J. 1996. V. 15. № 3. P. 675–683.

23. Li G.W., Burkhardt D., Gross C., Weissman J.S. // Cell. 2014. V. 157. № 3. P. 624–635.

24. Keene J.D., Tenenbaum S.A. // Mol. Cell. 2002. V. 9. № 6. P. 1161–1167.

25. Keene J.D. // Nat. Rev. Genet. 2007. V. 8. № 7. P. 533–543.

26. Mitchell S.F., Parker R. // Mol. Cell. 2014. V. 54. № 4. P. 547–558.

27. Nielsen F.C., Hansen H.T., Christiansen J. // Bioessays. 2016. V. 38. № 7. P. 674–681.

28. Kato M., Han T.W., Xie S., Shi K., Du X., Wu L.C., Mirzaei H., Goldsmith E.J., Longgood J., Pei J., et al. // Cell. 2012. V. 149. № 4. P. 753–767.

29. Hyman A.A., Simons K. // Science. 2012. V. 337. № 6098. P. 1047–1049.

30. Weber S.C., Brangwynne C.P. // Cell. 2012. V. 149. № 6. P. 1188–1191.

31. Uversky V.N., Kuznetsova I.M., Turoverov K.K., Zaslavsky B. // FEBS Lett. 2015. V. 589. № 1. P. 15–22.

32. Li P., Banjade S., Cheng H.C., Kim S., Chen B., Guo L., Llaguno M., Hollingsworth J.V., King D.S., Banani S.F., et al. // Nature. 2012. V. 483. № 7389. P. 336–340.

33. Aguzzi A., Altmeyer M. // Trends Cell Biol. 2016. V. 26. № 7. P. 547–558.

34. Brangwynne C.P. // J. Cell Biol. 2013. V. 203. № 6. P. 875–881.

35. Elbaum-Garfinkle S., Brangwynne C.P. // Dev. Cell. 2015. V. 35. № 5. P. 531–532.

36. Zhang H., Elbaum-Garfinkle S., Langdon E.M., Taylor N., Occhipinti P., Bridges A.A., Brangwynne C.P., Gladfelter A.S. // Mol. Cell. 2015. V. 60. № 2. P. 220–230.

37. Hyman A.A., Weber C.A., Jülicher F. // Annu. Rev. Cell Dev. Biol. 2014. V. 30. P. 39–58.

38. Dutertre M., Lambert S., Carreira A., Amor-Guéret M., Vagner S. // Trends Biochem. Sci. 2014. V. 39. № 3. P. 141–149.

39. Kleiman F.E., Manley J.L. // Cell. 2001. V. 104. P. 743–753.

40. Mirkin N., Fonseca D., Mohammed S., Cevher M.A., Manley J.L., Kleiman F.E. // Nucleic Acids Res. 2008. V. 36. № 6. P. 1792–1804.

41. Dutertre M., Sanchez G., Barbier J., Corcos L., Auboeuf D. // RNA Biol. 2011. V. 8. № 5. P. 740–747.

42. Ip J.Y., Schmidt D., Pan Q., Ramani A.K., Fraser A.G., Odom D.T., Blencowe B.J. // Genome Res. 2011. V. 21. № 3. P. 390–401.

43. Fan J., Yang X., Wang W., Wood W.H. 3rd, Becker K.G., Gorospe M. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. № 16. P. 10611–10616.

44. Braunstein S., Badura M.L., Xi Q., Formenti S.C., Schneider R.J. // Mol. Cell Biol. 2009. V. 29. № 21. P. 5645–5656.

45. Kruiswijk F., Yuniati L., Magliozzi R., Low T.Y., Lim R., Bolder R., Mohammed S., Proud C.G., Heck A.J., Pagano M., Guardavaccaro D. // Sci. Signal. 2012. V. 5. № 227. ra 40.

46. Powley I.R., Kondrashov A., Young L.A., Dobbyn H.C., Hill K., Cannell I.G., Stoneley M., Kong Y.W., Cotes J.A., Smith G.C., et al. // Genes Dev. 2009. V. 23. № 10. P. 1207–1220.

47. Mazan-Mamczarz K., Galbán S., López de Silanes I., Martindale J.L., Atasoy U., Keene J.D., Gorospe M. // Proc. Natl. Acad. Sci. USA. 2003. V. 100. № 14. P. 8354–8359.

48. Glorian V., Maillot G., Polès S., Iacovoni J.S., Favre G., Vagner S. // Cell Death Differ. 2011. V. 18. № 11. P. 1692–1701.

49. Wang W., Furneaux H., Cheng H., Caldwell M.C., Hutter D., Liu Y., Holbrook N., Gorospe M. // Mol. Cell Biol. 2000. V. 20. № 3. P. 760–769.

50. Hung T., Wang Y., Lin M.F., Koegel A.K., Kotake Y., Grant G.D., Horlings H.M., Shah N., Umbricht C., Wang P., et al. // Nat. Genet. 2011. V. 43. № 7. P. 621–629.

51. Hegde M.L., Banerjee S., Hegde P.M., Bellot L.J., Hazra T.K., Boldogh I., Mitra S. // J. Biol. Chem. 2012. V. 287. № 41. P. 34202–34211.

52. Anantha R.W., Alcivar A.L., Ma J., Cai H., Simhadri S., Ule J., König J., Xia B. // PLoS One. 2013. V. 8. № 4. e61368.

53. Hong Z., Jiang J., Ma J., Dai S., Xu T., Li H., Yasui A. // PLoS One. 2013. V. 8. № 4. e60208.

54. Rulten S.L., Rotheray A., Green R.L., Grundy G.J., Moore D.A., Gómez-Herreros F., Hafezparast M., Caldecott K.W. // Nucleic Acids Res. 2014. V. 42. № 1. P. 307–314.

55. Wei W., Ba Z., Gao M., Wu Y., Ma Y., Amiard S., White C.I., Rendtlew Danielsen J.M., Yang Y.G., et al. // Cell. 2012. V. 149. № 1. P. 101–112.

56. Azvolinsky A., Giresi P.G., Lieb J.D., Zakian V.A. // Mol. Cell. 2009. V. 34. № 6. P. 722–734.

57. Aguilera A., García-Muse T. // Mol. Cell. 2012. V. 46. № 2. P. 115–124.

58. Tuduri S., Crabbé L., Conti C., Tourrière H., Holtgreve-Grez H., Jauch A., Pantesco V., De Vos J., Thomas A., Theillet C., et al. // Nat. Cell Biol. 2009. V. 11. № 11. P. 1315–1324.

59. Drolet M. // Mol. Microbiol. 2006. V. 59. P. 723–730.

60. Bennetzen M.V., Larsen D.H., Bunkenborg J., Bartek J., Lukas J., Andersen J.S. // Mol. Cell Proteomics. 2010. V. 9. № 6. P. 1314–1323.

61. Bensimon A., Schmidt A., Ziv Y., Elkon R., Wang S.Y., Chen D.J., Aebersold R., Shiloh Y. // Sci. Signal. 2010. V. 3. № 151. rs3.

62. Jungmichel S., Rosenthal F., Altmeyer M., Lukas J., Hottiger M.O., Nielsen M.L. // Mol. Cell. 2013. V. 52. № 2. P. 272–285.

63. Beli P., Lukashchuk N., Wagner S.A., Weinert B.T., Olsen J.V., Baskcomb L., Mann M., Jackson S.P., Choudhary C. // Mol. Cell. 2012. V. 46. № 2. P. 212–225.

64. Koike K., Uchiumi T., Ohga T., Toh S., Wada M., Kohno K., Kuwano M. // FEBS Lett. 1997. V. 417. № 3. P. 390–394.

65. Cammas A., Lewis S.M., Vagner S., Holcik M. // Biochem. Pharmacol. 2008. V. 76. № 11. P. 1395–1403.

66. Lukong K.E., Chang K.W., Khandjian E.W., Richard S. // Trends Genet. 2008. V. 24. № 8. P. 416–425.

67. Cooper T.A., Wan L., Dreyfuss G. // Cell. 2009. V. 136. № 4. P. 777–793.

68. Braunschweig U., Gueroussov S., Plocik A.M., Graveley B.R., Blencowe B.J. // Cell. 2013. V. 152. № 6. P. 1252–1269.

69. Shi Y., Manley J.L. // Genes Dev. 2015. V. 29. № 9. P. 889–897.

70. Ha M., Kim V.N. // Nat. Rev. Mol. Cell Biol. 2014. V. 15. № 8. P. 509–524.

71. Rinn J.L. // Cold Spring Harb. Perspect. Biol. 2014. V. 6. № 8. P. a018614.

72. Lasda E., Parker R. // RNA. 2014. V. 20. № 12. P. 1829–1842.

73. Gerstberger S., Hafner M., Tuschl T. // Nat. Rev. Genet. 2014. V. 15. № 12. P. 829–845.

74. Neelamraju Y., Hashemikhabir S., Janga S.C. // J. Proteomics. 2015. V. 127. Pt A. P. 61–70.

75. Lunde B.M., Moore C., Varani G. // Nat. Rev. Mol. Cell Biol. 2007. V. 8. № 6. P. 479–490.

76. Leung A.K., Vyas S., Rood J.E., Bhutkar A., Sharp P.A., Chang P. // Mol. Cell. 2011. V. 42. № 4. P. 489–499.

77. Leung A., Todorova T., Ando Y., Chang P. // RNA Biol. 2012. V. 9. № 5. P. 542–548.

78. Wang X., McLachlan J., Zamore P.D., Hall T.M. // Cell. 2002. V. 110. № 4. P. 501–512.

79. Cheong C.G., Hall T.M. // Proc. Natl. Acad. Sci. USA. 2006. V. 103. № 37. P. 13635–13639.

80. Kielkopf C.L., Rodionova N.A., Green M.R., Burley S.K. // Cell. 2001. V. 106. № 5. P. 595–605.

81. Järvelin A.I., Noerenberg M., Davis I., Castello A. // Cell Commun. Signal. 2016. V. 14. № 9. doi: 10.1186/s12964-016-0132-3

82. Castello A., Fischer B., Eichelbaum K., Horos R., Beckmann B.M., Strein C., Davey N.E., Humphreys D.T., Preiss T., Steinmetz L.M., et al. // Cell. 2012. V. 149. № 6. P. 1393–1406.

83. Beckmann B.M., Horos R., Fischer B., Castello A., Eichelbaum K., Alleaume A.M., Schwarzl T., Curk T., Foehr S., Huber W., et al. // Nat. Commun. 2015. V. 6. № 10127. P. 1–9.

84. Livesay D.R. // Curr. Opin. Pharmacol. 2010. V. 10. № 6. P. 706–708.

85. Uversky V.N. // J. Biomol. Struct. Dyn. 2003. V. 21. № 2. P. 211–234.

86. Tsvetkov P., Asher G., Paz A., Reuven N., Sussman J.L., Silman I., Shaul Y. // Proteins. 2008. V. 70. № 4. P. 1357–1366.

87. Dunker A.K., Uversky V.N. // Curr. Opin. Pharmacol. 2010. V. 10. № 6. P. 782–788.

88. Uversky V.N. // J. Biol. Chem. 2016. V. 291. № 13. P. 6681–6688.

89. Uversky V.N., Gillespie J.R., Fink A.L. // Proteins. 2000. V. 41. № 3. P. 415–427.

90. Dunker A.K., Lawson J.D., Brown C.J., Williams R.M., Romero P., Oh J.S., Oldfield C.J., Campen A.M., Ratliff C.M., Hipps K.W., et al. // J. Mol. Graph. Model. 2001. V. 19. № 1. P. 26–59.

91. Uversky V.N. // Eur. J. Biochem. 2002. V. 269. № 1. P. 2–12.

92. Uversky V.N., Li J., Fink A.L. // J. Biol. Chem. 2001. V. 276. P. 10737–10744.

93. Goto Y., Takahashi N., Fink A.L. // Biochemistry. 1990. V. 29. № 14. P. 3480–3488.

94. Fink A.L., Calciano L.J., Goto Y., Kurotsu T., Palleros D.R. // Biochemistry. 1994. V. 33. № 41. P. 12504–12511.

95. Uversky V.N., Narizhneva N.V. // Biochemistry (Mosc.). 1998. V. 63. № 4. P. 420–433.

96. Witze E.S., Old W.M., Resing K.A., Ahn N.G. // Nat. Methods. 2007. V. 4. № 10. P. 798–806.

97. Walsh C.T., Garneau-Tsodikova S., Gatto G.J. Jr. // Angew Chem. Int. Ed. Engl. 2005. V. 44. № 45. P. 7342–7372.

98. Dunker A.K., Brown C.J., Obradovic Z. // Adv. Protein Chem. 2002. V. 62. P. 25–49.

99. Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Obradovic Z., Uversky V.N. // J. Proteome Res. 2007. V. 6. № 5. P. 1917–1932.

100. Cumberworth A., Lamour G., Babu M.M., Gsponer J. // Biochem J. 2013. V. 454. № 3. P. 361–369.

101. Ward J.J., Sodhi J.S., McGuffin L.J., Buxton B.F., Jones D.T. // J. Mol. Biol. 2004. V. 337. № 3. P. 635–645.

102. Xie H., Vucetic S., Iakoucheva L.M., Oldfield C.J., Dunker A.K., Uversky V.N., Obradovic Z. // J. Proteome Res. 2007. V. 6. № 5. P. 1882–1898.

103. Dunker A.K., Cortese M.S., Romero P., Iakoucheva L.M., Uversky V.N. // FEBS J. 2005. V. 272. № 20. P. 5129–5148.

104. Ekman D., Light S., Björklund A.K., Elofsson A. // Genome Biol. 2006. V. 7. № 6. R45.

105. Higurashi M., Ishida T., Kinoshita K. // Protein Sci. 2008. V. 17. № 1. P. 72–78.

106. Patil A., Kinoshita K., Nakamura H. // Protein Sci. 2010. V. 19. № 8. P. 1461–1468.

107. Singh G.P., Ganapathi M., Dash D. // Proteins. 2007. V. 66. № 4. P. 761–765.

108. Trudeau T., Nassar R., Cumberworth A., Wong E.T., Woollard G., Gsponer J. // Structure. 2013. V. 21. № 3. P. 332–341.

109. Dawicki-McKenna J.M., Langelier M.F., DeNizio J.E., Riccio A.A., Cao C.D., Karch K.R., McCauley M., Steffen J.D., Black B.E., Pascal J.M. // Mol. Cell. 2015. V. 60. № 5. P. 755–768.

110. Tapley T.L., Körner J.L., Barge M.T., Hupfeld J., Schauerte J.A., Gafni A., Jakob U., Bardwell J.C. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. № 14. P. 5557–5562.

111. Chakrabortee S., Tripathi R., Watson M., Schierle G.S., Kurniawan D.P., Kaminski C.F., Wise M.J., Tunnacliffe A. // Mol. Biosyst. 2012. V. 8. № 1. P. 210–219.

112. Buljan M., Chalancon G., Eustermann S., Wagner G.P., Fuxreiter M., Bateman A., Babu M.M. // Mol. Cell. 2012. V. 46. № 6. P. 871–883.

113. Ellis J.D., Barrios-Rodiles M., Colak R., Irimia M., Kim T., Calarco J.A., Wang X., Pan Q., O'Hanlon D., Kim P.M., et al. // Mol. Cell. 2012. V. 46. № 6. P. 884–892.

114. Hegde M.L., Izumi T., Mitra S. // Prog. Mol. Biol. Transl. Sci. 2012. V. 110. P. 123–153.

115. Hegde M.L., Hazra T.K., Mitra S. // Cell Mol. Life Sci. 2010. V. 67. № 21. P. 3573–3587.

116. Mittag T., Kay L.E., Forman-Kay J.D. // J. Mol. Recog. 2010. V. 23. № 2. P. 105–116.

117. Gunasekaran K., Tsai C.J., Kumar S., Zanuy D., Nussinov R. // Trends Biochem. Sci. 2003. V. 28. № 2. P. 81–85.

118. Krueger K.E., Srivastava S. // Mol. Cell Proteomics. 2006. V. 5. № 10. P. 1799–1810.

119. Seet B.T., Dikic I., Zhou M.M., Pawson T. // Nat. Rev. Mol. Cell Biol. 2006. V. 7. № 7. P. 473–483.

120. Brangwynne C.P., Eckmann C.R., Courson D.S., Rybarska A., Hoege C., Gharakhani J., Jülicher F., Hyman A.A. // Science. 2009. V. 324. № 5935. P. 1729–1732.

121. Brangwynne C.P., Mitchison T.J., Hyman A.A. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 11. P. 4334–4339.

122. Han T.W., Kato M., Xie S., Wu L.C., Mirzaei H., Pei J., Chen M., Xie Y., Allen J., Xiao G., et al. // Cell. 2012. V. 149. № 4. P. 768–779.

123. Zwicker D., Decker M., Jaensch S., Hyman A.A., Jülicher F. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. № 26. P. 2636–2645.

124. Altmeyer M., Neelsen K.J., Teloni F., Pozdnyakova I., Pellegrino S., Grøfte M., Rask M.B., Streicher W., Jungmichel S., et al. // Nat. Commun. 2015. V. 6. № 8088. P. 1–12.

125. Teloni F., Altmeyer M. // Nucleic Acids Res. 2016. V. 44. № 3. P. 993–1006.

126. Weber S.C., Brangwynne C.P. // Curr. Biol. 2015. V. 25. № 5. P. 641–646.

127. Nott T.J., Petsalaki E., Farber P., Jervis D., Fussner E., Plochowietz A., Craggs T.D., Bazett-Jones D.P., Pawson T., Forman-Kay J.D., et al. // Mol. Cell. 2015. V. 57. № 5. P. 936–947.

128. Shevtsov S.P., Dundr M. // Nat. Cell Biol. 2011. V. 13. № 2. P. 167–173.

129. Nott T.J., Petsalaki E., Farber P., Jervis D., Fussner E., Plochowietz A., Craggs T.D., Bazett-Jones D.P., Pawson T., Forman-Kay J.D., et al. // Mol. Cell. 2015. V. 57. № 5. P. 936–947.

130. Bürkle A. // FEBS J. 2005. V. 272. № 18. P. 4576–4589.

131. D'Annessa I., Coletta A., Desideri A. // Biopolymers. 2014. V. 101. № 1. P. 78–86.

132. Schultheisz H.L., Szymczyna B.R., Williamson J.R. // J. Am. Chem. Soc. 2009. V. 131. № 40. P. 14571–14578.

133. Minaga T., Kun E. // J. Biol. Chem. 1983. V. 258. № 2. P. 725–730.

134. D'Amours D., Desnoyers S., D'Silva I., Poirier G.G. // J. Biol. Chem. 1999. V. 342. Pt. 2. P. 249–268.

135. Wielckens K., George E., Pless T., Hilz H. // J. Biol. Chem. 1983. V. 25. № 7. P. 4098–4104.

136. Kreimeyer A., Wielckens K., Adamietz P., Hilz H. // J. Biol. Chem. 1984. V. 259. № 2. P. 890–896.

137. Alvarez-Gonzalez R., Althaus F.R. // Mutat. Res. 1989. V. 218. № 2. P. 67–74.

138. Bock F.J., Chang P. // FEBS J. 2016. V. 283. № 22. P. 4017–4031.

139. Hottiger M.O., Hassa P.O., Lüscher B., Schüler H., Koch-Nolte F. // Trends Biochem. Sci. 2010. V. 35. № 4. P. 208–219.

140. Beck C., Robert I., Reina-San-Martin B., Schreiber V., Dantzer F. // Exp. Cell Res. 2014. V. 329. № 1. P. 18–25.

141. Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S. // Mol. Cell Biol. 2002. V. 22. № 1. P. 332–342.

142. Chang P., Coughlin M., Mitchison T.J. // Nat. Cell Biol. 2005. V. 7. № 11. P. 1133–1139.

143. Ozaki Y., Matsui H., Asou H., Nagamachi A., Aki D., Honda H., Yasunaga S., Takihara Y., Yamamoto T., Izumi S., et al. // Mol. Cell. 2012. V. 47. № 5. P. 694–706.

144. Lonskaya I., Potaman V.N., Shlyakhtenko L.S., Oussatcheva E.A., Lyubchenko Y.L., Soldatenkov V.A. // J. Biol. Chem. 2005. V. 280. № 17. P. 17076–17083.

145. Langelier M.F., Planck J.L., Roy S., Pascal J.M. // Science. 2012. V. 336. № 6082. P. 728–732.

146. Daniels C.M., Ong S.E., Leung A.K. // Mol. Cell. 2015. V. 58. № 6. P. 911–924.

147. Altmeyer M., Messner S., Hassa P.O., Fey M., Hottiger M.O. // Nucleic Acids Res. 2009. V. 37. № 11. P. 3723–3738.

148. Zhang Y., Wang J., Ding M., Yu Y. // Nat. Methods. 2013. V. 10. № 10. P. 981–984.

149. Hottiger M.O. // Annu. Rev. Biochem. 2015. V. 84. P. 227–263.

150. Gagné J.P., Isabelle M., Lo K.S., Bourassa S., Hendzel M.J., Dawson V.L., Dawson T.M., Poirier G.G. // Nucleic Acids Res. 2008. V. 36. № 22. P. 6959–6976.

151. Krietsch J., Rouleau M., Pic É., Ethier C., Dawson T.M., Dawson V.L., Masson J.Y., Poirier G.G., Gagné J.P. // Mol. Aspects Med. 2013. V. 34. № 6. P. 1066–1087.

152. Krietsch J., Caron M.C., Gagné J.P., Ethier C., Vignard J., Vincent M., Rouleau M., Hendzel M.J., Poirier G.G., Masson J.Y. // Nucleic Acids Res. 2012. V. 40. № 20. P. 10287–10301.

153. Zhang F., Shi J., Chen S.H., Bian C., Yu X. // Nucleic Acids Res. 2015. V. 43. № 22. P. 10782–10794.

154. Izhar L., Adamson B., Ciccia A., Lewis J., Pontano-Vaites L., Leng Y., Liang A.C., Westbrook T.F., Harper J.W., Elledge S.J. // Cell Rep. 2015. V. 11. № 9. P. 1486–1500.

155. Isabelle M., Gagné J.P., Gallouzi I.E., Poirier G.G. // J. Cell Sci. 2012. V. 125. Pt. 19. P. 4555–4566.

156. Boamah E.K., Kotova E., Garabedian M., Jarnik M., Tulin A.V. // PLoS Genet. 2012. V. 8. № 1. e1002442.

157. Ji Y., Tulin A.V. // Int. J. Mol. Sci. 2013. V. 14. № 8. P. 16168–16183.

158. Kraus W.L., Hottiger M.O. // Mol. Aspects Med. 2013. V. 34. № 6. P. 1109–1123.

159. Kwon I., Kato M., Xiang S., Wu L., Theodoropoulos P., Mirzaei H., Han T., Xie S., Corden J.L., McKnight S.L. // Cell. 2013. V. 155. № 5. P. 1049–1060.

160. Andrabi S.A., Umanah G.K., Chang C., Stevens D.A., Karuppagounder S.S., Gagné J.P., Poirier G.G., Dawson V.L., Dawson T.M. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. № 28. P. 10209–10214.

161. Barkauskaite E., Jankevicius G., Ahel I. // Mol. Cell. 2015. V. 58. № 6. P. 935–946.

162. Barkauskaite E., Brassington A., Tan E.S., Warwicker J., Dunstan M.S., Banos B., Lafite P., Ahel M., Mitchison T.J., Ahel I., et al. // Nat. Commun. 2013. V. 4. № 2164. P. 1–8.

163. Dunstan M.S., Barkauskaite E., Lafite P., Knezevic C.E., Brassington A., Ahel M., Hergenrother P.J., Leys D., Ahel I. // Nat. Commun. 2012. V. 3. № 878. P. 1–6.

164. Jankevicius G., Hassler M., Golia B., Rybin V., Zacharias M., Timinszky G., Ladurner A.G. // Nat. Struct. Mol. Biol. 2013. V. 20. № 4. P. 508–514.

165. Tafuri S.R., Wolffe A.P. // New Biol. 1992. V. 4. № 4. P. 349–359.

166. Hasegawa S.L., Doetsch P.W., Hamilton K.K., Martin A.M., Okenquist S.A., Lenz J., Boss J.M. // Nucl. Acids Res. 1991. V. 19. № 18. P. 4915–4920.

167. Eliseeva I.A., Kim E.R., Guryanov S.G., Ovchinnikov L.P., Lyabin D.N. // Biochemistry (Mosc.). 2011. V. 76. № 13. P. 1402–1433.

168. Gaudreault I., Guay D., Lebel M. // Nucleic Acids Res. 2004. V. 32. № 1. P. 316–327.

169. En-Nia A., Yilmaz E., Klinge U., Lovett D.H., Stefanidis I., Mertens P.R. // J. Biol. Chem. 2005. V. 280. № 9. P. 7702–7711.

170. Lasham A., Moloney S., Hale T., Homer C., Zhang Y.F., Murison J.G., Braithwaite A.W., Watson J. // J. Biol. Chem. 2003. V. 278. № 37. P. 35516–35523.

171. Soop T., Nashchekin D., Zhao J., Sun X., Alzhanova-Ericsson A.T., Björkroth B., Ovchinnikov L., Daneholt B. // J. Cell Sci. 2003. V. 116. Pt. 8. P. 1493–1503.

172. Blobel G. // Biochem. Biophys. Res. Commun. 1972. V. 4. № 1. P. 88–95.

173. Evdokimova V.M., Kovrigina E.A., Nashchekin D.V., Davydova E.K., Hershey J.W., Ovchinnikov L.P. // J. Biol. Chem. 1998. V. 273. № 6. P. 3574–3581.

174. Liu T.T., Arango-Argoty G., Li Z., Lin Y., Kim S.W., Dueck A., Ozsolak F., Monaghan A.P., Meister G., DeFranco D.B., et al. // RNA. 2015. V. 21. № 6. P. 1159–1172.

175. Wu S.L., Fu X., Huang J., Jia T.T., Zong F.Y., Mu S.R., Zhu H., Yan Y., Qiu S., Wu Q., et al. // Nucleic Acids Res. 2015. V. 43. № 17. P. 8516–8528.

176. Hayakawa H., Uchiumi T., Fukuda T., Ashizuka M., Kohno K., Kuwano M., Sekiguchi M. // Biochemistry. 2002. V. 41. № 42. P. 12739–12744.

177. Das S., Chattopadhyay R., Bhakat K.K., Boldogh I., Kohno K., Prasad R., Wilson S.H., Hazra T.K. // J. Biol. Chem. 2007. V. 282. № 39. P. 28474–28484.

178. Ohga T., Koike K., Ono M., Makino Y., Itagaki Y., Tanimoto M., Kuwano M., Kohno K. // Cancer Res. 1996. V. 56. № 18. P. 4224–4228.

179. Koike K., Uchiumi T., Ohga T., Toh S., Wada M., Kohno K., Kuwano M. // FEBS Lett. 1997. V. 417. № 3. P. 390–394.

180. Fujita T., Ito K., Izumi H., Kimura M., Sano M., Nakagomi H., Maeno K., Hama Y., Shingu K., Tsuchiya S., et al. // Clin. Cancer Res. 2005. V. 11. № 24. P. 8837–8844.

181. Sorokin A.V., Selyutina A.A., Skabkin M.A., Guryanov S.G., Nazimov I.V., Richard C., Th'ng J., Yau J., Sorensen P.H., Ovchinnikov L.P., et al. // EMBO J. 2005. V. 24. № 20. P. 3602–3612.

182. Kretov D.A., Curmi P.A., Hamon L., Abrakhi S., Desforges B., Ovchinnikov L.P., Pastré D. // Nucleic Acids Res. 2015. V. 43. № 19. P. 9457–9473.

183. Selivanova O.M., Guryanov S.G., Enin G.A., Skabkin M.A., Ovchinnikov L.P., Serdyuk I.N. // Biochemistry (Mosc.). 2010. V. 75. № 1. P. 115–120.

184. Sengupta S., Mantha A.K., Mitra S., Bhakat K.K. // Oncogene. 2011. V. 30. № 4. P. 482–493.

185. Gaudreault I., Guay D., Lebel M. // Nucleic Acids Res. 2004. V. 32. № 1. P. 316–327.

186. Ise T., Nagatani G., Imamura T., Kato K., Takano H., Nomoto M., Izumi H., Ohmori H., Okamoto T., Ohga T., *et al.* // Cancer Res. 1999. V. 59. № 2. P. 342–346.

187. Alemasova E.E., Moor N.A., Naumenko K.N., Kutuzov M.M., Sukhanova M.V., Pestryakov P.E., Lavrik O.I. // Biochim. Biophys. Acta. 2016. V. 1864. № 12. P. 1631–1640.

188. Fomina E.E., Pestryakov P.E., Maltseva E.A., Petruseva I.O., Kretov D.A., Ovchinnikov L.P., Lavrik O.I. // Biochemistry (Mosc.). 2015. V. 80. № 2. P. 219–227.

189. Marenstein D.R., Ocampo M.T., Chan M.K., Altamirano A., Basu A.K., Boorstein R.J., Cunningham R.P., Teebor G.W. // Biol. Chem. 2001. V. 276. № 24. P. 21242–21249.

190. Pestryakov P., Zharkov D.O., Grin I., Fomina E.E., Kim E.R., Hamon L., Eliseeva I.A., Petruseva I.O., Curmi P.A., Ovchinnikov L.P., *et al.* // J. Mol. Recognit. 2012. V. 25. № 4. P. 224–233.

191. Fomina E.E., Pestryakov P.E., Kretov D.A., Zharkov D.O., Ovchinnikov L.P., Curmi P.A., Lavrik O.I. // J. Mol. Recognit. 2015. V. 28. № 2. P. 117–123.

192. Yang W.H., Bloch D.B. // RNA. 2007. V. 13. № 5. P. 704–712.

193. Kawaguchi A., Asaka M.N., Matsumoto K., Nagata K. // Sci. Rep. 2015. V. 5. P. 8768.

194. Gonda K., Wudel J., Nelson D., Katoku-Kikyo N., Reed P., Tamada H., Kikyo N. // J. Biol. Chem. 2006. V. 281. № 12. P. 8153–8160.

195. Chang P., Jacobson M.K., Mitchison T.J. // Nature. 2004. V. 432. № 7017. P. 645–649.

196. Alemasova E.E., Pestryakov P.E., Sukhanova M.V., Kretov D.A., Moor N.A., Curmi P.A., Ovchinnikov L.P., Lavrik O.I. // Biochimie. 2015. V. 119. P. 36–44.

197. Uversky V.N. // Biochim. Biophys. Acta. 2013. V. 1834. № 5. P. 932–951.

Glutamyl Endopeptidases: The Puzzle of Substrate Specificity

I.V. Demidyuk*, K.N. Chukhontseva, S.V. Kostrov
Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq., 2, Moscow, 123182, Russia
*E-mail: duk@img.ras.ru
Received November 2, 2016; in final form, February 22, 2017

ABSTRACT Glutamyl endopeptidases (GEPases) are chymotrypsin-like enzymes that preferentially cleave the peptide bonds of the α-carboxyl groups of glutamic acid. Despite the many years of research, the structural determinants underlying the strong substrate specificity of GEPases still remain unclear. In this review, data concerning the molecular mechanisms that determine the substrate preference of GEPases is generalized. In addition, the biological functions of and modern trends in the research into these enzymes are outlined.
KEYWORDS 3C-like serine protease, chymotrypsin-like protease, epidermolytic toxin, glutamyl endopeptidase, substrate specificity, V8 protease.
ABBREVIATIONS 3Cpro − picornaviral 3C protease; 3CLpro − 3C-like protease; 3CLSP − 3C-like serine protease; BIGEP − GEPase of *B. intermedius*; Boc-AAPE − *tert*-butyloxycarbonyl-Ala-Ala-Pro-Glu; CLP − chymotrypsin-like protease; EAV-nsp4 − nonstructural protein 4 of equine arteritis virus; Esp − extracellular serine protease of *S. epidermidis*; ET − epidermolytic toxin; ETA, ETB − epidermolytic toxins A and B of *S. aureus*; GEPase − glutamyl endopeptidase; (+)RNA-virus − positive-sense single-stranded RNA virus; Glu-SGP − GEPase of *Str. griseus;* Glu/Gln-P1 − an amino acid residue at position P1 of a substrate; Glu-V8 − protease V8 of *S. aureus;* HAstV-pro − human astrovirus protease; PDB ID − Protein Data Bank (http://www.rcsb.org) identifier; PRRSV-nsp4 − nonstructural protein 4 of porcine reproductive and respiratory syndrome; SeMV-pro − *Sesbania* mosaic virus protease.

GLUTAMYL ENDOPEPTIDASES AS MEMBERS OF THE STRUCTURAL CHYMOTRYPSIN FAMILY

Glutamyl endopeptidases (GEPases) are enzymes that preferentially cleave the bonds of the α-carboxyl groups of glutamic acid [1, 2]. GEPases from a number of gram-positive bacteria [23−25] and (+)RNA viruses have been characterized to date. All GEPases belong to the structural chymotrypsin family, which is one of the most extensive and well-studied families. Chymotrypsin-like protease (CLP) molecules share their spatial organization principle; the so-called chymotrypsin (or trypsin) fold (*Fig. 1*). The residue at the P1 position is a key determinant of the hydrolysis sites of CLPs (according to the Schechter and Berger nomenclature, the cleaved bond of the substrate is located downstream of the P1 residue, which corresponds to the S1-binding site of an enzyme [26]). Similar to pancreatic serine proteases, CLPs are conventionally classified into three main groups: 1) hydrolyzing bonds formed by the α-carboxyl groups of large hydrophobic amino acid residues (chymotrypsin-like specificity), 2) cleaving bonds downstream of positively charged residues (trypsin-like specificity), and 3) preferring small

hydrophobic residues at the P1 position (elastase-like specificity) [27]. Furthermore, CLPs with mixed specificity have been discovered. For example, collagenolytic enzymes isolated from crabs exhibit the combined specificity of trypsin, chymotrypsin, and elastase [28], while bovine duodenase [29] and cathepsin G [30] can efficiently hydrolyze the substrates of both trypsin and chymotrypsin. In addition, CLPs cleaving bonds preferentially downstream of the Gln residue (e.g., many 3C-like viral proteases [23]) and being specific to negatively charged amino acid residues (e.g., granzyme B that preferentially hydrolyzes bonds downstream of Asp residues [31] and the GEPases that this review focuses on) are known.

CLP molecules consist of two perpendicular β-cylindrical domains and a C-terminal α-helix (*Fig. 1*). The catalytic and substrate-binding sites reside in the cleft between the two β-cylinders. The functionally important residues are predominantly localized in the loops connecting the β-strands. The S1 pocket lying next to the catalytic residue Ser(Cys)195 (hereinafter, chymotrypsin numbering is used) is formed by the regions 189−192, 214−216, and 224−228. In most cases,
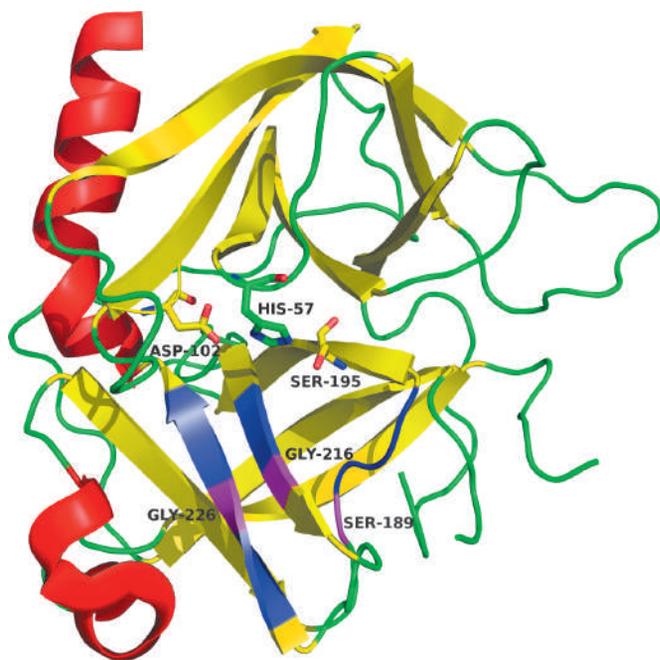
Fig. 1. Three-dimensional structure of chymotrypsin (PDB ID – 5cha). The catalytic triad residues are shown as sticks. The regions forming the S1 pocket are shown in blue; the positions of the key residues of the S1 pocket are shown in magenta. All 3D structure pictures were generated using the PyMOL Molecular Graphics System (www.pymol.org).

the residues at positions 189, 216, and 226 are the key determinants of substrate specificity [32, 33]. The enzymes capable of recognizing charged residues at position P1 carry residues compensating for the substrate charge at position 189 (Asp in trypsin [34]) or 226 (Arg in granzyme B [35], Glu in cathepsin G [36], and Asp in crab collagenase[37] and duodenase [38]). This gives grounds for believing that the primary substrate specificity of CLPs is controlled by a relatively small number of structural elements of the S1 site. However, the substrate specificity cannot be "switched" by just transferring these structural elements from one molecule into another.

As it has been demonstrated for the conversion of trypsin to chymotrypsin, specificity is also affected by a combination of remote structural elements that do not directly interact with the substrate. The S1 sites are similar in both enzymes. However, substitution of the main determinant of the binding of the charged substrates of trypsin Asp189 with Ser, which is typical of chymotrypsin, does not induce the corresponding specificity. Instead, a low-efficiency nonspecific protease is formed [39]. Ensuring chymotrypsin-like

specificity requires substitution of four residues in the S1 pocket and modification of the regions remote from the S1 site: two surface loops that do not come into direct contact with the substrate [40] and Tyr172 residue [41]. Comparison of the crystalline structures and kinetic characteristics of the resulting variants to those of chymotrypsin and trypsin demonstrates that additional modifications are important for accurate positioning of the bond being cleaved with respect to the catalytic center of the protein (the Ser195−His57 pair and the oxyanion hole) rather than for binding the P1 residue [40−43].

Hence, according to the data on the structural determinants of the substrate specificity of CLPs, one can expect that the preference of negatively charged amino acid residues at the P1 position by GEPases is determined by the same regions of the polypeptide chain as in other enzymes belonging to this group. The substrate charge compensator is expected to be the key structural determinant of specificity, as well as in all the CLPs recognizing charged P1 residues. Arg or Lys at position 189 or 226 can be suggested as candidates for this. Meanwhile, one should bear in mind that the structure of the regions remote from S1 plays a significant role in high-efficiency interaction with the P1 residue.

GLUTAMYL ENDOPEPTIDASE FROM
STREPTOMYCES GRISEUS

Glu-specific protease from *S. griseus* (Glu-SGP) (PDB ID − 1hpg) was the first GEPase whose spatial structure was determined [44]. The structure of this enzyme is generally typical for CLPs (*Fig. 2A*) and is the most similar to that of bacterial CLPs (proteases A and B from *Str. griseus* and α-lytic protease). The overall geometry of the S1 site is also very close to the geometry of this region in the aforelisted bacterial enzymes. Contrary to expectations, no explicit compensator for the negative charge of the substrate, Lys or Arg residue, was detected in the S1 site. The carboxyl group of Glu at position P1 of the substrate forms hydrogen bonds with Ser190 (192 if numbering [44] is used), Ser126, and His213. Hence, these residues probably play the key role in substrate recognition. The side chain of histidine can be positively charged. However, if pK_a of the side chain of His213 in the absence of the substrate is taken to be 6.4, the imidazole ring will be protonated by less than 1% at the pH 8.5 that is optimum for the functioning of Glu−SGP; therefore, histidine is expected to be neutrally charged [44]. Meanwhile, pK_a of amino acid residues in the proteins can vary significantly depending on the environment [45]. An analysis of the Glu-SGP structure has revealed that it carries the so-called histidine triad containing His199 and His228, along with His213. The three His residues permeate
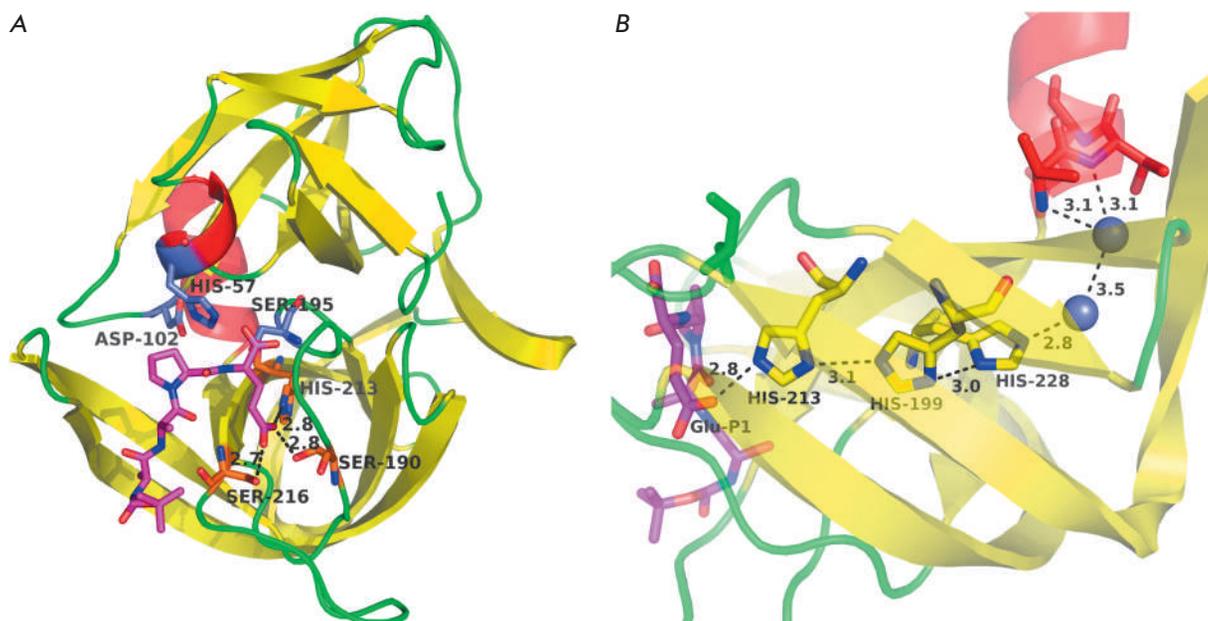
Fig. 2. Three-dimensional structure of the glutamyl endopeptidase of *Streptomyces griseus* (1hpg). A – general view. B – the histidine triad. The Boc-AAPE ligand (the structure of the protecting group is not shown) is colored in magenta; the catalytic triad residues, in blue; the residues directly interacting with the carboxyl group of Glu-P1, in orange; and the histidine triad, in yellow. Water molecules are represented as blue spheres. The distances are given in angstroms.

the C-terminal β-cylindrical domain to form a chain of hydrogen bonds that links the carboxyl group of the substrate Glu−P1 and, via two water molecules bound to the enzyme, the N-terminal rim of the C-terminal α-helix of the molecule (*Fig. 2B*). It was postulated that this very structure ensures the transfer of the positive charge compensating for the substrate charge from the microdipole of the α-helix to His213 of the substrate-binding site [44]. Let us mention that the histidine triad residues in GEPases are not conserved [46] and, in addition to Glu−SGP, have been found only in the highly homologous enzyme from *Str. fradiae* [13].

The role of the residues forming the S1 pocket and the histidine triad Glu-SGP was investigated by site-directed mutagenesis. Any modifications to Ser190(192) (Ala/Gly/Asn/Thr/Val) and His213 (Ala/Gly/Lys/Asn/Arg/Ser/Val) stop the autocatalytic processing (at Glu(-1)−Val1 bond) of the GEPase precursor, which proves that these residues play a fundamental role in the formation of the S1 site. Meanwhile, Ser216 seems to be less important, since its substitution for Ala or Gly does not result in a loss of activity by the enzyme. A similar result was observed for certain modifications of histidine triad residues: the mutations His199→Val and His228→Ala/Asp/Asn/Ser/Val do not impede enzyme processing. All the mutant proteins (His199→Val, Ser216→Ala, Ser216→Gly, and His228→Ala) whose specificities have been studied maintained their pref-

erence for the substrates carrying Glu-P1 [47]. Hence, the hypothesis of the significance of the histidine triad in charge compensation has not been confirmed experimentally and the Ser190(192) and His213 residues are now believed to play a key role in substrate recognition.

Thus, while the structure of the S1 site is already known, it remains unclear how the elements forming this site can ensure the observed substrate specificity. This controversy remains even more explicit once the data on the structure and specificity of viral 3C-like serine proteases are examined.

VIRAL 3C-LIKE SERINE PROTEASES
Processing of polyprotein precursors is an integral part of the life cycle of most (+)RNA viruses [48−50] and typically involves viral papain-like or chymotrypsin-like proteases, components of the polyprotein [51]. Most CLPs from (+)RNA viruses are cysteine proteases, such as 3C proteases (3Cpro) of picornaviruses or 3C-like proteases (3CLpro) of corona-, poty-, or comoviruses [49]. Meanwhile, some enzymes whose active sites contain the serine catalytic residue have been identified. These proteins are denoted as 3C-like serine proteases (3CLSP) [23]. CLPs from (+)RNA viruses exhibit a narrow substrate specificity. The hydrolysis sites of the 3C and 3C-like proteases are generally similar and usually contain a Gln or Glu residue at the P1 position along with a small amino acid residue located downstream
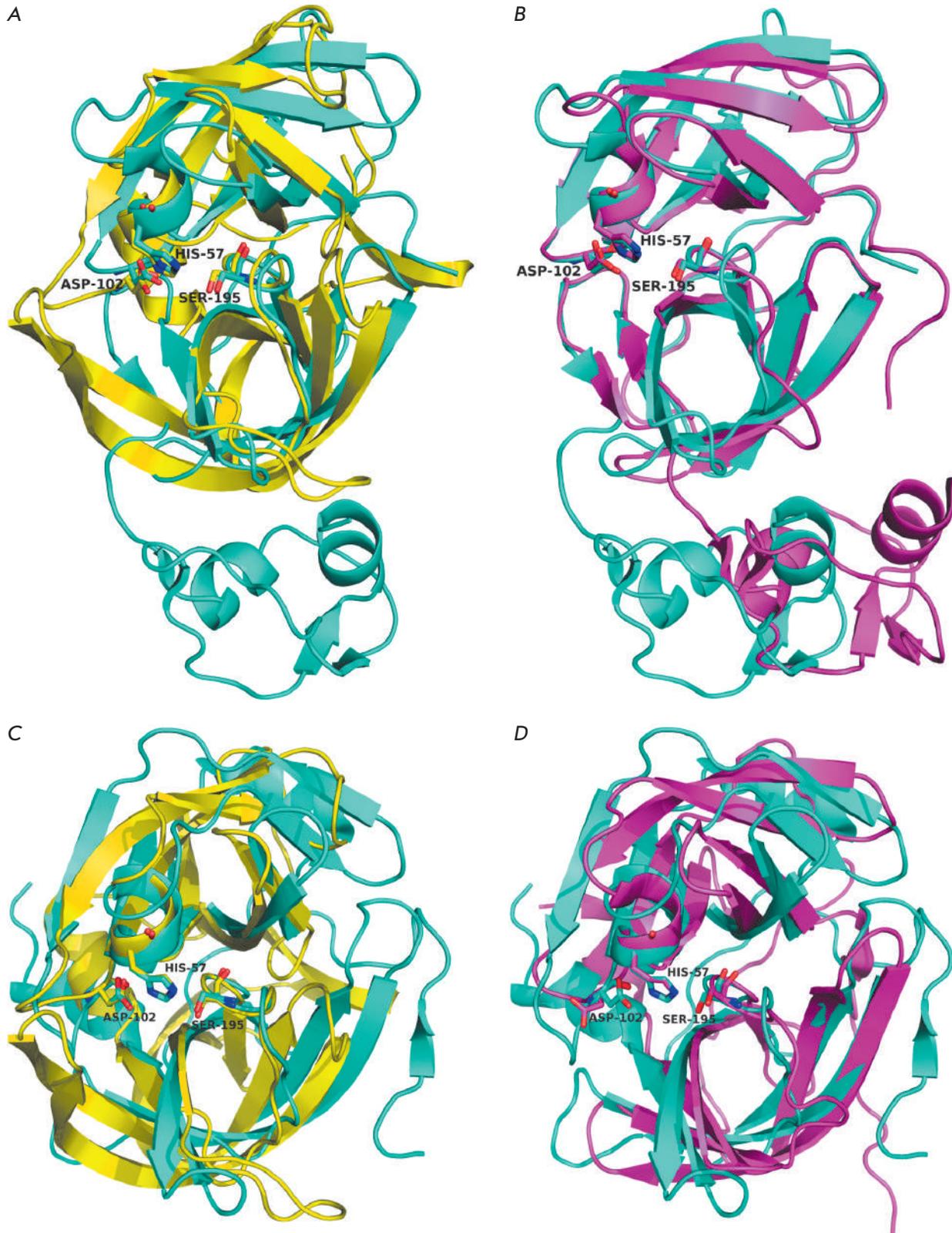
Fig. 3. Three-dimensional structures of viral glutamyl endopeptidases. A – EAV-nsp4 (PDB ID – 1mbm; blue) and Glu-SGP (1hpg; yellow). B – EAV-nsp4 (cyan) and PRRSV-nsp4 (3fan; magenta). C – Glu-SGP (yellow) and SeMV-pro (1zyo; cyan). D – SeMV-pro (cyan) and HAstV-pro (2w5e; magenta). The catalytic triad residues are designated.

(Gly, Ala, or Ser) [23, 52]. Some proteases cleave the bonds formed by both Gln and Glu [53−57], while others prefer Gln-P1 (e.g., 3Cpro or 3CLpro of picornaviruses and coronaviruses [23, 48]) or are true GEPases cleaving the polypeptide chain right after Glu. Such specificity is exhibited by CLPs of arteri-[23], sobemo-[25], and astroviruses [24].

Arteriviral GEPases denoted as Nsp4 (nonstructural protein 4) [23] are serine proteases [58, 59]. Their properties have been studied, and the spatial structures of Nsp4 of the equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV) have been identified. The 3D structure of EAV-Nsp4 is generally typical of CLPs (PDB ID − 1mbm). Meanwhile, the catalytic domain of the enzyme formed by two perpendicular β-cylinders also has a C-terminal extension (*Fig. 3A*) [60]. The structure of PRRSV-Nsp4 (PDB ID − 3fan) is similar to that of EAV-Nsp4; however, it noticeably differs in the mutual arrangement of the catalytic and C-terminal domains (*Fig. 3B*) [59].

The architectures of the S1 sites of EAV-Nsp4 and Glu-SGP are very similar (*Fig. 4A*). The S1 pocket contains the same three main structural elements: His213 (134 in EAV-Nsp4, 1198 in polyprotein), Thr190 (115, 1179) corresponding to Ser190 in Glu-SGP, and Ser216 (137, 1201) [60]. All three residues are also found in the primary structure of PRRSV-Nsp4 [58, 59]. However, the crystal structure analysis data show that the S1 site of the latter enzyme has a structure different from those of EAV-Nsp4 and Glu-SGP (*Fig. 4B*). The position of the polypeptide chain region 190−194 (113−117 in PRRSV-Nsp4) is altered compared to that in most CLPs, resulting in a nontypical configuration of the oxyanion hole and a significant distance between Thr190(113) and the carboxyl group of Glu-P1. Furthermore, the position of the Ser216-containing loop 216−220 (136−140) could not be detected by a crystal structure analysis, thus demonstrating that this region is highly flexible. The arrangement of the most conserved residue in the S1 site, His213(133), in the aforementioned three proteins is identical [59]. This situation probably does not describe the state of PRRSV-Nsp4 in the solution but is an artifact of free-enzyme crystallization.

The importance of the His213 and Thr190 residues for the functioning of EAV-Nsp4 was confirmed using site-directed mutagenesis experiments. It was demonstrated by modifying the catalytic triad residues that processing of the polyprotein involving cleavage of the bonds after Glu residues depends on the activity of EAV-Nsp4. The modifications His213(1198)→Lys/Arg/Tyr also terminated the processing. The same effect was observed with the Thr190(1179)→Asp substitution; however, the mutations Thr190(1179)→Ser/Gly only

slightly reduced the processing efficiency [58]. In combination with the data obtained using the Glu−SGP model, these results demonstrate the fundamental significance of His213 and the considerably smaller role of the residues 190 and 216 for the hydrolysis of specific substrates by GEPases. Meanwhile, it still remains unclear whether His213 is a key element in the recognition of the charged substrate and what contribution to the formation of substrate specificity is made by Thr/Ser190 and Ser216. An analysis of the structures of other viral GEPases will shed more light on some of these questions.

Sesbania mosaic virus protease (SeMV-pro) has a 3D structure typical of CLPs (PDB ID − 1zyo) that is more similar to those of cellular (in particular, Glu−SGP) rather than viral representatives of this family (*Fig. 3C*) [61]. The protease carries the conventional catalytic triad; modification of its residues terminates polyprotein processing [62]. Similar to all GEPases, the conserved residues His213(298) and Thr190(279) are maintained, located within the S1 site of the enzyme. However, position 216(301) is occupied by a large hydrophobic residue, Phe (*Fig. 4C*). Superimposition of the 3D structures of SeMV-pro and Glu-SGP complexed with the tetrapeptide product of proteolysis of *tert*-butyloxycarbonyl-Ala-Ala-Pro-Glu (Boc-AAPE) demonstrates that the side chain of the Glu−P1 residue fits well the S1 pocket of viral protease. In order for the volume of the S1 pocket to be retained if there is a residual with a bulky side chain, the main protein chain needs to be significantly shifted in the 214(299)−223(308) region and the resulting space needs to be filled with the side chain of the Asp223(308) residue that is involved in the formation of the bottom of the S1 pocket, but apparently does not directly interact with Glu-P1 (*Fig. 4C*). This situation demonstrates that Ser 216 and the hydrogen bond between residue 216 and the γ-carboxyl group of Glu−P1 play no role in ensuring glutamate specificity. Unfortunately, no experiments involving the modification of Phe216(301) within SeMV−pro have been carried out. Meanwhile, the substitutions of His213(298) and Thr190(279) for Ala, but not the Asp223(308)→Ala mutation, completely inhibit the processing *in cis* of the SeMV-pro/VPg fusion protein (VPg being the viral protein following SeMV-pro in the polyprotein) in the model system [61].

The substrate specificity of human astrovirus protease (HAstV-pro) has been poorly studied. There is a lack of consistency in the data on the processing sites of viral polyprotein performed by this enzyme [63]. Meanwhile, it was demonstrated by using a recombinant enzyme and a series of synthetic substrate *in vitro* that HAstV-pro cleaves only the bonds formed by the α-carboxyl groups of Glu and Asp [24]. The spatial
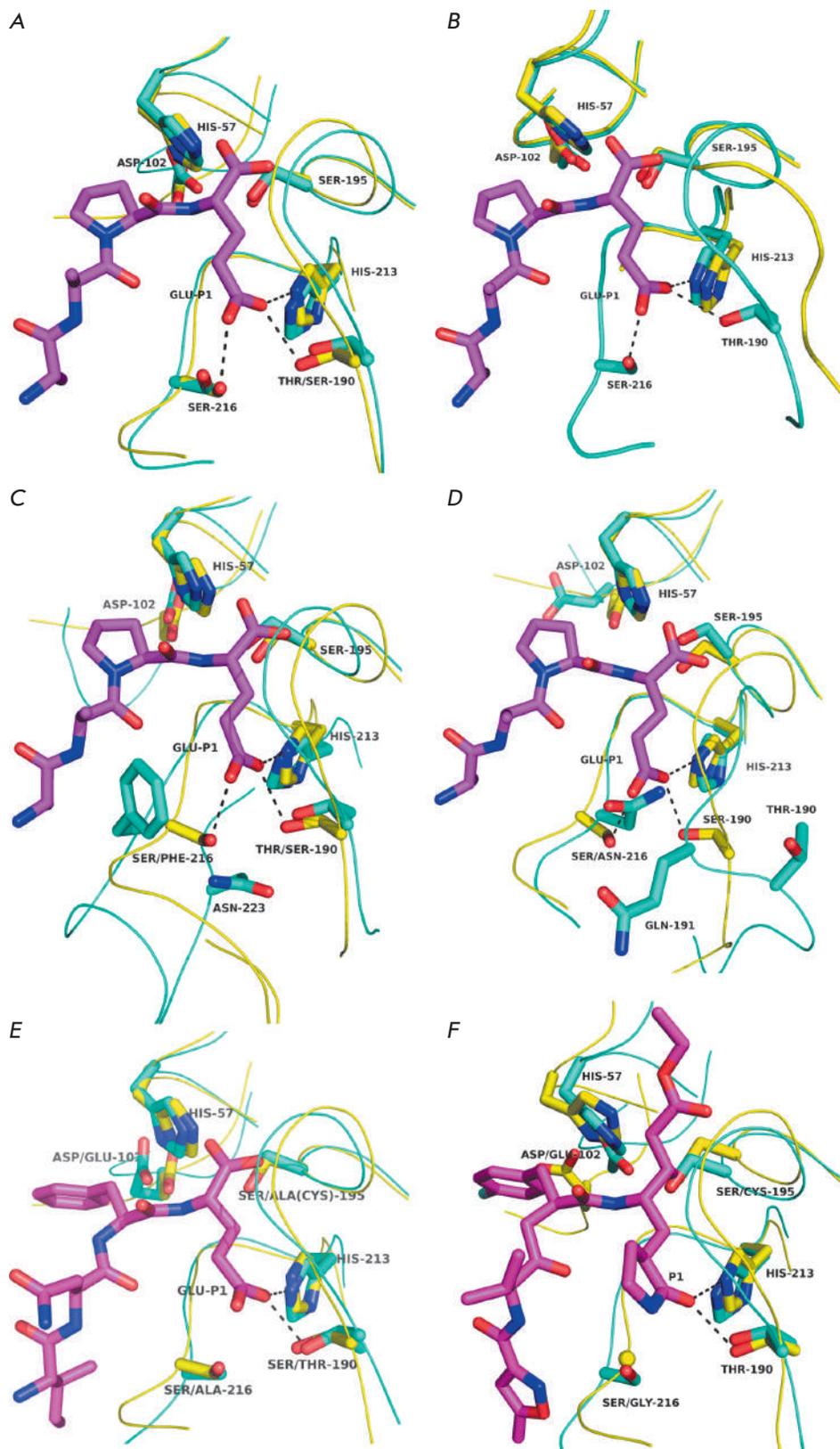
Fig. 4. S1 sites of viral Glu–
and Gln–specific proteases.
A – EAV-nsp4 (1mbm; cyan)
and Glu-SGP (1hpg; yellow)
complexed with Boc-AAPE
(magenta; the structure of
the protecting group is not
shown). B – EAV-nsp4 (cyan)
and PRRSV-nsp4 (3fan; yel-
low); the Boc-AAPE from the
Glu-SGP structure (magenta)
is inserted into the S1 site. C –
Glu-SGP (yellow) complexed
with Boc-AAPE (magenta)
and SeMV-pro (1zyo; cyan).
D – Glu-SGP (yellow) com-
plexed with Boc-AAPE (ma-
genta) and HAstV-pro (2w5e;
cyan). E – Glu-SGP (yellow)
and Gln / Glu-specific Norwalk
virus protease (4in1; cyan)
with Cys195→Ala substitution
complexed with tetrapeptide
Ile-Asn-Phe-Glu (magenta).
F – EAV-nsp4 (cyan) and
Gln-specific human rhinovirus
3C protease (1cqq, yellow)
complexed with inhibitor
AG7088 (magenta). Dashed
lines represent hydrogen
bonds.

structure of HAstV-pro (PDB ID − 2w5e) is generally similar to that of SeMV-pro (*Fig. 3D*) but has a number of specific features. Hence, the Asp102 residue (489 in polyprotein) of the catalytic triad that also contains Ser195(551) and His57(461) possesses a noncanonical conformation [24].

The structure of the S1 site also noticeably differs from the ones discussed above. Despite the fact that the His213 residue and its position are invariant, Ser at position 216 is substituted by Asn216(569), whose amide group actually occupies the place of the γ-carboxyl group of the substrate Glu-P1 as demonstrated by the superposition of the HAstV-pro structure and Glu-SGP complexed with the ligand (*Fig. 4D*). This significantly reduces the S1 pocket [24], whose volume does not match the Glu side radical. Furthermore, the conformation of the main-chain region 189−193 (545−549) differs from that in most CLP; thereafter, the conserved Thr190 residue lies far from the S1 site and is turned sideways. The position of the region 189−193 resembles the configuration of this region in PRRSV-Nsp4. Taking into account these differences from the structures of other GEPases and CLPs, it is rather arduous to draw any specific conclusions regarding the interactions between HAstV-pro and the P1 residue of the substrate.

Having summarized the data on viral GEPases and Glu−SGP, one can draw a conclusion that His213 is the shared element of the S1 pocket, while its modification causes enzyme inactivation in most cases. This residue can be positively charged; therefore, it is regarded as a candidate for being the key structural element that determines the substrate preferences of Glu-specific proteases. The Thr/Ser190 residue is also conserved in all GEPases, but its modification does not result in a loss of specific activity by the enzymes and possibly does not play any crucial role in the recognition of the Glu-P1 residue of the substrate. Finally, the nature of residue 216 is unessential in ensuring substrate specificity. As a result, GEPases carry residues with strongly different properties, Ser, Asn and Phe, at these positions. Additional information on the structural determinants of the substrate specificity of GEPases can be obtained by analyzing the viral 3C and 3C-like proteases that exhibit specificity to Gln at position P1.

Comparison of the primary and spatial structures of GEPases and 3C/3CLpro shows the similarity between their S1 sites (*Figs. 3E,F*). First, all 3C/3CLpro, identically to GEPases, contain the conserved His213 residue [64−76], whose modification results in enzyme inactivation [77−79]. This fact allows one to infer that this residue is not the key determinant of recognition of the substrate charge but is fundamental in ensuring a correct geometry of the S1 site. Second, most 3C/3CLpros retain the Thr/Ser190 residue that is typical of GEPases [58], thus confirming the conclusion that it is crucial for the formation of an adequate geometry of the S1 pocket rather than for charge recognition. The third element of the S1 site of GEPases at position 216, in 3C/3CLpro, is typically replaced with Gly (*Fig. 4F*) and sometimes Ala (*Fig. 4E*) residues, which have not been found in the known GEPases. The latter fact provides grounds for speculation about the involvement of Ser216 in the compensation for the substrate charge in GEPases [60]. However, mutagenesis in the Glu−SGP model shows that the Ser216→Ala/Gly substitution does not make the substrates with Gln−P1 the preferred ones, although it increases efficiency in their hydrolysis [47]. Furthermore, the data on GEPases with Phe/Asn216 residues that have been discussed do not support these assumptions. It is worth mentioning another hypothesis that still remains unverified. Since all GEPases are serine proteases, while Gln-specific enzymes are cysteine proteases, it is fair to assume that the difference in their substrate specificity depends on catalytic residues.

Hence, none of the detected conserved structural elements of the S1 site of Glu-SGP and viral 3CLSP seems to determine the preference of these enzymes for the Glu residue at the P1 position of the substrate. Therefore, this specificity of the GEPases of viruses and *Streptomyces* is ensured by structural determinants that do not directly reside in the substrate-binding site. However, the conventional research method combining the 3D structure analysis, site-directed mutagenesis, and studying the catalytic properties of enzymes has not identified these determinants yet. Studies focused on bacterial GEPases seem more successful.

STAPHYLOCOCCAL EPIDERMOLYTIC TOXINS

Staphylococci produce two types of GEPases: enzymes similar to V8 protease from *Staphylococcus aureus* (Glu-V8), which will be discussed below, and epidermolytic toxins (ETs). ETs are the key virulence factors responsible for the development of bullous impetigo and its generalized form, staphylococcal scalded skin syndrome, as well as similar animal diseases [80, 81]. The biological activity of ETs is associated with their ability to cleave with high specificity the Glu381-Gly bond in desmoglein 1, the desmosomal protein of cadherin type that mediates intercellular contacts (see more details in review [80]). In addition, ETs cleave the ester bonds formed by the carboxyl groups of Glu residues *in vitro* [82].

The spatial structures of epidermolytic toxins A [83, 84] and B [85] from *S. aureus* demonstrate that ETs belong to the CLPs family (*Fig. 5*). Meanwhile, these proteins exhibit unique features, the N-terminal α-helix
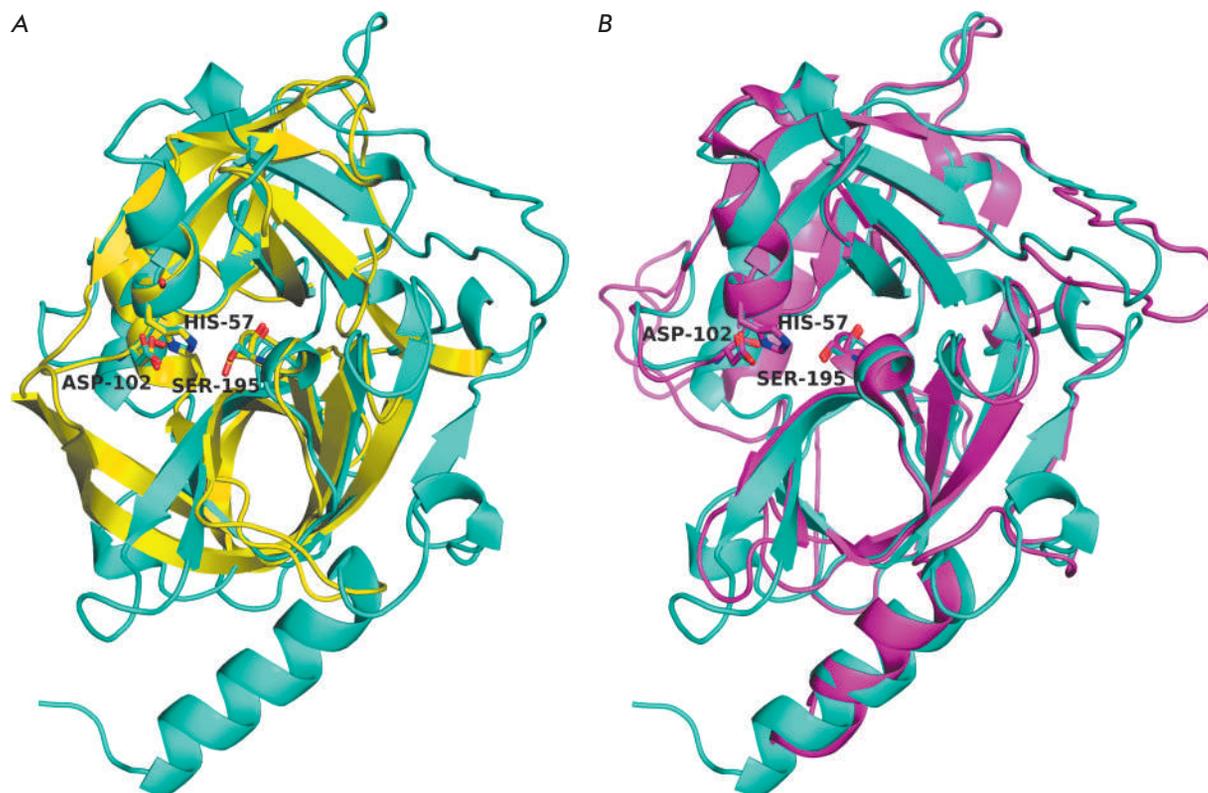
Fig. 5. Three-dimensional structures of the epidermolytic toxins of *S. aureus*. A — ETA (1agj; cyan) and Glu-SGP (1hpg; yellow). B — ETA (cyan) and ETB (1qtf; magenta). The catalytic triad residues are designated.

being one of them. The second feature consists in the unusual position of the residues forming the oxyanion hole: the Pro/Val192−Gly193 peptide bond (chymotrypsin numbering being used) is rotated 180° compared to other CLPs. As a result, a hydrogen bond is formed between the carbonyl oxygen of residue 192 and the hydroxyl group of catalytic Ser195 that seems to impede the manifestation of activity. After a structural analysis, a hypothesis has been put forward that binding between ET and the substrate (or a receptor) that the N-terminal α-helix is involved in results in a rearrangement of the active site and enzyme activation [83].

Identically to all the GEPases discussed above, the S1 pockets of ETs contain three key elements, two of which are the conserved His213 and Thr190 residues (*Fig. 6*). The third key element, as it has been predicted by simulation of the 3D structures [46], is Lys at position 216, which is an ideal candidate for compensating for the negative charge of Glu-P1 (Ser being typically found at this position in other GEPases). The Lys residue is conserved in most ETs from *S. aureus* and *S. hyicus*, while ExhA (an ET isolated from *S. hyicus*) contains Arg at position 216 [86, 87]. The significance of

Lys216 for the hydrolysis of substrates containing the Glu residue has been confirmed by site-directed mutagenesis experiments performed for the ETA model [88]. Any of the Lys216→Ala/Glu/Thr substitutions, identically to mutations in residues of the catalytic triad, resulted in a loss of the protein's ability to cleave N-Boc-*L*-glutamic acid α-phenyl ester and loss of epidermolytic activity.

Hence, in the case of ET, the positively charged residue that probably compensates for the substrate charge was detected directly in the S1 site, at position 216, which is important for substrate recognition by all CLPs. This compensator is critical for exhibiting enzymatic activity by ET. Meanwhile, there is no direct evidence yet that Lys/Arg216 in ET is responsible for glutamate specificity. The S1 sites of ET, except for Lys216, are very similar to the corresponding regions of GEPases from viruses and *Streptomyces* (*Fig. 6*). However, the findings presented above demonstrate that residue 216 is not significant in ensuring the substrate specificity of these enzymes. It should be inferred that different GEPase groups have different substrate recognition mechanisms. The standard charge compensator in the S1 pocket is the key structural element in
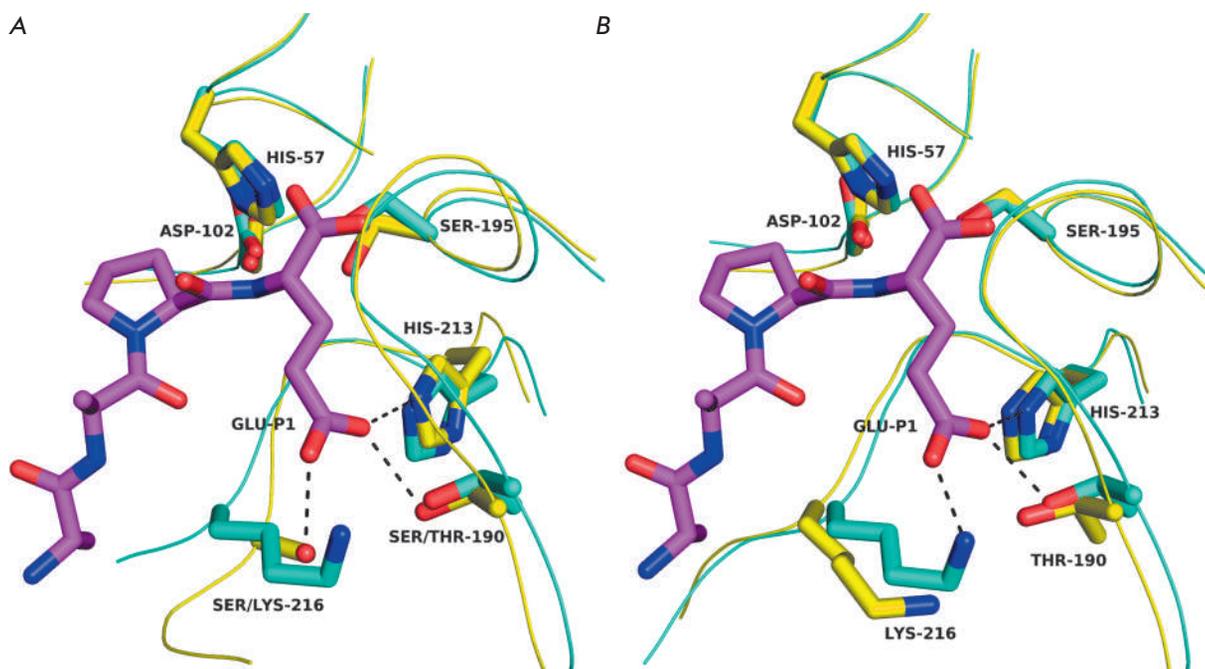
Fig. 6. S1 sites of the epidermolytic toxins of *S. aureus*. A – ETA (1agj; cyan) and Glu-SGP (1hpg; yellow) complexed with Boc-AAPE (magenta; the structure of the protecting group is not shown). B – ETA (cyan) and ETB (1qtf, yellow); the Boc-AAPE from the Glu-SGP structure (magenta) is inserted into the S1 site. Dashed lines represent hydrogen bonds.

ETs; in enzymes from viruses and *Streptomyces*, it is some other remote structural element. This conclusion has been supported by the data obtained for other bacterial GEPases.

OTHER BACTERIAL GLUTAMYL ENDOPEPTIDASES

In addition to GEPases from *Streptomyces* and ETs, a number of proteases secreted by gram-positive bacteria and possessing common structural features have been characterized. The simulation of the 3D structures of enzymes belonging to this group (Glu-V8, GEPases from *Bacillus licheniformis* and *B. subtilis*) conducted at early stages of the study of GEPases produced the assumption that compensation of the substrate charge in all three proteins is ensured by the α-amino group of residue 1 in the mature enzyme [46]. Localization of the N-terminal residue in the S1 site of GEPases of this group was verified later by experimental data on the tertiary structures of GEPase from *B. intermedius* (BIGEP) [89], Glu-V8 [90], and extracellular serine protease from *S. epidermidis* (Esp) [91].

The proteins under discussion possess high structural similarity with each other and with staphylococcal ETs; their structure is typical of CLPs. Their molecules consist of two β-domains separated by a deep cleft containing the active site (*Fig. 7*). The general architecture of the S1 sites in Glu-V8, BIGEP, and Esp is similar to

that of analogous regions in other GEPases and contain the mandatory elements: His213 and Ser/Thr190 (*Figs. 8A,B*). Meanwhile, the Gly residue is located in the third key position of the S1 pocket, which is a feature of viral Gln-specific 3C− and 3CLpro as discussed above. However, the absence of residue 216 side radical that can form a hydrogen bond with the carboxyl group oxygen of the substrate is compensated for, as predicted earlier, by the α-amino group of Val1, which occupies a position corresponding to that of the ε-amino group of the Lys216 residue in ET (*Fig. 8C*). Hence, a unique situation seems to take place for Glu-V8, BIGEP, and Esp, when protease specificity is determined by the N-terminus of the polypeptide chain. The originality of this "design concept" consists in the fact that Glu-V8, BIGEP, and Esp are synthesized by the cell as precursors that involve the signal peptide and propeptide, in addition to the catalytic domain. Hence, the N-terminus of a mature protein and, therefore, the S1 pocket are formed only after processing. This situation resembles the mechanism of activation of mammalian CLPs: after the propeptide was removed, the N-terminal NH_2-group of the mature protein formed a salt bridge with the Asp194 residue, thus triggering structural rearrangements in the enzyme molecule that result in its activation due to the formation of a proper structure of the S1 site and an oxyanion hole [92−96].

A

B

C

D
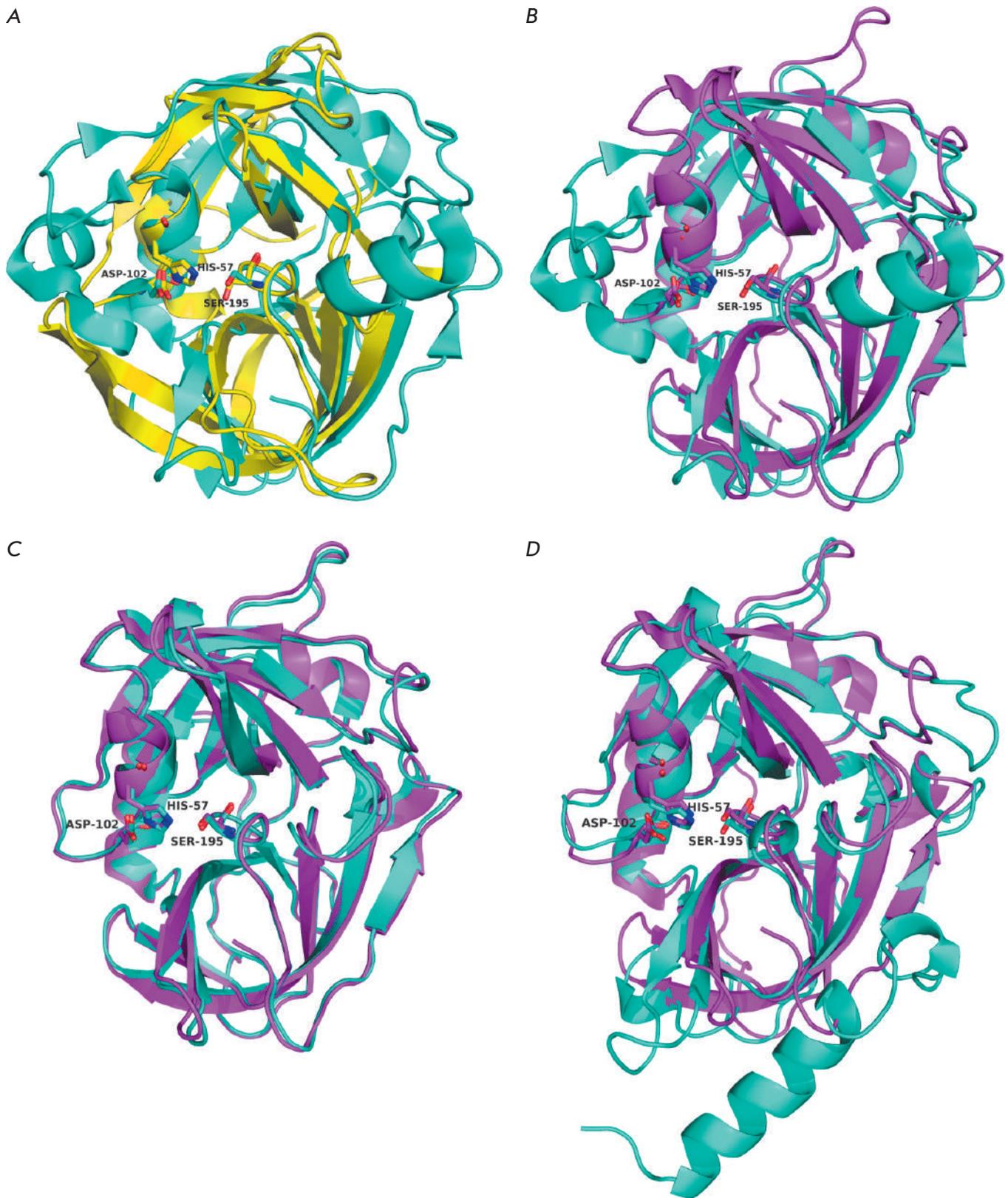
Fig. 7. Three-dimensional structures of bacterial glutamyl endopeptidases. A – BIGEP (1p3c, cyan) and Glu-SGP (1hpg; yellow). B – BIGEP (cyan) and Glu-V8 (1qy6; magenta). C – Glu-V8 (magenta) and Esp (4jcn; cyan). D – Glu-V8 (magenta) and ETA *S. aureus* (1agj; cyan). The catalytic triad residues are designated.

Site-directed modification of the residues in the S1 sites of BIGEP and Glu-V8 provided interesting results. First of all, the GEPase variant with a modification of the His213 residue was studied for the first time. Mutations of this type had been inserted earlier [47, 58], but no proteins were obtained. It was demonstrated that BIGEP with the His213(186 in BIGEP)→Thr substitution does not alter substrate preference and cleaves the protein substrate only after Glu residues. Meanwhile, modification significantly affects the catalysis effectiveness (the k_{cat} decreases more than 600-fold) but has a relatively low impact on substrate binding (the K_M increases approximately fivefold) [97]. Interestingly, a similar effect is also observed when the substrates containing the Asp residue at the P1 position are cleaved by native GEPases: the K_M increases approximately sixfold, while the k_{cat} declines by the same order of magnitude (~150-fold) [98]. These findings allow one to conclude that the conserved His213 residue is not the key element that determines the recognition of the negative charge of the substrate by GEPases but seems to be significant for accurate positioning of the cleaved bond with respect to the nucleophile (oxygen of the hydroxyl group of Ser195). This conclusion is consistent with the fact that His213 is the common structural element for Glu- and Gln-specific proteases.

The data on the role of the N-terminal residue in the functioning of GEPases were obtained for the Glu−V8 model. The substitution of the N-terminal Val for Leu/Ala/Phe/Gly/Ser was shown to reduce the efficiency of hydrolysis of the substrates carrying the Glu residue approximately 3-, 20-, 50-, 100-, and 200-fold, respectively [9, 99]. The more properties of the residue are similar to those of Val, the smaller the decrease in activity is. This result indicates that residue 1 is important for enzyme function and can be explained by the fact that deviations of the position of the α-amino group of this residue from the optimal position are different in mutants. Furthermore, Glu-V8 variants with additional amino acid residues, propeptide fragments, at the N-terminus, have been successfully obtained. Insertion of additional residues (from 1 to 39) in all cases significantly reduced enzymatic activity in hydrolyzing the substrates containing Glu-P1 [9, 100] but had a smaller impact on the efficiency of hydrolysis of similar substrates carrying Gln-P1. The mutants maintained their preference for substrates containing Glu-P1; cleavage efficiency was 10−20 times higher [100]. Hence, the α-amino group of the N-terminal Val residue probably makes a very significant contribution to the recognition of the charged substrate by the bacterial GEPases under discussion but is not fully responsible for enzyme specificity.

Summarizing all the available data regarding GEPases, a conclusion can be drawn about the dif-
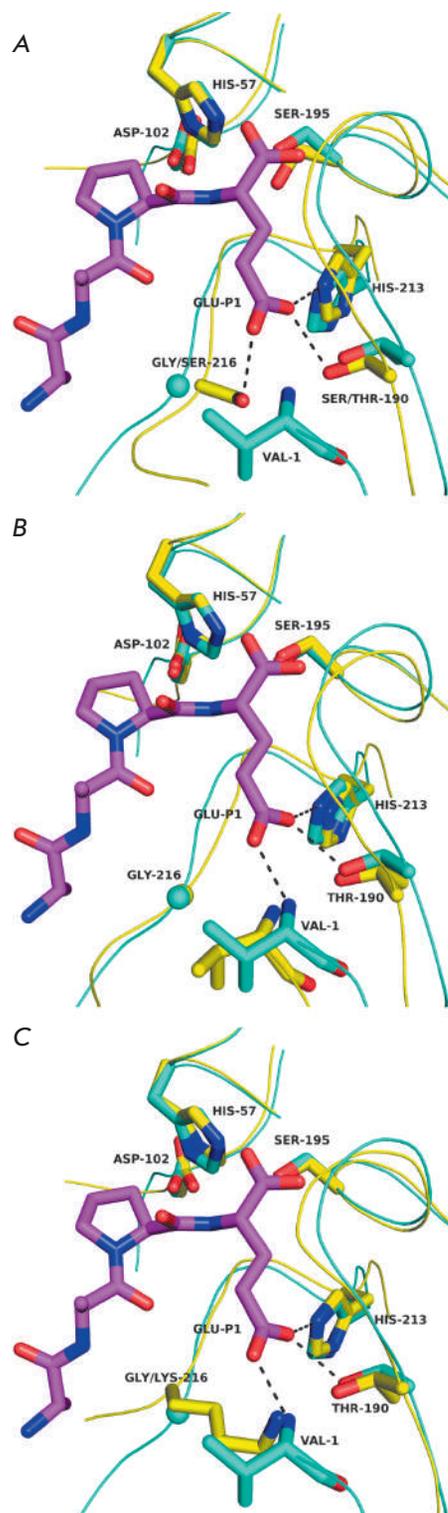
Fig. 8. S1 sites of bacterial glutamyl endopeptidases. A − BIGEP (1p3c, cyan) and Glu-SGP (1hpg; yellow) complexed with Boc-AAPE (magenta; the structure of the protecting group is not shown). B − BIGEP (cyan) and Glu-V8 (1qy6, yellow). C − BIGEP (cyan) and ETA (1agj, yellow). In B and C, the Boc-AAPE from the Glu-SGP structure (magenta) is inserted into the S1 site. Dashed lines represent hydrogen bonds.

ferences in the mechanisms of recognition of charged substrates by enzymes belonging to various groups. This indicates that GEPase branches have appeared several times in the evolutionary tree of CLPs, probably on the basis of the fundamental structure of the S1 pocket that is equally suitable for ensuring both glutamate and glutamine specificities and is most similar to the structure of the S1 regions of viral enzymes. The necessity for several structural variants of specificity optimization apparently is supposed to be caused by the differences in the functions of proteases belonging to different groups. An analysis of the published data on GEPases reveals that variations in the structure of the S1 sites in these enzymes correlate with the differences in the maturation mechanisms of their precursors. This observation allows one to put forward a hypothesis that the charge-compensation method depends on the maturation mechanism of the precursor protein.

STRUCTURAL DETERMINANTS OF SUBSTRATE SPECIFICITY AND MATURATION OF GLUTAMYL ENDOPEPTIDASE PRECURSORS

All GEPases are synthesized as precursors. However, the enzyme processing mechanisms significantly differ and can be subdivided into three groups. GEPases from *Streptomyces* and viruses are processed autocatalytically [47, 51]. ET precursors contain only a secretory leader [3] and, therefore, are processed by signal peptidase. For bacterial GEPases similar to Glu-V8 and BIGEP, propeptide is removed heterocatalytically by different proteases [22, 100−104], with just one exception [21]. Comparison of the structures of the S1 sites of GEPases and the processing mechanisms shows that no explicit substrate charge compensator is revealed in the S1 pocket in autoactivated enzymes; the S1 site of ET is characterized by the presence of the Lys216 residue, while the GEPases similar to Glu−V8 and BIGEP processed heterocatalytically contain an α-amino group of the N-terminal residue. Let us discuss these matches in the context of the biological functions of proteases belonging to each group.

Viral GEPases are synthesized as part of the long polyprotein, its selective hydrolysis being the main function of these enzymes [25, 51, 105, 106]. Hence, viral GEPases function inside the cell and start acting immediately after the polyprotein is synthesized. Therefore, the active site of the enzyme, including the specificity-determining regions, needs to form and be able to perform high-specificity hydrolysis already as part of the precursor protein, maintaining its structure after processing. The function of GEPases from *Streptomyces* appears to be fundamentally different. These extracellular enzymes are synthesized as conventional protease precursors carrying prepropeptide. The func-

tions of the prosequences of GEPases from *Streptomyces* are yet to be elucidated; however, one can assume that propeptides ensure the kinetic stability of mature molecules and partake in their secretion, by analogy with the closely related protease B from *Str. griseus* [107, 108]. Meanwhile, autocatalytic processing and the lack of a noticeable post-translational regulation of activity make this situation similar to that reported for viral enzymes: the active site needs to have completely formed within a precursor and maintained intact after a mature molecule has formed. In both cases, this problem seems to have one structural solution (*Fig. 4*). The S1 pocket does not have a direct charge compensator. The N-terminus is remote from the active site in the mature protein. Therefore, it does not partake in the formation of the S1 site as it is involved in processing. The structural elements responsible for glutamate specificity, which have not been identified yet, reside outside the S1 region and probably form before precursor processing. Hence, the structural elements that change during maturation are not involved in the formation of the molecule sites important for catalysis.

The opposite is observed for GEPases synthesized as preproprotein (e.g., Glu-V8). Not only are these proteases subjected to heteroactivation [101−103, 109], but they are also involved in regulatory activation cascades as it was demonstrated for Glu-V8 [110−112]. This implies that activity is strictly controlled via a rather complex and somewhat controversial mechanism. At first glance, the Glu−V8, BIGEP, and Esp precursors are supposed to be inactive, since the S1 site in these proteins is formed only in the mature molecule (*Figs. 7 and 8*). Meanwhile, data have been published demonstrating that the precursors of Glu-V8 [113], BIGEP [109], Esp [7], as well as GEPases from *B. licheniformis* [114], *B. subtilis* [102], and *Thermoactinomyces* sp. [21] are capable of autoprocessing; in most cases, it is the bonds corresponding to the specificity of the mature enzymes that are cleaved [7, 21, 109, 114]. Furthermore, glutamate activity *in trans* of precursor analogues has been detected [100]. These facts cast doubt on the mere possibility of regulating the activity of the proteases under discussion, although a closer look at the precursor activation mechanism demonstrates that the situation is more complex.

Autoprocessing (maybe intramolecular) of native enzymes that spontaneously occurs both *in vitro* and *in vivo* results in the formation of protein species with propeptide fragments usually 3−15 a.a.r. long rather than in complete deletion of the prosequence [7, 100, 109, 113, 114] that corresponds to the size of propeptides in mammalian CLPs. These species exhibit no activity with respect to protein substrates and low activity with respect to peptides *in trans* and can be
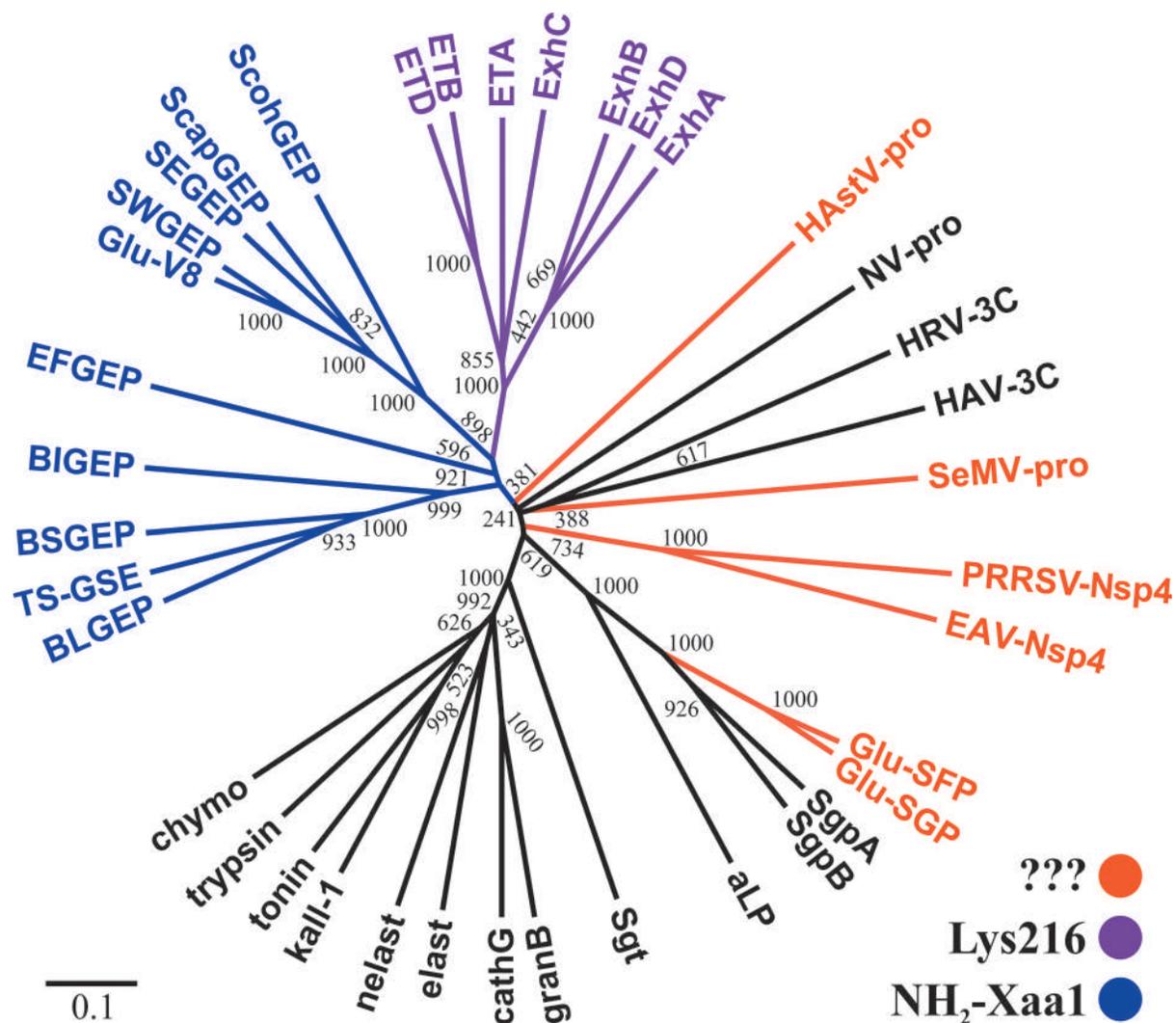
Fig. 9. Phylogenetic tree of chymotrypsin-like proteases. Branches corresponding to GEPases are colored: in orange − a compensator of the substrate charge at the enzyme S1 site has not been identified; magenta − Lys216 at the S1 site; and blue − α-amino group of the N-terminal residue at the S1 site. GEPases: Glu-SFP of *Str. fradiae*, SEGEP of *S. epidermidis*, SWGEP of *S. warneri*, ScohGEP of *S. cohnii*, ScapGEP of *S. caprae*, BLGEP of *B. licheniformis*, BSGEP of *B. subtilis*, TS-GSE of *Thermoactimomyces* sp., EFGEP of *Ent. faecalis*; ExhA, ExhB, ExhC and ExhD − epidermolytic toxins A, B, C and D of *S. hyicus*. NV-pro − Norwalk virus protease; HRV-3C, HAV-3C − proteases 3C of human rhinovirus and hepatitis A virus; aLP − α-lytic protease of *Lysobacter enzymogenes*; Sgt, SgpA, SgpB − trypsin, proteases A and B of *Str. griseus*; kall-1, trypsin, nelast, cathG, granB − human kallikrein 1, trypsin 1, neutrophil elastase, cathepsin G and granzyme B; chymo, elast − bovine chymotrypsin A and elastase 1; tonin − rat tonin. Sequence alignment and neighbor-joining tree reconstruction was carried out using ClustalX 2.1 (www.clustal.org). The tree was visualized with the use of FigTree software (tree.bio.ed.ac.uk /software/figtree/). The numbers represent the number of dendrograms in which the individual bifurcations were reproduced during bootstrap sampling of 1,000 trees.

activated only heterocatalytically [7, 100, 109, 113, 114]. To make the picture complete, we would like to add that data on the enzyme from *Thermoactinomyces* sp. carrying Glu1, which can be autoactivated *in vitro* in a heterologous expression system [21], identically to the previously artificially obtained mutants of other GE-Pases [109, 114], have been published recently. Hence,

maturation of enzymes similar to Glu−V8 is a stepwise process. These proteins seem to contain two propeptides. The first one is a long folding assistant [99, 109] that ensures the kinetic stability of a mature protein as often occurs in bacterial proteases [115]. The second propeptide, which is short and forms after the first processing step, is the activation unit [109, 113, 114] that

maintains the inactive state of the enzyme. Furthermore, it cannot be ruled out that the structure of the active site of proteases changes after the first propeptide portion is removed. It is fair to say that propeptides of the discussed group of GEPases simultaneously combine properties typical of the propeptides of bacterial CLPs and mammalian enzymes. Hence, the need for strict regulation of the activity of Glu-V8-like enzymes is satisfied through the formation of the S1 pocket only after the propeptide has been deleted. A mechanism similar to the activation mechanism of mammalian CLPs is used: involvement of the N-terminal amino group in the structure of the molecule elements essential for catalysis. However, the folding assistant is deleted autocatalytically due to the basic specificity of the enzymes.

Staphylococcal ETs are the intermediate variant. On the one hand, their precursors are processed heterocatalytically. On the other hand, processing is not related to activity regulation, since it only involves signal peptide deletion. Hence, neither the formation of a functionally active enzyme before processing nor strict activity regulation is required. The variant observed in GEPases from viruses and *Streptomyces* would be suitable here. However, a phylogenetic analysis demonstrates that ETs are most likely to be Glu−V8 paralogues (*Fig. 9*, see discussion below); i.e., these proteins are "engineered" on the same basis as Glu−V8 and employ essentially the same architecture of the S1 site (*Fig. 8*). Meanwhile, unlike Glu−V8, ETs contain no propeptides, being indicative of a different folding mechanism [115], and exhibit a much narrower specificity. They are inactive with respect to most proteins and peptides, which is possibly attained through inserting the Lys216 residue and reducing the volume of the S1 pocket, as well as due to the unusual conformation of the oxyanion hole [83].

In the context of our discussion, it would be interesting to trace the phylogeny of GEPases. The only attempt at a phylogenetic analysis of enzymes belonging to this group was found in a study published 20 years ago [46]. Therefore, in this review we compared the sequences of the characterized GEPases and some CLPs with different specificities in order to build a phylogenetic tree (*Fig. 9*). First, we would like to mention, as the authors of [46] did, that there is an impression that GEPases have appeared in the phylogenetic tree of GLPs at least twice. This is indicated by the presence of two remote branches of bacterial GEPases: one branch contains proteins similar to Glu-V8 and ET, while the second one corresponds to enzymes from *Streptomyces*. (The phylogenetic position of viral proteases is difficult to infer, since the topology of the resulting tree in the portion concerning these proteins is unreliable.)

It is especially illustrative that GEPases from *Streptomyces* are just a small sprout in the branch of bacterial proteases exhibiting broad specificity. This observation gives grounds for assuming that there is quite a high probability that glutamate specificity (actually, any other specificity) develops via the chymotrypsin fold. Modification of the key residues of the S1 pocket (His213, Thr/Ser190) that provide the required geometry and minimal interactions for the binding of Glu/Gln residues is apparently needed for that. However, this basic specificity probably needs to be enhanced, which can be achieved through different mechanisms, in particular by inserting a compensator into the S1 site. However, this is not the only possibility as demonstrated by the analysis of enzymes from viruses and *Streptomyces*. Special attention should be focused on the branch combining all bacterial GEPases, except for enzymes isolated from *Streptomyces*. As expected, the topology of this branch corresponds to the taxonomy of producer bacteria. ETs and staphylococcal enzymes, such as Glu-V8, share the phylogenetic tree's branch; i.e., they are structurally closer to each other than they are to the remaining bacterial GEPases.

CONCLUSIONS

Our analysis demonstrates that all known GEPases belong to the structural family of chymotrypsin and possess a similar overall structure of the S1 substrate-binding site. Enzymes in this group have several different systems of substrate charge compensation. The differences in the mechanisms of negative charge recognition correlate with the differences in the architecture and processing pathways of the precursors, which is probably determined by the biological functions of the corresponding proteases. All these facts provide grounds for assuming that GEPases have emerged in the phylogenetic tree of CLP at least twice. However, we have to admit that the data on the structure and mechanisms of action of GEPases available today are not sufficient to solve the puzzle of their strict substrate specificity.

It should be emphasized that the focus of studies devoted to GEPases shifts from the investigation of enzymes towards analyzing their biological functions, typically because of the pathogenesis. Thus, the involvement of staphylococcal GEPases in the regulation of biofilm growth is studied intensively today, primarily due to the hope of finding new strategies to combat staphylococcal infection [116]. Viral GEPases are being thoroughly studied in connection with attempts to design effective antiviral drugs. Meanwhile, GEPases are usually not isolated from the entire pool of 3C-like proteases in pursuit of universal inhibitors of the processing of viral polyproteins. Engineering of inhibitors

requires extensive investigation into protein–ligand interactions, which implies obtaining a large body of structural data (e.g., [76, 117]). The study of the role of GEPases in the viral life cycle is still underway [24, 118]. The recent studies devoted to viral 3C and 3C-like proteases, including GEPases, as apoptosis inductors deserve special mention [119–122]. Research into GEPases in the medical context will undoubtedly continue. It should be emphasized that, since strict substrate specificity underlies the biological activity of GEPases, novel data on its structural determinants will be inevitably collected during these studies. ●

REFERENCES

1. Demidiuk I.V., Kostrov S.V. // Mol Biol (Mosk). 1999. V. 33. № 1. P. 100–105.
2. Mil'gotina E.I., Voiushina T.L., Chestukhina G.G. // Bioorg Khim. 2003. V. 29. № 6. P. 563–576.
3. Dancer S.J., Garratt R., Saldanha J., Jhoti H., Evans R. // FEBS Lett. 1990. V. 268. № 1. P. 129–132.
4. Hanakawa Y., Schechter N.M., Lin C., Nishifuji K., Amagai M., Stanley J.R. // J. Biol. Chem. 2004. V. 279. № 7. P. 5268–5277.
5. Dubin G., Chmiel D., Mak P., Rakwalska M., Rzychon M., Dubin A. // Biol. Chem. 2001. V. 382. № 11. P. 1575–1582.
6. Moon J.L., Banbula A., Oleksy A., Mayo J.A., Travis J. // Biol. Chem. 2001. V. 382. № 7. P. 1095–1099.
7. Ohara-Nemoto Y., Ikeda Y., Kobayashi M., Sasaki M., Tajika S., Kimura S. // Microb. Pathog. 2002. V. 33. № 1. P. 33–41.
8. Yokoi K., Kakikawa M., Kimoto H., Watanabe K., Yasukawa H., Yamakawa A., Taketo A., Kodaira K.I. // Gene. 2001. V. 281. № 1–2. P. 115–122.
9. Ono T., Ohara-Nemoto Y., Shimoyama Y., Okawara H., Kobayakawa T., Baba T.T., Kimura S., Nemoto T.K. // Biol. Chem. 2010. V. 391. № 10. P. 1221–1232.
10. Fudaba Y., Nishifuji K., Andresen L.O., Yamaguchi T., Komatsuzawa H., Amagai M., Sugai M. // Microb. Pathog. 2005. V. 39. № 5–6. P. 171–176.
11. Nishifuji K., Fudaba Y., Yamaguchi T., Iwasaki T., Sugai M., Amagai M. // Vet. Dermatol. 2005. V. 16. № 5. P. 315–323.
12. Yoshida N., Tsuruyama S., Nagata K., Hirayama K., Noda K., Makisumi S. // J. Biochem. 1988. V. 104. № 3. P. 451–456.
13. Kitadokoro K., Nakamura E., Tamaki M., Horii T., Okamoto H., Shin M., Sato T., Fujiwara T., Tsuzuki H., Yoshida N., et al. // Biochim. Biophys. Acta. 1993. V. 1163. № 2. P. 149–157.
14. Khaidarova N.V., Rudenskaya G.N., Revina L.P., Stepanov V.M., Egorov N.S. // Biokhimiia. 1989. V. 54. № 1. P. 46–53.
15. Svendsen I., Breddam K. // Eur. J. Biochem. 1992. V. 204. № 1. P. 165–171.
16. Rufo G.A., Jr., Sullivan B.J., Sloma A., Pero J. // J. Bacteriol. 1990. V. 172. № 2. P. 1019–1023.
17. Leshchinskaya I.B., Shakirov E.V., Itskovitch E.L., Balaban N.P., Mardanova A.M., Sharipova M.R., Viryasov M.B., Rudenskaya G.N., Stepanov V.M. // FEBS Lett. 1997. V. 404. № 2–3. P. 241–244.
18. Rebrikov D.V., Akimkina T.V., Shevelev A.B., Demidyuk I.V., Bushueva A.M., Kostrov S.V., Chestukhina G.G., Stepanov V.M. // J. Protein Chem. 1999. V. 18. № 1. P. 21–27.
19. Mosolova J.V., Rudenskaya G.N., Stepanov V.M., Khodova O.M., Tsaplina I.A. // Biokhimiia. 1987. V. 52. № 3. P. 414–422.
20. Demidyuk I.V., Nosovskaya E.A., Tsaplina I.A., Karavaiko G.I., Kostrov S.V. // Biochemistry (Mosc). 1997. V. 62. № 2. P. 171–175.
21. Liu F., Zhao Z.S., Ren Y., Cheng G., Tang X.F., Tang B. // Appl. Microbiol. Biotechnol. 2016. V. 100. № 24. P. 10429–10441.
22. Kawalec M., Potempa J., Moon J.L., Travis J., Murray B.E. // J. Bacteriol. 2005. V. 187. № 1. P. 266–275.
23. Ziebuhr J., Snijder E.J., Gorbalenya A.E. // J. Gen. Virol. 2000. V. 81. № 4. P. 853–879.
24. Speroni S., Rohayem J., Nenci S., Bonivento D., Robel I., Barthel J., Luzhkov V.B., Coutard B., Canard B., Mattevi A. // J. Mol. Biol. 2009. V. 387. № 5. P. 1137–1152.
25. Somera M., Sarmiento C., Truve E. // Viruses. 2015. V. 7. № 6. P. 3076–3115.
26. Schechter I., Berger A. // Biochem. Biophys. Res. Commun. 1967. V. 27. № 2. P. 157–162.
27. Stroud R.M. // Sci. Am. 1974. V. 231. № 1. P. 74–88.
28. Rudenskaia G.N. // Bioorg Khim. 2003. V. 29. № 2. P. 117–128.
29. Zamolodchikova T.S., Sokolova E.A., Alexandrov S.L., Mikhaleva, I.I., Prudchenko I.A., Morozov I.A., Kononenko N.V., Mirgorodskaya O.A., Da U., Larionova N.I., et al. // Eur. J. Biochem. 1997. V. 249. № 2. P. 612–621.
30. Polanowska J., Krokoszynska I., Czapinska H., Watorek W., Dadlez M., Otlewski J. // Biochim. Biophys. Acta. 1998. V. 1386. № 1. P. 189–198.
31. Poe M., Blake J.T., Boulton D.A., Gammon M., Sigal N.H., Wu J.K., Zweerink H.J. // J. Biol. Chem. 1991. V. 266. № 1. P. 98–103.
32. Perona J.J., Craik C.S. // Protein Sci. 1995. V. 4. № 3. P. 337–360.
33. Hedstrom L. // Chem. Rev. 2002. V. 102. № 12. P. 4501–4524.
34. Krieger M., Kay L.M., Stroud R.M. // J. Mol. Biol. 1974. V. 83. № 2. P. 209–230.
35. Waugh S.M., Harris J.L., Fletterick R., Craik C.S. // Nat. Struct. Biol. 2000. V. 7. № 9. P. 762–765.
36. Hof P., Mayr I., Huber R., Korzus E., Potempa J., Travis J., Powers J.C., Bode W. // EMBO J. 1996. V. 15. № 20. P. 5481–5491.
37. Tsu C.A., Perona J.J., Fletterick R.J., Craik C.S. // Biochemistry. 1997. V. 36. № 18. P. 5393–5401.
38. Pletnev V.Z., Zamolodchikova T.S., Pangborn W.A., Duax W.L. // Proteins. 2000. V. 41. № 1. P. 8–16.
39. Graf L., Jancso A., Szilagyi L., Hegyi G., Pinter K., Naray-Szabo G., Hepp J., Medzihradszky K., Rutter W.J. // Proc. Natl. Acad. Sci. USA. 1988. V. 85. № 14. P. 4961–4965.
40. Hedstrm L., Szilagyi L., Rutter W.J. // Science. 1992. V. 255. № 5049. P. 1249–1253.
41. Hedstrom L., Perona J.J., Rutter W.J. // Biochemistry. 1994. V. 33. № 29. P. 8757–8763.
42. Hedstrom L., Farr-Jones S., Kettner C.A., Rutter W.J. // Biochemistry. 1994. V. 33. № 29. P. 8764–8769.
43. Perona J.J., Hedstrom L., Rutter W.J., Fletterick R.J. // Biochemistry. 1995. V. 34. № 5. P. 1489–1499.

44. Nienaber V.L., Breddam K., Birktoft J.J. // Biochemistry. 1993. V. 32. № 43. P. 11469−11475.

45. Cortes A., Emery D.C., Halsall D.J., Jackson R.M., Clarke A.R., Holbrook J.J. // Protein Sci. 1992. V. 1. № 7. P. 892−901.

46. Barbosa J.A., Saldanha J.W., Garratt R.C. // Protein Eng. 1996. V. 9. № 7. P. 591−601.

47. Stennicke H.R., Birktoft J.J., Breddam K. // Protein Sci. 1996. V. 5. № 11. P. 2266−2275.

48. Dougherty W.G., Semler B.L. // Microbiol. Rev. 1993. V. 57. № 4. P. 781−822.

49. Gorbalenya A., Snijder E. // Perspectives Drug Discov. Design. 1996. V. 6. № 1. P. 64−86.

50. Spall V.E., Shanks M., Lomonossoff G.P. // Semin. Virol. 1997. V. 8. № 1. P. 15−23.

51. Gorbalenya A.E., Koonin E.V. // Sov. Sci. Rev. Sect. D Physicochem. Biol. 1993. V. 11. № 3. P. 1−84.

52. Blanck S., Stinn A., Tsiklauri L., Zirkel F., Junglen S., Ziebuhr J. // J. Virol. 2014. V. 88. № 23. P. 13747−13758.

53. Meyers G., Wirblich C., Thiel H.J., Thumfart J.O. // Virology. 2000. V. 276. № 2. P. 349−363.

54. Belliot G., Sosnovtsev S.V., Mitra T., Hammer C., Garfield M., Green K.Y. // J. Virol. 2003. V. 77. № 20. P. 10957−10974.

55. Sosnovtsev S.V., Belliot G., Chang K.O., Prikhodko V.G., Thackray L.B., Wobus C.E., Karst S.M., Virgin H.W., Green K.Y. // J. Virol. 2006. V. 80. № 16. P. 7816−7831.

56. Robel I., Gebhardt J., Mesters J.R., Gorbalenya A., Coutard B., Canard B., Hilgenfeld R., Rohayem J. // J. Virol. 2008. V. 82. № 16. P. 8085−8093.

57. Wei C., Meller J., Jiang X. // Virology. 2013. V. 436. № 1. P. 24−32.

58. Snijder E.J., Wassenaar A.L., van Dinten L.C., Spaan W.J., Gorbalenya A.E. // J. Biol. Chem. 1996. V. 271. № 9. P. 4864−4871.

59. Tian X., Lu G., Gao F., Peng H., Feng Y., Ma G., Bartlam M., Tian K., Yan J., Hilgenfeld R., et al. // J. Mol. Biol. 2009. V. 392. № 4. P. 977−993.

60. Barrette-Ng I.H., Ng K.K., Mark B.L., van Aken D., Cherney M.M., Garen C., Kolodenko Y., Gorbalenya A.E., Snijder E.J., James M.N. // J. Biol. Chem. 2002. V. 277. № 42. P. 39960−39966.

61. Gayathri P., Satheshkumar P.S., Prasad K., Nair S., Savithri H.S., Murthy M.R. // Virology. 2006. V. 346. № 2. P. 440−451.

62. Satheshkumar P.S., Lokesh G.L., Savithri H.S. // Virology. 2004. V. 318. № 1. P. 429−438.

63. Bosch A., Pinto R.M., Guix S. // Clin. Microbiol. Rev. 2014. V. 27. № 4. P. 1048−1074.

64. Matthews D.A., Smith W.W., Ferre R.A., Condon B., Budahazi G., Sisson W., Villafranca J.E., Janson C.A., McElroy H.E., Gribskov C.L., et al. // Cell. 1994. V. 77. № 5. P. 761−771.

65. Mosimann S.C., Cherney M.M., Sia S., Plotch S., James M.N. // J. Mol. Biol. 1997. V. 273. № 5. P. 1032−1047.

66. Bergmann E.M., Mosimann S.C., Chernaia M.M., Malcolm B.A., James M.N. // J. Virol. 1997. V. 71. № 3. P. 2436−2448.

67. Matthews D.A., Dragovich P.S., Webber S.E., Fuhrman S.A., Patick A.K., Zalman L.S., Hendrickson T.F., Love R.A., Prins T.J., Marakovits J.T., et al. // Proc. Natl. Acad. Sci. USA. 1999. V. 96. № 20. P. 11000−11007.

68. Anand K., Palm G.J., Mesters J.R., Siddell S.G., Ziebuhr J., Hilgenfeld R. // EMBO J. 2002. V. 21. № 13. P. 3213−3224.

69. Anand K., Ziebuhr J., Wadhwani P., Mesters J.R., Hilgenfeld R. // Science. 2003. V. 300. № 5626. P. 1763−1767.

70. Yang H., Yang M., Ding Y., Liu Y., Lou Z., Zhou Z., Sun L., Mo L., Ye S., Pang H., et al. // Proc. Natl. Acad. Sci. USA. 2003. V. 100. № 23. P. 13190−13195.

71. Birtley J.R., Knox S.R., Jaulent A.M., Brick P., Leatherbarrow R.J., Curry S. // J. Biol. Chem. 2005. V. 280. № 12. P. 11520−11527.

72. Nakamura K., Someya Y., Kumasaka T., Ueno G., Yamamoto M., Sato T., Takeda N., Miyamura T., Tanaka N. // J. Virol. 2005. V. 79. № 21. P. 13685−13693.

73. Zeitler C.E., Estes M.K., Venkataram Prasad B.V. // J. Virol. 2006. V. 80. № 10. P. 5050−5058.

74. Leen E.N., Baeza G., Curry S. // PLoS One. 2012. V. 7. № 6. P. e38723.

75. Muhaxhiri Z., Deng L., Shanker S., Sankaran B., Estes M.K., Palzkill T., Song Y., Prasad B.V. // J. Virol. 2013. V. 87. № 8. P. 4281−4292.

76. Damalanka V.C., Kim Y., Alliston K.R., Weerawarna P.M., Galasiti Kankanamalage A.C., Lushington G.H., Mehzabeen N., Battaile K.P., Lovell S., Chang K.O., et al. // J. Med. Chem. 2016. V. 59. № 5. P. 1899−1913.

77. Ziebuhr J., Heusipp G., Siddell S.G. // J. Virol. 1997. V. 71. № 5. P. 3992−3997.

78. Hegyi A., Friebe A., Gorbalenya A.E., Ziebuhr J. // J. Gen. Virol. 2002. V. 83. № 3. P. 581−593.

79. Someya Y., Takeda N., Miyamura T. // J. Virol. 2002. V. 76. № 12. P. 5949−5958.

80. Nishifuji K., Sugai M., Amagai M. // J. Dermatol. Sci. 2008. V. 49. № 1. P. 21−31.

81. Bukowski M., Wladyka B., Dubin G. // Toxins (Basel). 2010. V. 2. № 5. P. 1148−1165.

82. Bailey C.J., Redpath M.B. // Biochem. J. 1992. V. 284. № 1. P. 177−180.

83. Vath G.M., Earhart C.A., Rago J.V., Kim M.H., Bohach G.A., Schlievert P.M., Ohlendorf D.H. // Biochemistry. 1997. V. 36. № 7. P. 1559−1566.

84. Cavarelli J., Prevost G., Bourguet W., Moulinier L., Chevrier B., Delagoutte B., Bilwes A., Mourey L., Rifai S., Piemont Y., et al. // Structure. 1997. V. 5. № 6. P. 813−824.

85. Vath G.M., Earhart C.A., Monie D.D., Iandolo J.J., Schlievert P.M., Ohlendorf D.H. // Biochemistry. 1999. V. 38. № 32. P. 10239−10246.

86. Yamaguchi T., Nishifuji K., Sasaki M., Fudaba Y., Aepfelbacher M., Takata T., Ohara M., Komatsuzawa H., Amagai M., Sugai M. // Infect. Immun. 2002. V. 70. № 10. P. 5835−5845.

87. Ahrens P., Andresen L.O. // J. Bacteriol. 2004. V. 186. № 6. P. 1833−1837.

88. Rago J.V., Vath G.M., Bohach G.A., Ohlendorf D.H., Schlievert P.M. // J. Immunol. 2000. V. 164. № 4. P. 2207−2213.

89. Meijers R., Blagova E.V., Levdikov V.M., Rudenskaya G.N., Chestukhina G.G., Akimkina T.V., Kostrov S.V., Lamzin V.S., Kuranova I.P. // Biochemistry. 2004. V. 43. № 10. P. 2784−2791.

90. Prasad L., Leduc Y., Hayakawa K., Delbaere L.T. // Acta Crystallogr. D Biol. Crystallogr. 2004. V. 60. № 2. P. 256−259.

91. Chen C., Krishnan V., Macon K., Manne K., Narayana S.V., Schneewind O. // J. Biol. Chem. 2013. V. 288. № 41. P. 29440−29452.

92. Bode W. // J. Mol. Biol. 1979. V. 127. № 4. P. 357−374.

93. Bode W., Schwager P., Huber R. // J. Mol. Biol. 1978. V. 118. № 1. P. 99−112.

94. Wang D., Bode W., Huber R. // J. Mol. Biol. 1985. V. 185. № 3. P. 595−624.

95. Wroblowski B., Diaz J.F., Schlitter J., Engelborghs Y. // Protein Eng. 1997. V. 10. № 10. P. 1163−1174.

96. Gomis-Ruth F.X., Bayes A., Sotiropoulou G., Pampalakis G., Tsetsenis T., Villegas V., Aviles F.X., Coll M. // J. Biol. Chem. 2002. V. 277. № 30. P. 27273−27281.

97. Demidyuk I.V., Romanova D.V., Nosovskaya E.A., Ches-
tukhina G.G., Kuranova I.P., Kostrov S.V. // Protein Eng.
Des. Sel. 2004. V. 17. № 5. P. 411−416.

98. Breddam K., Meldal M. // Eur. J. Biochem. 1992. V. 206.
№ 1. P. 103−107.

99. Nemoto T.K., Ohara-Nemoto Y., Ono T., Kobayakawa
T., Shimoyama Y., Kimura S., Takagi T. // FEBS J. 2008.
V. 275. № 3. P. 573−587.

100. Rouf S.M., Ohara-Nemoto Y., Shimoyama Y., Kimura
S., Ono T., Nemoto T.K. // Indian J. Biochem. Biophys. 2012.
V. 49. № 6. P. 421−427.

101. Drapeau G.R. // J. Bacteriol. 1978. V. 136. № 2. P. 607−613.

102. Park C.H., Lee S.J., Lee S.G., Lee W.S., Byun S.M. // J.
Bacteriol. 2004. V. 186. № 19. P. 6457−6464.

103. Velishaeva N.S., Gasanov E.V., Gromova T.Iu., Demidiuk
I.V. // Bioorg Khim. 2008. V. 34. № 6. P. 786−791.

104. Demidyuk I.V., Gasanov E.V., Safina D.R., Kostrov S.V. //
Protein J. 2008. V. 27. № 6. P. 343−354.

105. Geigenmuller U., Chew T., Ginzton N., Matsui S.M. // J.
Virol. 2002. V. 76. № 4. P. 2003−2008.

106. Kiang D., Matsui S.M. // J. Gen. Virol. 2002. V. 83. № 1.
P. 25−34.

107. Truhlar S.M., Cunningham E.L., Agard D.A. // Protein
Sci. 2004. V. 13. № 2. P. 381−390.

108. Baardsnes J., Sidhu S., MacLeod A., Elliott J., Morden
D., Watson J., Borgford T. // J. Bacteriol. 1998. V. 180. № 12.
P. 3241−3244.

109. Gasanov E.V., Demidyuk I.V., Shubin A.V., Kozlovskiy
V.I., Leonova O.G., Kostrov S.V. // Protein Eng. Des. Sel.
2008. V. 21. № 11. P. 653−658.

110. Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.
// Infect. Immunol. 2001. V. 69. № 1. P. 159−169.

111. Massimi I., Park E., Rice K., Muller-Esterl W., Sau-
der D., McGavin M.J. // J. Biol. Chem. 2002. V. 277. № 44.
P. 41770−41777.

112. Shaw L., Golonka E., Potempa J., Foster S.J. // Microbi-
ology. 2004. V. 150. № 1. P. 217−228.

113. Nickerson N.N., Prasad L., Jacob L., Delbaere L.T.,
McGavin M.J. // J. Biol. Chem. 2007. V. 282. № 47. P. 34129−
34138.

114. Trachuk L.A., Shcheglov A.S., Milgotina E.I., Chestukhi-
na G.G. // Biochimie. 2005. V. 87. № 6. P. 529−537.

115. Demidyuk I.V., Shubin A.V., Gasanov E.V., Kostrov S.V. //
BioMolecular Concepts. 2010. V. 1. № 3−4. P. 305−322.

116. Iwase T., Uehara Y., Shinji H., Tajima A., Seo H., Takada
K., Agata T., Mizunoe Y. // Nature. 2010. V. 465. № 7296.
P. 346−349.

117. Kim Y., Lovell S., Tiew K.C., Mandadapu S.R., Alliston
K.R., Battaile K.P., Groutas W.C., Chang K.O. // J. Virol.
2012. V. 86. № 21. P. 11754−11762.

118. Nair S., Savithri H.S. // Virology. 2010. V. 396. № 1.
P. 106−117.

119. Calandria C., Irurzun A., Barco A., Carrasco L. // Virus
Res. 2004. V. 104. № 1. P. 39−49.

120. Chau D.H., Yuan J., Zhang H., Cheung P., Lim T., Liu Z.,
Sall A., Yang D. // Apoptosis. 2007. V. 12. № 3. P. 513−524.

121. Ma Z., Wang Y., Zhao H., Xu A.T., Wang Y., Tang J., Feng
W.H. // PLoS One. 2013. V. 8. № 7. P. e69387.

122. Shubin A.V., Demidyuk I.V., Lunina N.A., Komissarov
A.A., Roschina M.P., Leonova O.G., Kostrov S.V. // BMC
Cell Biol. 2015. V. 16. № 1. P. 4.

Huntington's Disease: Calcium Dyshomeostasis and Pathology Models

Y.A. Kolobkova, V.A. Vigont, A.V. Shalygin, E.V. Kaznacheyeva*

Institute of cytology of the Russian Academy of Sciences, Tikhoretsky ave. 4., Saint-Petersburg, 194064, Russia

*E-mail: evkazn@incras.ru

ABSTRACT Huntington's disease (HD) is a severe inherited neurodegenerative disorder characterized by motor dysfunction, cognitive decline, and mental impairment. At the molecular level, HD is caused by a mutation in the first exon of the gene encoding the huntingtin protein. The mutation results in an expanded polyglutamine tract at the N-terminus of the huntingtin protein, causing the neurodegenerative pathology. Calcium dyshomeostasis is believed to be one of the main causes of the disease, which underlies the great interest in the problem among experts in molecular physiology. Recent studies have focused on the development of animal and insect HD models, as well as patient-specific induced pluripotent stem cells (HD-iPSCs), to simulate the disease's progression. Despite a sesquicentennial history of HD studies, the issues of diagnosis and manifestation of the disease have remained topical. The present review addresses these issues.

KEYWORDS calcium, induced pluripotent stem cells (HD-iPSCs), huntington's disease, neurodegeneration, huntingtin, SOC.

ABBREVIATIONS HD − Huntington's disease; PM − plasma membrane; ER − endoplasmatic reticulum; [Ca^{2+}] − calcium concentration; (HD)iPSCs − (HD specific) induced pluripotent stem cells; InsP$_3$ − inositol-1,4,5-tris-phosphate; InsP$_3$R − inositol-1,4,5-trisphosphate receptor; (m)Htt − (mutant) huntingtin; Q − glutamine residue; MSNs − striatal medium spiny neurons; SOC(E) − store-operated calcium (entry).

INTRODUCTION

The inherited nature of Huntington's disease (HD) was discovered and described by George Huntington in his original paper almost a century and a half ago [1]. HD has an autosomal dominant type of inheritance and is caused by a mutation that leads to an increased number of CAG-repeats in the huntingtin (Htt) protein gene localized on chromosome 4p16.3. This mutation increases the number of glutamine (Q) residues in the N-terminal region of Htt, which, in different ways, leads to the observed pathologies [2]. Normally, the polyglutamine tract contains no more than 35 glutamines [3]. Huntington's disease is characterized by selective death of GABAergic striatal neurons [3], while dopaminergic neurons of the substance nigra are what are mainly affected in Parkinson's disease [4], and preferential loss of hippocampal neurons occurs in Alzheimer's disease [5]. To date, several mechanisms are believed to contribute to the pathogenesis of HD, including the new toxic properties of mutant Htt (mHtt), concomitantly with the dysfunction of normal Htt [6]. These changes lead to a dysregulation of the transcription of the gene encoding Htt [7], synaptic dysfunction and excitotoxicity [8, 9], mHtt dyshomeostasis [10], intracellular transport defects [11], mitochondrial dysfunction [12−14], and calcium signaling disturbances [15−17].

MANIFESTATION AND DIAGNOSIS OF HD

The prevalence of HD is quite high: the disease incidence rate is approximately 1 per 1,000,000 people of Asian and African descent and 5−10 per 100,000 Caucasians, besides the many people who are at risk. HD is more common in males than in females, manifests itself primarily at age older than 30 years, and usually leads to death 15−20 years after the onset of the first symptoms. At the same time, long polyglutamine tracts may be the cause of juvenile or even infantile HD. Mutations increasing the length of glutamine repeats up to 36−40Q are associated with incomplete penetrance; if repeats are longer than 41Q, the disease is fully penetrant [18].

The polyglutamine tract length of mHtt directly correlates with the disease's severity and in most cases inversely correlates with the age of onset of the first symptoms [19]. However, there is a significant variability between the expected and actual age of manifestations [20]. For example, for the same length of polyglutamine tract, especially in the range of 40−44Q, the

age of manifestations may differ by 20 years [21]. This difference may be explained by the presence of some genetic modifiers that regulate the expression of both Htt and other proteins and, thereby, mediate increased sensitivity or resistance to the disease. For example, the S18Y polymorphism in the gene encoding ubiquitin C-terminal hydrolase L1 is associated with late manifestations of HD [22]. In patients with the M441T mutation in the gene encoding the Htt-associated protein (Hap1), HD manifested itself at an earlier age due to a weakened interaction between Hap1 and mHtt and, thereby, increased Htt-mediated toxicity [23]. Recently, a single nucleotide polymorphism in the NF-κB binding site located in the *Htt* gene promoter was shown to reduce the promoter activity and, as a consequence, Htt expression, which led to late manifestations of HD [24].

However, a genetic mutation is not sufficient for both predicting the individual risk to a disease and assessing the current physiological processes in a body. Therefore, identification of biomarkers of HD progression, which may indicate pathological processes before the manifestation of clinical symptoms, is important for the development of new drugs and evaluation of treatment efficacy and the effect of environmental factors. Recent advances in the diagnosis include quantification of the mHtt level by a hypersensitive immunological analysis of single molecules in the cerebrospinal fluid samples of subjects with a mutation in the gene [25].

Because Htt is expressed in almost all body tissues, mHtt-induced changes can be detected even in the blood. Involvement of leukocytes in the immune response makes a blood test an ideal method to identify pathological processes, such as peripheral inflammation, in HD. In HD, expression of the *H2A histone family, member Y* gene is increased in the blood [26]. Clinical trials demonstrated that the expression of this gene both in the blood samples and brain tissues of HD patients was 1.6-fold higher than that in controls. Next-generation sequencing and Fluidigm technologies were used to identify five genes that encode the potential HD biomarkers detected in the blood of patients [27]. A correlation between cognitive impairment in HD and the levels of the peptide hormone prokineticin 2 (PROK2) involved in the regulation of circadian rhythms was revealed [28]. Therefore, PROK2 is considered as one of the promising markers of HD progression. Also, an elevated level of aquaporin 9 mRNA was detected in the blood of HD patients [29].

The variability of the clinical HD phenotype and the potential effect of some environmental and pharmacological factors lead to the need to combine different markers of HD progression. A decreased level of N-acetylaspartate (NAA) in brain tissues is considered a reliable indicator of neuronal dysfunction and death

and can be measured noninvasively by MRI, which is important for a clinical diagnosis [30]. The NAA level in patients with early HD manifestations is lower than that in a control group. At the same time, the level of a gliosis marker, myo-inositol, is significantly increased in these patients [31]. A relationship between the NAA level and the disease severity opens the opportunity to use this metabolite as an identifier of neurochemical reactions in evaluating the effectiveness of potential therapeutic agents.

In HD, there is an increase in the serum concentrations of vasopressin that play an important role in the homeostasis of body fluids [32], 8-hydroxy-2-deoxyguanosine (an indicator of oxidative DNA damage), and lipid peroxidation products (lactic acid, 4-hydroxynonenal, and malondialdehyde), which makes these compounds potential biomarkers [33]. Reduced levels of glutathione peroxidase and Cu,Zn-superoxide dismutase were detected in the erythrocytes of HD patients [34], and elevated levels of cytokines, including interleukins 4, 6, 8, 10, and 23, TNF-A, as well as clusterin, were found in postmortem brain sections and plasma samples [35].

The use of all these biomarkers will provide an accurate assessment of the efficacy of new treatments and increase the safety and efficacy of preclinical and clinical trials.

HUNTINGTIN PROTEIN

The development of HD is associated with a mutation in the huntingtin protein gene. Huntingtin is a protein with a molecular weight of about 350 kDa and a polyglutamine tract at the N-terminus. In the same region, there is a proline-rich domain involved in protein-protein interactions and protecting huntingtin from aggregation [36]. *Figure 1* presents the domain structure of human huntingtin.

In the cell, huntingtin functions as a scaffold protein; i.e., it provides colocalization of the proteins interacting with it, helping them to perform their functions. Huntingtin (especially its N-terminal region) interacts with numerous proteins, performing a wide variety of functions ranging from vesicular transport and endocytosis to the regulation of transcription and apoptosis [37].

The huntingtin molecule looks like a solenoid with a hydrophobic core composed of docked HEAT repeats. These repeats, together with the proline-rich region, participate in protein-protein interactions. The name HEAT is an acronym for four proteins in which the repeat structure was first identified (huntingtin, elongation factor 3, PR65/A (a phosphatase 2A subunit), and lipid kinase TOR) [38]. The structure of the short N-terminal fragments of huntingtin was studied by X-ray diffraction [39] and nuclear magnetic reso-
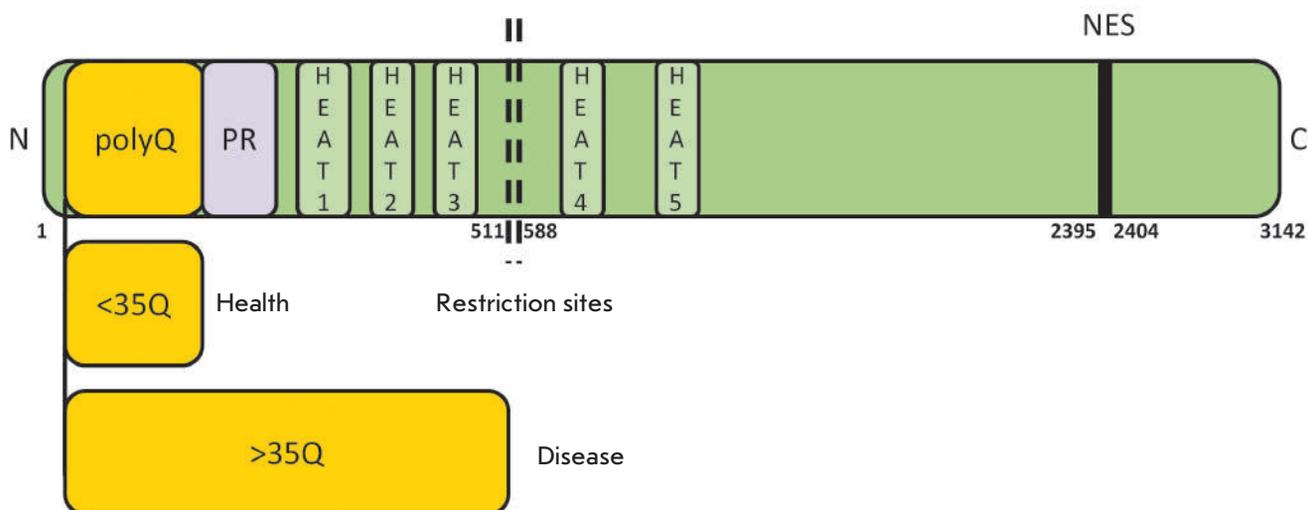
Fig. 1. The domain structure of human huntingtin: PolyQ — polyglutamine tract; PR — proline-rich domain responsible for protein-protein interactions; HEAT repeats; protease cleavage region; NES — nuclear export signal.

nance [40]. Recently, the secondary structure of huntingtin was shown to correlate with the length of the polyglutamine tract [41]. Images of normal and mutant huntingtins having a spherical structure with a cavity were obtained by electron microscopy [41]. Images of Htt23Q and Htt78Q are very similar, but the impact of the polyglutamine tract on the huntingtin structure suggests that huntingtin undergoes dramatic conformational changes upon interaction with its binding partners [41]. Despite these facts, it is not yet fully understood how the structure of huntingtin is related to its functions, and how mutation-induced changes in its structure lead to the observed pathologies.

HD is believed to be associated with cleavage of the N-terminal fragment from mutant huntingtin, which is encoded by the first exon and contains the polyQ tract. The cleaved fragment accumulates in the nucleus, while wild-type huntingtin is localized mainly in the cytosol [42, 43]. Posttranslational modifications of huntingtin control its localization [44]. Accumulation of aggregated N-terminal mHtt fragments and associated proteins, such as various transcription factors, heat shock proteins, and proteasome components, in the nucleus complicates their functioning and, as a consequence, leads to various cell pathologies [45].

Neuropathological markers of HD include intracellular inclusions formed by N-terminal mHtt fragments, which were found in a postmortem study of the brains of HD patients, as well as in animal and cell models of HD [42, 46, 47]. The formation of insoluble aggregates in HD leaves no doubt, but many studies have demonstrated that this process is not directly associated with neuronal degeneration. For example, expression of mHtt in a striatal neuron culture demonstrated an accumulation of insoluble protein aggregates, which did not correlate with neuronal death. Furthermore, a decrease in intranuclear inclusions of mHtt coincided with an aggravation of neurodegenerative processes [48]. A study of neurons expressing the first exon of *mHtt* also showed that neuronal death correlates with an increase in the polyglutamine tract length and with the amount of diffuse mHtt in the cell, while accumulation of aggregates just reduces the level of dissolved mHtt, thereby increasing the survival of neurons [49]. It is believed that unstable heterogeneous prefibrillar aggregates are responsible for amyloid toxicity, whereas mature fibrils are stable and harmless reservoirs of toxic species [50].

These facts suggest that the formation of aggregates in HD cannot be the sole cause of pathology development, and elucidating the molecular basis of HD remains a topical issue.

HD MODELING

Generation of adequate disease models is very important for studying the molecular mechanisms of neuro-

degeneration and searching for new drugs. Since HD is a hereditary disease caused by a mutation in a single gene, genetic manipulations can be used to create various models that accurately simulate the disease (*Fig. 2*).

R6/2 mice have a stable phenotype that includes impaired coordination and gait, hypoactivity, and cognitive dysfunction. The disease manifestation age in this model is about 4 weeks [51]. R6/2 mice were detected with aggregates containing intracellular inclusions similar to those found in the biopsy specimens of the brain tissues of HD patients [52]. However, despite the stability of the phenotype, R6/2 mice cannot be an accurate HD model, because they express only the N-terminal fragment of a mutant protein. Nevertheless, R6/2 mice are widely used to simulate the common features of polyglutamine diseases, including the abnormal protein conformation due to an expanded polyQ tract.

YAC128 and BACHD transgenic mice containing 128Q and 97Q in a full-length mutant protein, respec-

tively, have a milder HD phenotype compared to that of the R6/2 model [53].

Mouse knock-in models have the weakest HD phenotype. Even upon 150Q expression, HdhQ150/Q150 mice had fewer abnormalities than R6/2 mice. In HdhQ150/Q150 mice, the first disease symptoms, including motor dysfunction and gait disturbances, developed at a later age [54].

Despite the fact that mouse models are based on a disease-inducing mutation, most of them lack the stable neuronal loss that occurs in patients. To overcome this problem, other model organisms are required. Toxicity of the N-terminal mHtt fragment is more pronounced in large mammals, such as pigs and monkeys, while sheep expressing full-length mHtt lack marked phenotypic signs of the disease [55]. However, despite a number of advantages, these models have serious drawbacks, such as high cost and the need for specialized laboratory animal care equipment.
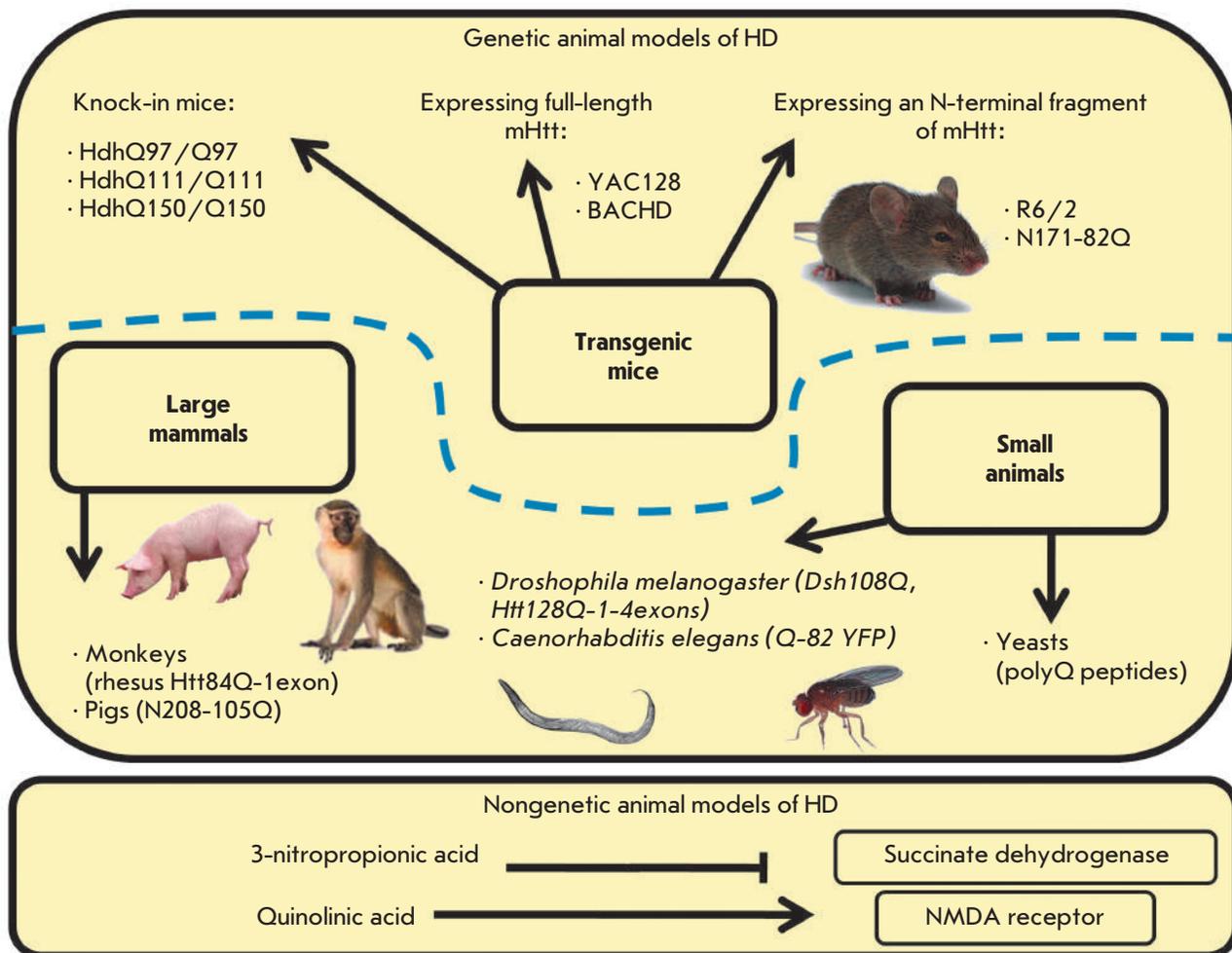
Fig. 2. Animal models of Huntington's disease
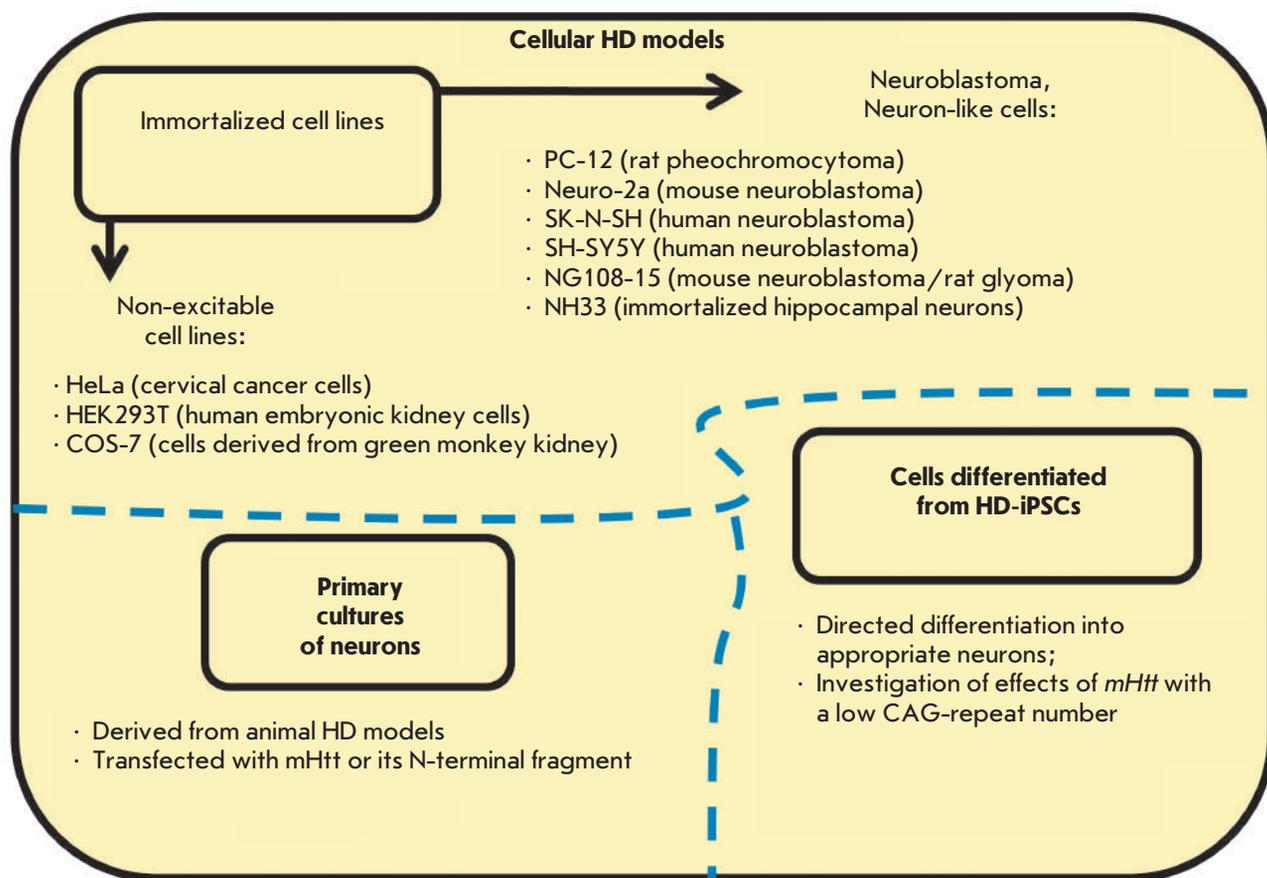
Fig. 3. Cellular models of Huntington's disease

Drosophila melanogaster and *Caenorhabditis elegans* are also used to model HD. The advantage of these organisms is a short lifespan and rapid reproduction. Identification of a human Htt ortholog in *Drosophila* suggests that these insects have the pathways necessary for the normal functioning of Htt, which makes *Drosophila* a good model for studying HD [56]. Another interesting feature of *Drosophila* as a HD model is easy visual evaluation of neurodegeneration. Overexpression of mHtt in *Drosophila* leads to the formation of aggregates, neuronal death, and decreased survival [57]. In addition, fly models of HD reproduce symptoms such as motor dysfunction and impairment of cognitive abilities and memory [58]. In the body wall muscle cells of *C. elegans* expressing the polyQ tract fused to the yellow fluorescent protein (YFP), the formation of aggregates, cellular toxicity, and paralysis were directly correlated with the age and the number of Q repeats [59]. Both mentioned models are actively used for testing potential drugs against HD. However, these models, which are based on species distantly related to humans, cannot fully reproduce the clinical picture observed in HD patients. For example, expression of Htt fragments containing polyglutamine tracts with 88Q or 128Q in *C. elegans* resulted in significant neuronal dysfunction and touch insensitivity, without causing neuronal death [60].

A yeast model of HD is often used. For example, yeasts were used to demonstrate various pathological effects of mHtt aggregation: disruptions in endocytosis, tryptophan metabolism, cell cycle, and protein degradation [61−63].

There are also nongenetic animal models of HD, which are based on the use of chemical compounds (*Fig. 2*). For example, 3-nitropropionic acid and quinolinic acid are used as excitotoxic agents in animal models of HD. The first compound is a toxin that acts on mitochondria and induces neurotoxicity by irreversible inhibition of succinate dehydrogenase, the key respiratory chain enzyme responsible for the oxidation of succinate to fumarate. Quinolinic acid is an agonist of the N-methyl-D-aspartate receptor. The excitotoxicity induced by these compounds was studied in striatum slices, sagittal slices of the hippocampus [64], and in slices of the hippocampus of transgenic R6/2 mice [65].

It should be noted that many pathological manifestations of HD can be studied at the cellular level (*Fig. 3*).

Cells can be transfected with both full-length mHtt and its fragments with a polyQ tract of a different length. For example, transfection of PC-12 cells with the first *mHtt* exon resulted in the nuclear localization of mHtt, changes in the morphology and expression of genes, and a lower rate of survival [66].

A large number of immortalized cell lines modeling HD have been generated, but not all pathological manifestations of HD can be revealed by these models. For this reason, primary neuronal cultures derived from transgenic mouse models of HD [67–69] or neurons isolated from wild-type animals and transfected with a vector for the expression of mHtt or its fragment have been used quite often [16].

An interesting and promising model of HD may be corticostriatal slices of a rat brain which are transfected with constructs expressing human mHtt. This model has an advantage over simple cellular models, because it maintains permanent cell-cell interactions, which is important in studying HD pathogenesis [70]. This model may be used to study the effect of potential therapeutic agents effective in HD.

One of the most advanced and promising approaches to the modeling of HD and other neurodegenerative diseases is the use of patient-specific induced pluripotent stem cells (HD-iPSCs) that endogenously express mutant huntingtin. Protocols for differentiating iPSCs into a phenotype similar to the phenotype of striatal medium spiny neurons (MSNs) [71–73], the cells most affected in HD, have been developed. One of the advantages of HD-iPSCs is the opportunity to study the pathological processes associated with the expression of mHtt with a short polyQ tract [73], which usually does not cause pathological changes in other models.

The expression of genes and proteins in HD-iPSCs differed from that in the controls; changes in proteostasis, neuronal development, intracellular transport, RNA metabolism, and cellular metabolism were observed [74]. In addition, the degree of expression disturbance was directly correlated with the polyQ tract length. Neurons differentiated from HD-iPSCs had a disease-associated phenotype, including electrophysiological changes and changes in metabolism, cell adhesion, and cellular toxicity. Cells containing the longest polyQ tract were the most sensitive to stress: e.g., to the absence of the brain-derived neurotrophic factor (BDNF) in the cell medium. Studies of neurons differentiated from HD-iPSCs revealed changes in the lysosomal activity [73, 75], mitochondrial fragmentation [76], and transcriptional repressor activity [77].

Another area of HD-iPSC application is cell transplantation for replacing diseased cells. Neuronal precursors differentiated from iPSCs were implanted into rat HD models. In this case, restoration of normal behavior was observed [78]. HD-iPSC-derived neural precursors were found to similarly restore the population of GABAergic striatal neurons and normalize the behavior of rats; however, the transplanted cells began to exhibit pathological properties at later stages [78], emphasizing the need for preliminary genetic correction in autologous transplantation.

CALCIUM DYSHOMEOSTASIS IN HD

Calcium-signaling disruptions are characteristic of various neurodegenerative diseases, such as HD, Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis [16, 79–81]. In animal HD models, which were created using genetically delivered mHtt or induced by 3-nitropropionic acid (3-NPA), calcium-signaling disruptions were shown to be a hallmark of HD.

MHtt affects calcium signaling in the cell in many directions, including interactions with calcium-binding proteins and mitochondrial membranes, regulation of calcium influx from the extracellular medium, and release of calcium from intracellular stores (*Fig. 4*).

The main participants in neuronal calcium signaling include the calcium-binding proteins activated by binding to Ca^{2+} and regulating the free Ca^{2+} level, proteins exporting Ca^{2+} from the cytosol to the extracellular medium (plasma membrane ATPase, Na^+/Ca^{2+} exchangers) or organelle cavity (SERCA), and the calcium channels involved in Ca^{2+} delivery to the cytoplasm [82, 83].

MHtt directly interacts with calcium-binding proteins [84], which may lead to an increase in the intracellular Ca^{2+} concentration and dysfunction of the proteins [85]. In particular, interaction between mHtt and calmodulin was found to occur in large molecular-weight-protein complexes [84], and disturbance of this interaction had a neuroprotective effect [85, 86]. One of the causes of an adverse effect of a prolonged increased Ca^{2+} level in the cytosol is the activation of calpain, a Ca^{2+}-activated cysteine protease the action of which is almost irreversible. Calpain destroys cytoskeletal proteins and other perimembrane proteins. In a *Drosophila* model of HD, inhibition of calpain was shown to prevent the aggregation and toxicity of mHtt, stimulating autophagy. Overexpression of a calpain inhibitor, calpastatin, increases the number of autophagosomes and has a positive effect on mouse models of HD, which makes this process appropriate for developing approaches to HD therapy [87].

It is important to note that calcium-signaling disturbance in HD occurs at the transcription level, because mHtt fragments change the expression of some calcium homeostasis genes both in mouse models and in HD patients [6, 88]. Genomic studies conducted in various HD models have revealed significant differ-
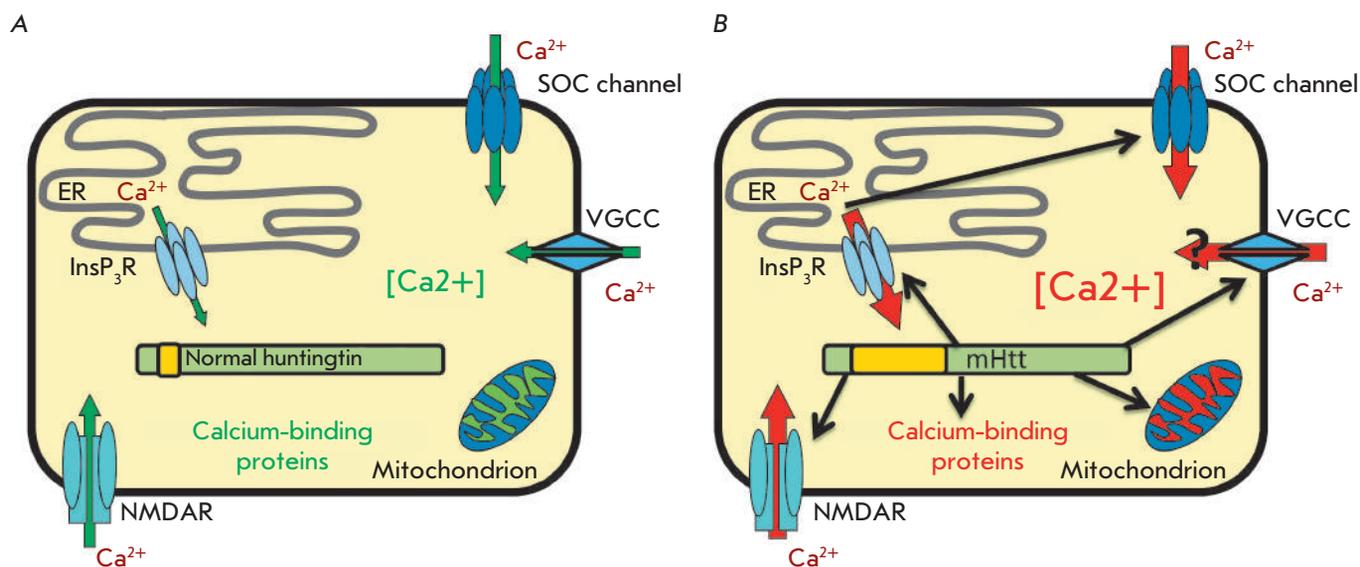
Fig. 4. Calcium signaling disturbances in cells expressing mHtt. *A* – major pathways of calcium homeostasis regulation in normal cells. Green arrows indicate main calcium flows in health. *B* – the effect of mHtt expression on calcium homeostasis in the cell. Red arrows denote main impaired or potentially impaired calcium flows in HD. Changes in mitochondrial membrane permeability and disturbances in the expression and function of calcium-binding proteins are also shown in red. Black arrows indicate the calcium-signaling mechanisms affected by mHtt expression. ER – endoplasmic reticulum, NMDAR – N-methyl D-aspartate receptor, VGCC – voltage-gated calcium channel, SOC channel – store-operated calcium channel, InsP$_3$R – inositol-1,4,5-trisphosphate receptor.

ences in the mRNA levels of the genes encoding the proteins involved in intracellular Ca^{2+} regulation, including calcium-binding proteins such as parvalbumin, calmodulin, calbindin, and hypocalcin, as well as ryanodine receptor type 1, the inositol trisphosphate receptor (InsP$_3$R1), and different subunits of voltage-gated calcium channels (VGCCs) [88−91]. In particular, the level of mRNA in the sarco-endoplasmic reticulum-associated ATP2A2 calcium pump (SERCA2) was reduced in peripheral blood mononuclear cells in HD [92].

Recently, a number of genes encoding calcium-signaling proteins the expression of which was disturbed in neurons differentiated from HD-iPSCs were identified using an analysis of gene ontology categories [73].

Transcription disturbances can be additionally enhanced by calcium-dependent control mechanisms. This may occur due to abnormal calcium-dependent regulation of the activity and stability of transcription factors, as well as changes in the functions of some calcium-binding proteins: e.g., the transcriptional repressor DREAM (downstream responsive element antagonist modulator), which is translocated to the nucleus in response to an increase in [Ca^{2+}] in the cytosol [93, 94], as well as the cofactor LMO4, the activity of which is induced by Ca^{2+} influx via VGCC [95].

Also, the activity of glutamate receptors increases in HD, which leads to a significant calcium influx into the

cytosol via the plasma membrane (PM), neuronal disturbances, and cell death. A relationship between polyglutamine expansion and neuronal sensitivity to glutamate-mediated excitotoxicity has been established [96]. Increased calcium influx into the cytosol via NMDA receptors (NMDARs) is associated with the potentiating effect of mHtt on the transport and incorporation of NMDAR into PM [97]. In this case, differences in the expression level and subunit composition of NMDARs in different cells may be one of the causes for the selective death of MSNs in HD [98]. Pharmacological inhibition of NMDARs exerted a neuroprotective effect on a primary culture of MSNs of HD mouse models [99, 100]. It should also be noted that YAC128 mice were characterized by an increased expression of the extrasynaptic NMDAR, which resulted in disruption of the p38 MAPK and CREB signaling pathways, as well as dysfunction and atrophy of the striatum [101].

Also, mHtt was shown to affect VGCC by binding directly to the auxiliary α2/δ subunit of VGCC [102]. The association of the N-terminal domain of huntingtin (both mutant and normal) with the pore-forming CaV2.2 subunit of N-type VGCC leads to a displacement of the syntaxin 1A that negatively regulates the channel and, as a result, to an increase in the activity of N-type VGCC [103]. This example indicates the potential physiological functions of cleavage of the N-

terminal fragment from normal huntingin, whereas further research is needed to understand the role of N-type VGCC in the disease. At the same time, a potential hyperfunction of VGCC in HD is confirmed by the results obtained in *Drosophila*, which demonstrate that removal of Dmca1D (the L-type VGCC channel in *Drosophila*) leads to decreased photoreceptor neurodegeneration [104].

Overexpression of mHtt fragments in striatal neuronal precursor cells (Q7/7) resulted in a significant decrease in $[Ca^{2+}]$ in the endoplasmic reticulum (ER), while $[Ca^{2+}]$ in the cytosol remained the same as in the controls [105]. Application of cyclopiazonic acid induced an increased release of Ca^{2+} from the ER to the cytosol in a striatal cell line derived from knock-in mouse embryos expressing mHtt with 111Q [106]. At the same time, expression of mHtt in PC-12 cells did not lead to statistically significant changes in the ER calcium level [91].

MHtt (but not wild-type Htt) was shown to directly interact with the C-terminal region of InsP$_3$R1, increasing its sensitivity to InsP$_3$ [107], thereby promoting the outflow of Ca^{2+} from the ER. An important role of InsP$_3$R1 in polyglutamine expansion-induced neurotoxicity was experimentally confirmed in a primary culture of MSNs from a HD mouse model [99, 102] and in *Drosophila* [108]. Also, a peptide that disrupts the interaction between mHtt and InsP$_3$R1 was found to exert a neuroprotective effect on MSN cells from a HD model [109]. In addition, inhibition of *InsP$_3$R* gene expression reduces mHtt aggregation [110], which emphasizes the importance of the interaction of two proteins in the pathogenesis of HD.

MHtt that interacts with InsP$_3$R1 and, thereby, affects the ER Ca^{2+} level may disrupt the functions of store-operated calcium (SOC) channels. These channels are activated in response to a decrease in the calcium concentration in intracellular calcium stores, the most common of which is the ER. Thus, the activation of InsP$_3$R1 will result not only in store depletion, but also in subsequent store-operated calcium entry via the PM. It is important to note that disruption of SOCE (SOC Entry) has been established in many neurodegenerative diseases, including Alzheimer's disease, spinocerebellar ataxia, and HD [80, 111−113].

SOCE disruption may be caused by a change in the level of STIM1/2 proteins containing EF-hand domains and acting as calcium sensors in the ER lumen. These changes can be caused by the impaired proteasomal degradation that occurs in neurodegeneration [114].

A significant increase in SOCE was found in SK-N-SH neuroblastoma cells expressing mHtt 138Q [113]. It was suggested that the significant increase in SOCE in cell models of HD was mediated not by changes in the properties of SOC channels but by an increase in their number; however, it should be noted that no direct experimental evidence of this hypothesis was provided.

A significant increase in SOCE was also found in SK-N-SH cells expressing not full-length mHtt, but its N-terminal fragment. Additionally, the STIM1 protein was shown to be required for SOCE activation. Suppression of STIM1 was accompanied by a decrease in SOCE, and detected currents might be divided into two types: high and low reversal potentials, which implies competition of at least two types of SOC channels for interaction with STIM1 [115]. The data indicating that at least two different proteins mediate calcium entry by the store-operated mechanism was also obtained in HD models: Neuro-2a mouse neuroblastoma cells and a primary culture of mouse striatal neurons [16]. Using patch-clamp and RNA interference, the authors found that the pore-forming proteins Orai1 and TRPC1 together maintain SOCE in cells that express an N-terminal fragment of mHtt with 138Q, which may be explained by the existence of a heteromeric channel containing subunits of Orai1 and TRPC1 [16]. This heteromeric channel was hypothesized as early as 2007 [116], but no experimental evidence confirming this idea was presented. At the same time, calcium entry via Orai1-formed channels was shown to be necessary for the incorporation of TRPC1 proteins into the PM [117]. Therefore, it may be assumed that TRPC1 proteins largely contribute to the amplitude of store-operated currents in a Neuro-2a cell model of HD, which is confirmed by a dramatic current drop upon TRPC1 suppression. However, upon Orai1 suppression, a significant reduction in the SOCE amplitude was also observed, which may now be explained not only by a decrease in the current through Orai1, but also by a decreased TRPC1-mediated current component due to a disruption of TRPC1 traffic to the plasma membrane [16]. The importance of TRPC1 in the pathogenesis of HD is also confirmed by data demonstrating that TRPC1 suppression by a short interfering RNA has a significant protective effect on MSNs of YAC128 mice in a model of glutamate-induced apoptosis. In this case, suppression of TRPC1 in the neurons of wild-type mice had practically no effect on glutamate-mediated cell death [111].

Expression of the N-terminal fragment of mHtt in a primary culture of MSNs also results in abnormally large store-operated calcium entry into the cytosol [16]. These results are confirmed by measurements of the intracellular calcium concentration using a calcium probe, FURA-2, in MSN cells isolated from YAC128 mice [111]. Furthermore, the effect of a NF-κB signaling pathway inhibitor, EVP4593, on these cells was studied. There is a close relationship between activa-

tion of NF-κB and store-operated calcium entry [118, 119]. NF-κB is able to bind to the *Htt* gene and enhance the activity of its promoter in mouse striatal neurons [24]. MHtt can also bind to one of the key enzymes of the NF-κB signaling pathway, IKK, thereby increasing its activity [120].

EVP4593 was shown to reversibly reduce abnormally large SOCE to control values both in SK-N-SH cells expressing mHtt with 138Q and in MSNs of YAC128 mice [111]. EVP4593 exerted a similar effect on MSN cells expressing an N-terminal fragment of mHtt [16]. Now, EVP4593 is proven to act as an inhibitor of SOCE necessary for the initial stages of the NF-κB signaling pathways; however, the molecular target of EVP4593 remains unknown.

It should be noted that EVP4593 has a high therapeutic potential, because it exerts a neuroprotective effect in glutamate-induced apoptosis of MSNs from YAC128 mice and induces a positive effect in motor assays in fly models of HD [111]. Cytofluorimetric measurements demonstrated that incubation of Neuro-2a cells (HD model) with EVP4593 results in increased survival of the cells [16].

The published data suggest that the neuroprotective effect of EVP4593 is based on a negative feedback present in mHtt influencing the cell. Since NF-κB is able to bind directly to *Htt* and enhance the activity of its promoter [24], and EVP4593 inhibits NF-κB signal transduction, a potential result of EVP4593 application may be a decreased mHtt expression and, as a consequence, a decrease in the toxic functions of mHtt. Nevertheless, additional research is required to confirm this idea.

Of special interest is a study of the effect of mHtt expression on SOCE which was performed in HD-specific human neurons differentiated from iPSCs and expressing mHtt with a low Q number in the tract. Despite the fact that the polyglutamine tract of mHtt in this HD model contained only 40−47Q, which was close to the normal value, changes in SOCE were as significant as those in other models with a tract length exceeding 100Q [73]. In this case, EVP4593 decreased the SOCE amplitude both in pathology and in controls and likewise had a neuroprotective effect upon exposure to the proteasome inhibitor MG132 [73].

In general, the conducted studies demonstrate that SOCE abnormalities are systemic and occur in various cellular models of HD (*Fig. 5*) [16, 73, 111]. This fact may indicate that SOCE abnormalities precede other pathological processes in HD and, probably, are of the central mechanisms underlying neurodegeneration. Thus, SOCE may be considered a promising target for the development of approaches to HD therapy and the data obtained in various cellular models may be used for the development of EVP4593-based drugs.

An increase in SOCE is supposed to directly affect the ability of mitochondria to store Ca^{2+}, since mitochondria are located in immediate vicinity to the site of ER Ca^{2+} release [121]. The mitochondrion is one of the main regulators of the intracellular Ca^{2+} level. A significant increase in $[Ca^{2+}]$ in the cytosol, in the immediate vicinity of the mitochondrion, is accompanied by the activation of the low affinity mitochondrial Ca^{2+} uniporter (MCU) mediating Ca^{2+} influx into the matrix. The mitochondria release Ca^{2+} via the Na^+/Ca^{2+} exchanger [122] or, in the case of calcium overload, via megapores (PTP), the activation of which leads to a membrane potential jump, rupture of the outer membrane, and release of cytochrome C and caspases, which results in apoptotic cell death [123, 124]. The involvement of mitochondrial dysfunction in the pathogenesis of HD is confirmed, in particular, by the fact that 3-nitropropionic acid, which is used as an inhibitor of the mitochondrial respiratory chain complex II, causes impairments typical of HD [125]. An additional piece of evidence of the important role of mitochondria in the pathogenesis of HD is the neuroprotective effect of mitochondrial membrane permeability inhibitors, which has been demonstrated in both cellular and animal models [99, 126].

Expression of mHtt was also accompanied by defects in mitochondria morphology. In a cell line derived from knock-in mouse embryos expressing mHtt with 111Q, mitochondria are more prone to fragmentation because abnormal $[Ca^{2+}]$ in the cytosol promotes an increase in the activity of a calcium-dependent phosphatase, calcineurin, dephosphorylating (and, thereby, activating) the Drp1 protein responsible for mitochondrial division. Finally, enhanced mitochondrial fragmentation promotes cell apoptosis [106].

Impairment of Ca^{2+} buffering and calcium metabolism in mitochondria was detected at both early and late stages of HD, which indicates the key role of these impairments in the pathogenesis of HD. Mitochondria isolated from the brain cells of HD patients and from the cells of HD mouse models were more sensitive to Ca^{2+} stress and tended to form megapores [127, 128]. Similar results were obtained later in an immortalized line of striatal neuronal precursor cells derived from knock-in KI-Hdh^{Q111} mice [129]. However, the susceptibility of mitochondria to calcium stress was reproduced not in all experimental models. For example, striatal mitochondria isolated from knock-in mice expressing different mHtt variants (80, 92, or 111Q), R6/2 mice, and YAC128 mice were equally, and in some cases even less, susceptible to Ca^{2+} stress than the control wild-type samples [130, 131]. In addition, the sensitivity of mitochondria to Ca^{2+} stress in some HD models decreased proportionally to the age and polyQ tract
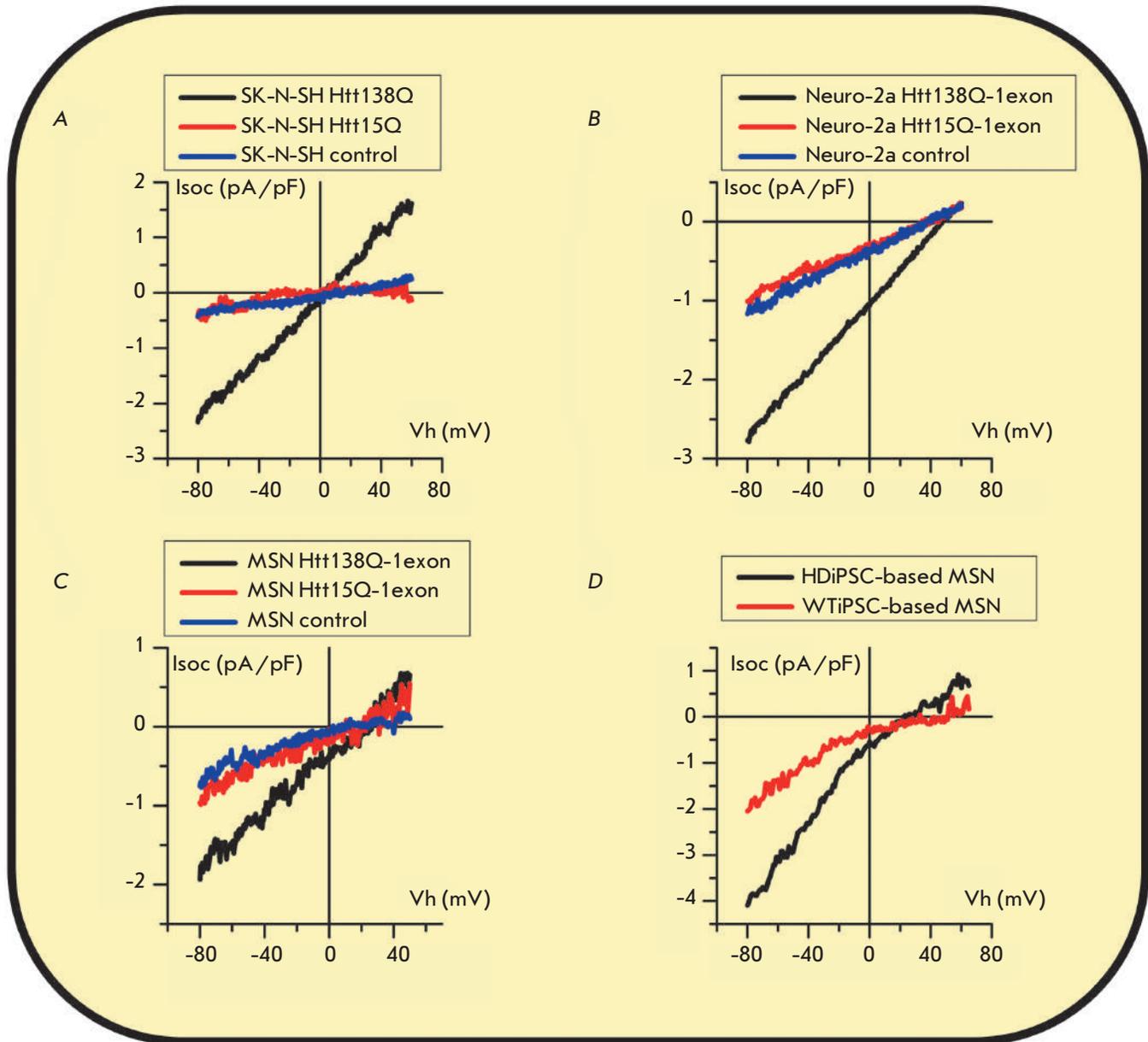
Fig. 5. Abnormal increase in store-operated calcium entry in cellular models of Huntington's disease. Mean current–voltage curves presented at maximum level of current development and normalized to cell capacitance, which represent store-operated calcium entry in SK–N–SH human neuroblastoma cells expressing full-length huntingtin 138Q, 15Q, or an empty control vector (*A*) [111]; in Neuro-2a mouse neuroblastoma cells expressing the first exon of huntingtin containing 138Q, 15Q, or an empty control vector (*B*) [16]; in a primary culture of mouse striatal MSNs expressing the first exon of huntingtin containing 138Q, 15Q, or an empty control vector (*C*) [16]; in human neurons differentiated from HD-specific iPSCs or wild-type iPSCs (*D*) [73].

length [130], which suggests the presence of protective compensatory mechanisms. A recent study of isolated mitochondria and striatal neurons in R6/2 mice has also revealed the absence of respiratory chain dysfunction and increased mitochondrial sensitivity to calcium stress [132]. Therefore, the role of mitochondria in the pathogenesis of HD remains controversial. Further research is needed to elucidate the molecular mechanisms of the disease.

CONCLUSION

Despite the long history of HD research, the issues of manifestation, simulation, and investigation of the molecular basis of the disease remain topical. This

review has described the cellular and animal models widely used in investigations of the pathological processes in HD, as well as in screening for potential drugs. Of particular interest are models based on endogenous expression of mutant huntingtin in neurons differentiated from patient-specific iPSCs. The analysis of recent publications indicates that abnormal calcium signaling is one of the central links that mediate the development of the pathology and lead to neuronal death. One of the most important elements of calcium signaling, which is impaired in HD, is the store-operated calcium entry, whose pathological increase was demonstrated in many of the models described in this review. It is likely that an abnormal ER calcium level, together with coupled excessive store-operated calcium entry via the PM, may af-

fect the mitochondria that activate the process of cell death, being unable to store excessive calcium.

In summary, it should be noted that the investigation of neurodegeneration is a research field that is developing intensively, which gives hope that a complete picture of neurodegeneration processes could be built and that new drugs effective against HD, Alzheimer's disease, Parkinson's disease, and other pathologies would be developed. ●

REFERENCES

1. Huntington G. // Med. Surg. Rept. Philadelphia. 1872. V. 26. № 15. P. 317–321.
2. The Huntington's Disease Collaborative Research Group // Cell. 1993. V. 72. № 6. P. 971–983.
3. Vonsattel J.P., Difiglia M. // J. Neuropathol. Exp. Neurol. 1998. V. 57. № 5. P. 369–384.
4. Ugryumov M.V. // Vestn. Ross. Akad. Med. Nauk. 2010. V. 8. P. 6–19.
5. Sarkar A., Irwin M., Singh A., Riccetti M., Singh A. // Neural Regen. Res. 2016. V. 11. № 5. P. 693–697.
6. Zuccato C., Valenza M., Cattaneo E. // Physiol. Rev. 2010. V. 90. № 3. P. 905–981.
7. Francelle L., Galvan L., Brouillet E. // Front Cell Neurosci. 2014. V. 8. P. 295.
8. Plotkin J.L., Surmeier D.J. // Curr. Opin. Neurobiol. 2015. V. 33. P. 53–62.
9. Sepers M.D., Raymond L.A. // Drug Discov. Today. 2014. V. 19. № 7. P. 990–996.
10. Margulis J., Finkbeiner S. // Front Cell Neurosci. 2014. V. 8. P. 218.
11. Hinckelmann M., Zala D., Saudou F. // Trends Cell Biol. 2013. V. 23. № 12. P. 634–643.
12. Panov A.V., Gutekunst C.A., Leavitt B.R., Hayden M.R., Burke J.R., Strittmatter W.J., Greenamyre J.T. // Nat. Neurosci. 2002. V. 5. № 8. P. 731–736.
13. Costa V., Scorrano L. // EMBO J. 2012. V. 31. № 8. P. 1853–1864.
14. Reddy P.H., Shirendeb U.P. // Biochim. Biophys. Acta. 2012. V. 1822. № 2. P. 101–110.
15. Giacomello M., Oliveros J.C., Naranjo J.R., Carafoli E. // Prion. 2013. V. 7. № 1. P. 76–84.
16. Vigont V., Kolobkova Y., Skopin A., Zimina O., Zenin V., Glushankova L., Kaznacheyeva E. // Front. Physiol. 2015. V. 6. P. 337.
17. Wu J., Ryskamp D.A., Liang X., Egorova P., Zakharova O., Hung G., Bezprozvanny I. // J. Neurosci. 2016. V. 36. № 1. P. 125–141.
18. Walker F.O. // Lancet. 2007. V. 369. № 9557. P. 218–228.
19. Andrew S., Theilmann J., Almqvist E., Norremolle A., Lucotte G., Anvret M., Sorensen S.A., Turpin J.C., Hayden M.R. // Clin. Genet. 1993. V. 43. № 6. P. 286–294.
20. Langbehn D.R., Brinkman R.R., Falush D., Paulsen J.S., Hayden M.R., International Huntington's Disease Collaborative Group. // Clin. Genet. 2004. V. 65. № 4. P. 267–277.
21. Brinkman R.R., Mezei M.M., Theilmann J., Almqvist E., Hayden M.R. // Am. J. Hum. Genet. 1997. V. 60. № 5. P. 1202–1210.
22. Nazé P., Vuillaume I., Destée A., Pasquier F., Sablonnière B. // Neurosci. Lett. 2002. V. 328. № 1. P. 1–4.
23. Metzger S., Rong J., Nguyen H.P., Cape A., Tomiuk J., Soehn A.S., Propping P., Freudenberg-Hua Y., Freudenberg J., Tong L., et al. // Hum. Mol. Genet. 2008. V. 17. № 8. P. 1137–1146.
24. Bečanović K., Nørremølle A., Neal S.J., Kay C., Collins J.A., Arenillas D., Lilja T., Gaudenzi G., Manoharan S., Doty C.N., et al. // Nat. Neurosci. 2015. V. 18. № 6. P. 807–816.
25. Wild E.J., Boggio R., Langbehn D., Robertson N., Haider S., Miller J.R., Zetterberg H., Leavitt B.R., Kuhn R., Tabrizi S.J., et al. // J. Clin. Invest. 2015. V. 125. № 5. P. 1979–1986.
26. Hu Y., Chopra V., Chopra R., Locascio J.J., Liao Z., Ding H., Zheng B., Matson W.R., Ferrante R.J., Rosas H.D., et al. // Proc. Natl. Acad. Sci. USA. 2011. V. 108. № 41. P. 17141–17146.
27. Mastrokolias A., Ariyurek Y., Goeman J.J., van Duijn E., Roos R.A., van der Mast R.C., van Ommen G.B., den Dunnen J.T., 't Hoen P.A., van Roon-Mom W.M. // Eur. J. Hum. Genet. 2015. V. 23. № 10. P. 1349–1356.
28. Aziz N.A., Anguelova G.V., Marinus J., Lammers G.J., Roos R.A. // Parkinsonism Relat. Disord. 2010. V. 16. № 5. P. 345–350.
29. Mesko B., Poliska S., Szegedi A., Szekanecz Z., Palatka K., Papp M., Nagy L. // BMC Med. Genomics. 2010. V. 3. P. 15.
30. Demougeot C., Garnier P., Mossiat C., Bertrand N., Giroud M., Beley A., Marie C. // J. Neurochem. 2001. V. 77. № 2. P. 408–415.
31. Sturrock M., Hao W., Schwartzbaum J., Rempala G.A. // J. Theor. Biol. 2015. V. 380. P. 299–308.
32. Wood N.I., Goodman A.O., van der Burg J.M., Gazeau V., Brundin P., Björkqvist M., Petersén A., Tabrizi S.J., Barker R.A., Morton A.J. // Brain Res. Bull. 2008. V. 76. № 1–2. P. 70–79.
33. Weir D.W., Sturrock A., Leavitt B.R. // Lancet Neurol. 2011. V. 10. № 6. P. 573–590.

34. Chen C.M., Wu Y.R., Cheng M.L., Liu J.L., Lee Y.M., Lee P.W., Soong B.W., Chiu D.T. // Biochem. Biophys. Res. Commun. 2007. V. 359. № 2. P. 335−340.

35. Dalrymple A., Wild E.J., Joubert R., Sathasivam K., Björkqvist M., Petersén A., Jackson G.S., Isaacs J.D., Kristiansen M., Bates G.P., et al. // J. Proteome Res. 2007. V. 6. № 7. P. 2833−2840.

36. Ehrnhoefer D.E., Sutton L., Hayden M.R. // Neuroscientist. 2011. V. 17. № 5. P. 475−492.

37. Harjes P., Wanker E.E. // Trends Biochem. Sci. 2003. V. 28. V. 8. P. 425−433.

38. Warby S.C., Doty C.N., Graham R.K., Carroll J.B., Yang Y.Z., Singaraja R.R., Overall C.M., Hayden M.R. // Hum. Mol. Genet. 2008. V. 17. № 15. P. 2390−2404.

39. Kim M.W., Chelliah Y., Kim S.W., Otwinowski Z., Bezprozvanny I. // Structure. 2009. V. 17. № 9. P. 1205−1212.

40. Michalek M., Salnikov E.S., Bechinger B. // Biophys. J. 2013. V. 105. № 3. P. 699−710.

41. Vijayvargia R., Epand R., Leitner A., Jung T.Y., Shin B., Jung R., Lloret A., Singh Atwal R., Lee H., Lee J.M., et al. // Elife. 2016. pii: e11184.

42. Difiglia M., Sapp E., Chase K.O., Davies S.W., Bates G.P., Vonsattel J.P., Aronin N. // Science. 1997. V. 277. № 5334. P. 1990−1993.

43. Davies S.W., Turmaine M., Cozens B.A., DiFiglia M., Sharp A.H., Ross C.A., Scherzinger E., Wanker E.E., Mangiarini L., Bates G.P. // Cell. 1997. V. 90. № 3. P. 537−548.

44. Rockabrand E., Slepko N. Pantalone A., Nukala V.N., Kazantsev A., Marsh J.L., Sullivan P.G., Steffan J.S., Sensi S.L., Thompson L.M. // Hum. Mol. Genet. 2007. V. 16. № 1. P. 61−77.

45. Sadri-Vakili G., Cha J.H. // Nat. Clin. Pract. Neurol. 2006. V. 2. № 6. P. 330−338.

46. Gutekunst C.A., Li S.H., Yi H., Mulroy J.S., Kuemmerle S., Jones R., Rye D., Ferrante R.J., Hersch S.M., Li X.J. // J. Neurosci. 1999. V. 19. № 7. P. 2522−2534.

47. Juenemann K., Weisse C., Reichmann D., Kaether C., Calkhoven C.F., Schilling G. // Neurotox. Res. 2011. V. 20. № 2. P. 120−133.

48. Saudou F., Finkbeiner S., Devys D., Greenberg M.E. // Cell. 1998. V. 95. № 1. P. 55−66.

49. Arrasate M., Mitra S., Schweitzer E.S., Segal M.R., Finkbeiner S. // Nature. 2004. V. 431. № 7010. P. 805−810.

50. Stefani M. // Prog. Neurobiol. 2012. V. 99. № 3. P. 226−245.

51. Carter R.J., Lione L.A., Humby T., Mangiarini L., Mahal A., Bates G.P., Dunnett S.B., Morton A.J. // J. Neurosci. 1999. V. 19. № 8. P. 3248−3257.

52. Li H., Li S.H., Cheng A.L., Mangiarini L., Bates G.P., Li X.J. // Hum. Mol. Genet. 1999. V. 8. № 7. P 1227−1236.

53. Menalled L.B., Chesselet M.F. // Trends Pharmacol. Sci. 2002. V. 23. № 1. P. 32−39.

54. Woodman B., Butler R., Landles C., Lupton M.K., Tse J., Hockly E., Moffitt H., Sathasivam K., Bates G.P. // Brain Res. Bull. 2007. V. 72. № 2−3. P. 83−97.

55. Li X.J., Li S. // J. Genet. Genomics. 2012. V. 39. № 6. P. 239−245.

56. Li Z., Karlovich C.A., Fish M.P., Scott M.P., Myers R.M. // Hum. Mol. Genet. 1999. V. 8. № 9. P. 1807−1815.

57. Marsh J.L., Walker H., Theisen H., Zhu Y.Z., Fielder T., Purcell J., Thompson L.M. // Hum. Mol. Genet. 2000. V. 9. № 1. P. 13−25.

58. Kahsai L., Zars T. // Int. Rev. Neurobiol. 2011. V. 99. P. 139−167.

59. Morley J.F., Brignull H.R., Weyers J.J., Morimoto R.I. // Proc. Natl. Acad. Sci. USA. 2002. V. 99. № 16. P. 10417−10422.

60. Parker J.A., Connolly J.B., Wellington C., Hayden M., Dausset J., Neri C. // Proc. Natl. Acad. Sci. USA. 2001. V. 98. № 23. P. 13318−13323.

61. Giorgini F., Guidetti P., Nguyen Q., Bennett S.C. Muchowski P.J. // Nat. Genet. 2005. V. 37. № 5. P. 526−531.

62. Kochneva-Pervukhova N.V., Alexandrov A.I., Ter-Avanesyan M.D. // PLoS One. 2012. V. 7. № 1. e29832.

63. Duennwald M.L., Lindquist S. // Genes Dev. 2008. V. 22. № 23. P. 3308−3319.

64. Colle D., Hartwig J.M., Soares F.A., Farina M. // Brain Res. Bull. 2012. V. 87. № 4−5. P. 397−405.

65. Smith D.L., Portier R., Woodman B., Hockly E., Mahal A., Klunk W.E., Li X.J., Wanker E., Murray K.D., Bates G.P. // Neurobiol. Dis. 2001. V. 8. № 6. P. 1017−1026.

66. Li S.H., Cheng A.L., Li H., Li X.J. // J. Neurosci. 1999. V. 19. № 13. P. 5159−5172.

67. Sarantos M.R., Papanikolaou T., Ellerby L.M., Hughes R.E. // J. Huntingtons Dis. 2012. V. 1. № 2. P. 195−210.

68. Milnerwood A.J., Kaufman A.M., Sepers M.D., Gladding C.M., Zhang L., Wang L., Fan J., Coquinco A., Qiao J.Y., Lee H., et al. // Neurobiol. Dis. 2012. V. 48. № 1. P. 40−51.

69. Doria J.G., Silva F.R., de Souza J.M., Vieira L.B., Carvalho T.G., Reis H.J., Pereira G.S., Dobransky T., Ribeiro F.M. // Br. J. Pharmacol. 2013. V. 169. № 4. P. 909−921.

70. Reinhart P.H., Kaltenbach L.S., Essrich C., Dunn D.E., Eudailey J.A., DeMarco C.T., Turmel G.J., Whaley J.C., Wood A., Cho S., et al. // Neurobiol. Dis. 2011. V. 43. № 1. P. 248−256.

71. Ebert A.D., Shelley B.C., Hurley A.M., Onorati M., Castiglioni V., Patitucci T.N., Svendsen S.P., Mattis V.B., McGivern J.V., Schwab A.J., et al. // Stem Cell Res. 2013. V. 10. № 3. P. 417−427.

72. Carri A.D., Onorati M., Lelos M.J., Castiglioni V., Faedo A., Menon R., Camnasio S., Vuono R., Spaiardi P., Talpo F., et al. // Development. 2013. V. 140. № 2. P. 301−312.

73. 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. Neurodegener. 2016. V. 11. P. 27. doi: 10.1186/s13024-016-0092-5.

74. HD iPSC Consortium. // Cell Stem Cell. 2012. V. 11. № 2. P. 264−278.

75. Camnasio S., Delli Carri A., Lombardo A., Grad I., Mariotti C., Castucci A., Rozell B., Lo Riso P., Castiglioni V., Zuccato C. et al. // Neurobiol. Dis. 2012. V. 46. № 1. P. 41−51.

76. Guo X., Disatnik M.-H., Monbureau M., Shamloo M., Mochly-Rosen D., Qi X. // J. Clin. Invest. 2013. V. 123. № 12. P. 5371−5388.

77. Charbord J., Poydenot P., Bonnefond C., Feyeux M., Casagrande F., Brinon B., Francelle L., Auregan G., Guillermier M., Cailleret M., et al. // Stem Cells. 2013. V. 31. № 9. P. 1816−1828.

78. Jeon I., Lee N., Li J.-Y., Park I.-H., Park K.S., Moon J., Shim S.H., Choi C., Chang D.J., Kwon J., et al. // Stem Cells. 2012. V. 30. № 9. P. 2054−2062.

79. Bezprozvanny I. // Trends Mol. Med. 2009. V. 15. № 3. P. 89−100.

80. Ryazantseva M., Skobeleva K., Glushankova L., Kaznacheyeva E. // J. Neurochem. 2016. V. 136. № 5. P. 1085−1095.

81. Abeti R., Abramov A.Y. // Pharmacol. Res. 2015. V. 99. P. 377−381.

82. Saris N.E., Carafoli E. // Biochemistry (Mosc). 2005. V. 70 № 2. P. 187−194.

83. Naranjo J.R., Mellström B. // J. Biol. Chem. 2012. V. 287. № 38. P. 31674−31680.

84. Bao J., Sharp A.H., Wagster M.V., Becher M., Schilling G., Ross C.A., Dawson V.L., Dawson T.M. // Proc. Natl. Acad. Sci. USA. 1996. V. 93. № 10. P. 5037–5042.

85. Kreutz M.R., Naranjo J.R., Koch K.W., Schwaller B. // Front. Mol. Neurosci. 2012. V. 5. P. 92.

86. Dudek N.L., Dai Y., Muma N.A. // Brain Pathol. 2010. V. 20. № 1. P. 176–189.

87. Menzies F.M., Garcia-Arencibia M., Imarisio S., O'Sullivan N.C., Ricketts T., Kent B.A., Rao M.V., Lam W., Green-Thompson Z.W., Nixon R.A., et al. // Cell Death Differ. 2015. V. 22. № 3. P. 433–444.

88. Luthi-Carter R, Hanson S.A., Strand A.D., Bergstrom D.A., Chun W., Peters N.L., Woods A.M., Chan E.Y., Kooperberg C., Krainc D., et al. // Hum. Mol. Genet. 2002. V. 11. № 17. P. 1911–1926.

89. Hodges A., Strand A.D., Aragaki A.K., Kuhn A., Sengstag T., Hughes G., Elliston L.A., Hartog C., Goldstein D.R., Thu D., et al. // Hum. Mol. Genet. 2006. V. 15. № 6. P. 965–977.

90. Kuhn A., Goldstein D.R., Hodges A., Strand A.D., Sengstag T., Kooperberg C., Becanovic K., Pouladi M.A., Sathasivam K., Cha J.H., et al. // Hum. Mol. Genet. 2007. V. 16. № 15. P. 1845–1861.

91. Czeredys M., Gruszczynska-Biegala J., Schacht T., Methner A., Kuznicki J. // Front. Mol. Neurosci. 2013. V. 6. P. 42.

92. Cesca F., Bregant E., Peterlin B., Zadel M., Dubsky de Wittenau G., Siciliano G., Ceravolo R., Petrozzi L., Pauletto G., Verriello L., et al. // PLoS One. 2015. V. 10. № 4. e0125259.

93. Ledo F., Kremer L., Mellström B., Naranjo J.R. // EMBO J. 2002. V. 21. № 17. P. 4583–4592.

94. Savignac M., Pintado B., Gutierrez-Adan A., Palczewska M., Mellström B., Naranjo J.R. // EMBO J. 2005. V. 24. № 20. P. 3555–3564.

95. Kashani A.H., Qiu Z., Jurata L., Lee S.K., Pfaff S., Goebbels S., Nave K.A., Ghosh A. // J. Neurosci. 2006. V. 26. № 32. P. 8398–8408.

96. Sun Y., Savanenin A., Reddy P.H., Liu Y.F. // J. Biol. Chem. 2001. V. 276. № 27. P. 24713–24718.

97. Fan M.M., Fernandes H.B., Zhang L.Y., Hayden M.R., Raymond L.A. // J. Neurosci. 2007. V. 27. № 14. P. 3768–3779.

98. Fan M.M., Raymond L.A. // Prog. Neurobiol. 2007. V. 81. № 5–6. P. 272–293.

99. Tang T.S., Slow E., Lupu V., Stavrovskaya I.G., Sugimori M., Llinás R., Kristal B.S., Hayden M.R., Bezprozvanny I. // Proc. Natl. Acad. Sci. USA. 2005. V. 102. № 7. P. 2602–2607.

100. Shehadeh J., Fernandes H.B., Zeron Mullins M.M., Graham R.K., Leavitt B.R., Hayden M.R., Raymond L.A. // Neurobiol. Dis. 2006. V. 21. № 2. P. 392–403.

101. Dau A., Gladding C.M., Sepers M.D., Raymond L.A. // Neurobiol. Dis. 2014. V. 62. P. 533–542.

102. Kaltenbach L.S., Romero E., Becklin R.R., Chettier R., Bell R., Phansalkar A., Strand A., Torcassi C., Savage J., Hurlburt A., et al. // PLoS Genet. 2007. V. 3. № 5. e82.

103. Swayne L.A., Chen L., Hameed S., Barr W., Charlesworth E., Colicos M.A., Zamponi G.W., Braun J.E. // Mol. Cell. Neurosci. 2005. V. 30. № 3. P. 339–351.

104. Romero E., Cha G.H., Verstreken P., Ly C.V., Hughes R.E., Bellen H.J., Botas J. // Neuron. 2008 V. 57. № 1. P. 27–40.

105. De Mario A., Scarlatti C., Costiniti V., Primerano S., Lopreiato R., Calì T., Brini M., Giacomello M., Carafoli E. // PLoS Curr. 2016. V. 8. doi: 10.1371/currents.hd.37fcb1c9a275 03dc845594ee4a7316c3.

106. Costa V., Giacomello M., Hudec R., Lopreiato R., Ermak G., Lim D., Malorni W., Davies K.J., Carafoli E., Scorrano L. // EMBO Mol. Med. 2010. V. 2. № 12. P. 490–503.

107. Tang T.S., Tu H., Chan E.Y., Maximov A., Wang Z., Wellington C.L., Hayden M.R., Bezprozvanny I. // Neuron. 2003. V. 39. № 2. P. 227–239.

108. Zhang H., Li Q., Graham R.K., Slow E., Hayden M.R., Bezprozvanny I. // Neurobiol. Dis. 2008. V. 31. № 1. P. 80–88.

109. Tang T.S., Guo C., Wang H., Chen X., Bezprozvanny I. // J. Neurosci. 2009. V. 29. № 5. P. 1257–1266.

110. Bauer P.O., Hudec R., Ozaki S., Okuno M., Ebisui E., Mikoshiba K., Nukina N. // Biochem. Biophys. Res. Commun. 2011. V. 416. № 1–2. P. 13–17.

111. Wu J., Shih H., Vigont V., Hrdlicka L., Diggins L., Singh C., Mahoney M., Chesworth R., Shapiro G., Ahlijanian M., et al. // Chem. Biol. 2011. V. 18. № 6. P. 777–793.

112. Egorova P., Popugaeva E., Bezprozvanny I. // Semin. Cell Dev. Biol. 2015. V. 40. P. 127–133.

113. Glushankova L.N., Zimina O.A., Vigont V.A., Mozhaeva G.N., Bezprozvanny I.B., Kaznacheeva E.V. // Dokl. Biol. Sci. 2010. V. 433. P. 293–295.

114. Kuang X.L., Liu Y., Chang Y., Zhou J., Zhang H., Li Y., Qu J., Wu S. // Cell Calcium. 2016. V. 59. № 4. P. 172–180.

115. Vigont V.A., Zimina O.A., Glushankova L.N., Kolobkova J.A., Ryazantseva M.A., Mozhayeva G.N., Kaznacheyeva E.V. // Acta Naturae. 2014. V. 6. № 4. P. 40–47.

116. Liao Y., Erxleben C., Yildirim E., Abramowitz J., Armstrong D.L., Birnbaumer L. // Proc. Natl. Acad. Sci. USA. 2007. V. 104. № 11. P. 4682–4687.

117. Cheng K.T., Ong H.L., Liu X., Ambudkar I.S. // Curr. Top Membr. 2013. V. 71. P. 149–179.

118. Dolmetsch R.E., Xu K., Lewis R.S. // Nature. 1998. V. 392. № 6679. P. 933–936.

119. Chang W.S. // Acta Pharmacol. Sin. 2006. V. 27. № 7. P. 813–820.

120. Khoshnan A., Ko J., Watkin E.E., Paige L.A., Reinhart P.H., Patterson P.H. // J. Neurosci. 2004. V. 24. № 37. P. 7999–8008.

121. Kipanyula M.J., Contreras L., Zampese E., Lazzari C., Wong A.K., Pizzo P., Fasolato C., Pozzan T. // Aging Cell. 2012. V. 11. № 5. P. 885–893.

122. Carafoli E. // Biochem. Soc. Symp. 1974. № 39. P. 89–109.

123. Giacomello M., Drago I., Pizzo P., Pozzan T. // Cell Death Differ. 2007. V. 14. № 7. P. 1267–1274.

124. Giacomello M., Oliveros J.C., Naranjo J.R., Carafoli E. // Prion. 2013. V. 7. № 1. P. 76–84.

125. Brouillet E., Jacquard C., Bizat N., Blum D. // J. Neurochem. 2005. V. 95. № 6. P. 1521–1540.

126. Wang H., Lim P.J., Karbowski M., Monteiro M.J. // Hum. Mol. Genet. 2009. V. 18. № 4. P. 737–752.

127. Panov A.V., Gutekunst C.A., Leavitt B.R., Hayden M.R., Burke J.R., Strittmatter W.J., Greenamyre J.T. // Nat. Neurosci. 2002. V. 5. № 8. P. 731–736.

128. Choo Y.S., Johnson G.V., MacDonald M., Detloff P.J., Lesort M. // Hum. Mol. Genet. 2004. V. 13. № 14. P. 1407–1420.

129. Lim D., Fedrizzi L., Tartari M., Zuccato C., Cattaneo E., Brini M., Carafoli E. // J. Biol. Chem. 2008. V. 283. № 9. P. 5780–5789.

130. Brustovetsky N., LaFrance R., Purl K.J., Brustovetsky T., Keene C.D., Low W.C., Dubinsky J.M. // J. Neurochem. 2005. V. 93. № 6. P. 1361–1370.

131. Oliveira J.M., Jekabsons M.B., Chen S., Lin A., Rego A.C., Gonçalves J., Ellerby L.M., Nicholls D.G. // J. Neurochem. 2007. V. 101. № 1. P. 241–249.

132. Hamilton J., Pellman J.J., Brustovetsky T., Harris R.A., Brustovetsky N. // Hum. Mol. Genet. 2016. pii: ddw133. (Epub).

C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors

A. A. Fedotova, A. N. Bonchuk, V. A. Mogila, P. G. Georgiev*

Institute of Gene Biology, Russian Academy of Sciences, Vavilov Str., 34/5, Moscow, 119334, Russia

E-mail: georgiev_p@mail.ru

Received: October 27, 2016; in final form February 22, 2017

ABSTRACT The emergence of whole-genome assays has initiated numerous genome-wide studies of transcription factor localizations at genomic regulatory elements (enhancers, promoters, silencers, and insulators), as well as facilitated the uncovering of some of the key principles of chromosomal organization. However, the proteins involved in the formation and maintenance of the chromosomal architecture and the organization of regulatory domains remain insufficiently studied. This review attempts to collate the available data on the abundant but still poorly understood family of proteins with clusters of the C2H2 zinc finger domains. One of the best known proteins of this family is a well conserved protein known as CTCF, which plays a key role in the establishment of the chromosomal architecture in vertebrates. The distinctive features of C2H2 zinc finger proteins include strong and specific binding to a long and unique DNA recognition target sequence and rapid expansion within various animal taxa during evolution. The reviewed data support a proposed model according to which many of the C2H2 proteins have functions that are similar to those of the CTCF in the organization of the chromatin architecture.

KEYWORDS architectural proteins, CTCF, KRAB domain, SCAN domain, transcription factors, ZAD.

ABBREVIATIONS aa − amino acid, bp − base pair, C2H2 proteins − proteins containing clusters of C2H2 zinc fingers, TAD − topologically associated domain, TF − transcription factor.

INTRODUCTION

Recent genome-wide studies of intra- and interchromosomal interactions have revealed that the human, mouse, and *Drosophila* chromosomes are organized into large topologically associated domains (TADs) [1−4]. Long-distance interactions between promoters, enhancers, and silencers can occur within topological domains, which affect the regulation of gene expression [5, 6]. However, the mechanisms that underlie the organization and maintenance of the chromosomal architecture remain poorly understood [7]. It has been posited that there is a special class of architectural proteins whose inactivation significantly affects the distribution of inter- and intrachromosomal contacts [8, 9].

Vertebrates have a highly conserved transcription factor (TF), CTCF, which is considered to be the main architectural protein of chromosomes [10, 11]. CTCF, along with the cohesin complex, participates in the formation of topological domain boundaries and also maintains the long-distance interactions between the regulatory elements within the domains [12−14]. CTCF contains a cluster of C2H2 zinc finger domains, some of which are responsible for a highly specific binding

of the protein to DNA. Proteins containing C2H2 zinc fingers (C2H2 proteins) emerged early during evolution and are found in many eukaryotes [15, 16]. Many of them are structurally similar to CTCF. C2H2 proteins could be divided into three groups [17]: 1) proteins with one, two, or several randomly distributed C2H2 domains; 2) proteins with three C2H2 domains organized into a C-terminal cluster; and 3) proteins with more than three C2H2 domains, forming one or more clusters. The best studied group includes conserved TFs with three C2H2 domains, with many of them playing a critical role in the regulation of gene expression in all higher eukaryotes [18, 19]. This review is devoted to the poorly studied TFs that contain more than three C2H2 domains.

THE STRUCTURE AND FUNCTIONAL ROLE OF THE C2H2 DOMAIN

C2H2 zinc fingers (Cys2-His2) represent one of the most common domains found in the TFs of higher eukaryotes. The classical C2H2 domain of 28−30 aa includes a β-hairpin (antiparallel β-sheet consisting of two β-strands), followed by an α-helix, which form a
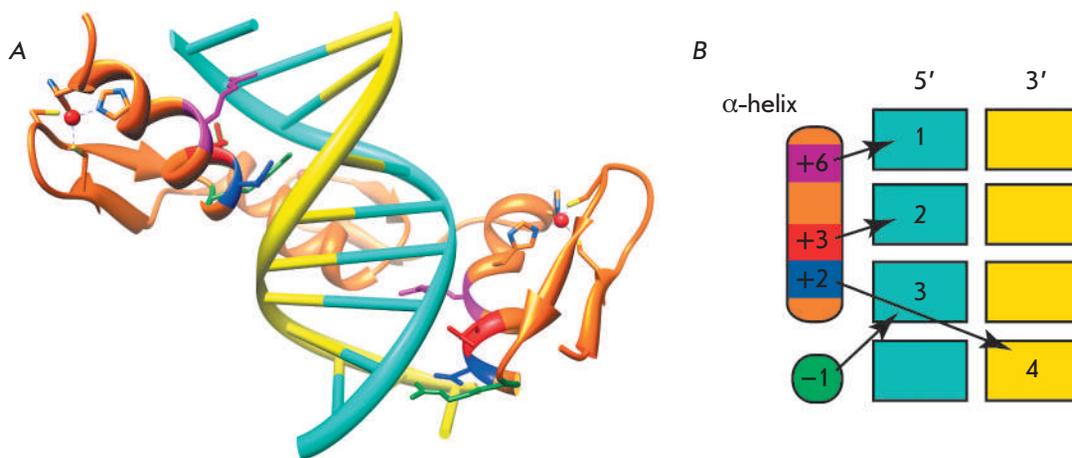
Fig. 1. A model of the site-specific DNA recognition by C2H2 zinc finger domains. (A) The crystal structure of three zinc fingers of the Zif268 protein bound to DNA [20]. The amino acids involved in the site-specific DNA recognition are color-coded: −1 – green, +2 – blue, +3 – red, and +6 – purple. (B) A model of the site-specific DNA recognition by α-helical amino acids (adapted from [24]).

left-handed ββα structure (*Fig. 1A*). The zinc finger structure is stabilized by the coordination of a zinc atom with two conserved cysteine residues at one end of the β sheet and with two conserved histidine residues at the α-helix C-terminus. The cysteine and histidine pairs are conserved, as well as the hydrophobic core forming the α-helix. The other amino acid residues in C2H2 domains are very variable.

One of the first structures to be determined was that of a complex of three tandem C2H2 domains of the mammalian Zif268 protein [20]. The three zinc fingers were found to form a semicircle located in the major DNA groove (*Fig. 1A*). Each of the three C2H2 domains binds to three or four DNA nucleotides via amino acids at the same α-helical positions (*Fig. 1B*): arginine at position −1, as well as amino acid residues at positions 2, 3, and 6. Biochemical and structural studies of the C2H2 domains confirmed the key role of the amino acids at these positions for the specific binding to DNA. According to the canonical model, the amino acids at positions 6, 3, and −1 are responsible for recognition of the first, second, and third nucleotides at the 5'-end, respectively, and the amino acid at position 2 recognizes the fourth nucleotide on the complementary strand (*Fig. 1B*).

Structural studies of C2H2 domains have revealed a new principle of DNA recognition. A distinctive feature of the C2H2 proteins is their specific binding to long (20−40 bp) DNA sequences, which distinguishes this class of proteins from the other TFs that usually recognize relatively short, degenerate DNA sequences. Typically, the tandem C2H2 domains involved in DNA recognition are separated by conserved sequences of 5 aa [21]. The existing algorithms can predict very ac-

curately the binding sequence for a cluster of C2H2 domains and, conversely, to select C2H2 domain combinations that recognize a target DNA sequence [22, 23]. However, the interference between neighboring C2H2 domains in large clusters (more than three C2H2 domains) complicates an accurate prediction of the binding site [24].

In contrast to the invariant mechanism of the interactions between the C2H2 domains and DNA, contacts between the domains and proteins and RNAs form via various amino acid combinations, which has been detailed in other reviews [25, 26]. Typically, the C2H2 clusters are located in the middle or at the C-terminus of a protein. Most proteins that contain a C2H2 cluster in the middle position do not have other conserved domains. At the same time, proteins with a cluster at the C-terminus often contain additional N-terminal domains (*Fig. 2*). The KRAB and SCAN domains are typical of vertebrates, while the ZAD is typical of insects [27, 28]. A small group of C2H2 proteins has a conservative BTB/POZ domain at the N-terminus. This domain is often found in different classes of proteins. Therefore, we have excluded this group of C2H2 proteins from the present review. Comprehensive information on BTB-containing proteins is available in detailed reviews [29, 30].

TRANSCRIPTION FACTORS CONTAINING ONLY A SINGLE CLUSTER OF C2H2 DOMAINS

The group of TFs containing only a single cluster of C2H2 domains includes the best studied and highly conserved CTCF protein (CCCTC-binding factor) [31] that was first described as a negative regulator of *myc* gene expression [32]. Later, a binding site for CTCF was
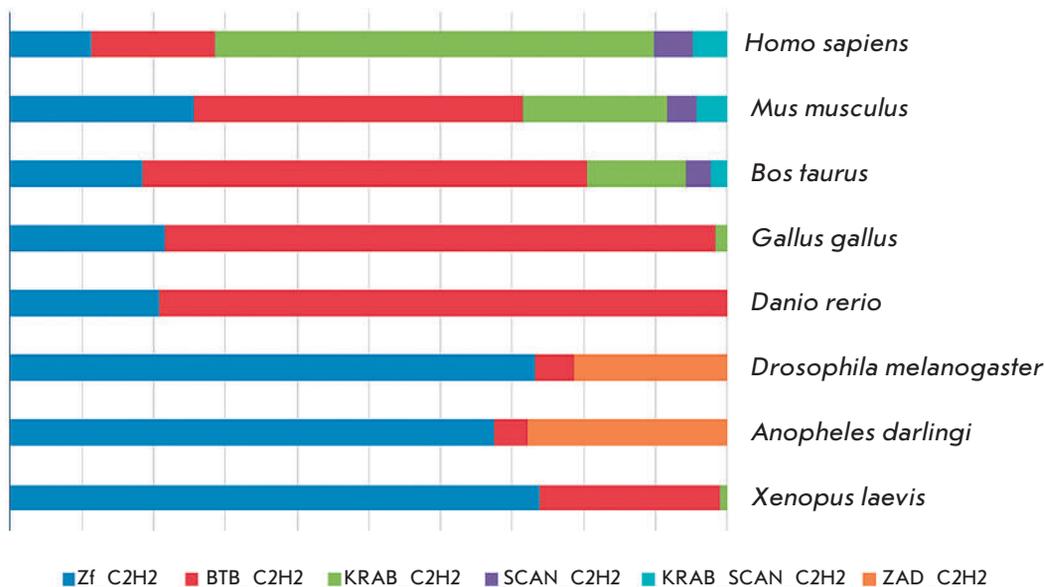
found in an insulator located at the 5'-end of the chicken β-globin locus [33]. The CTCF binding sites are often located at the boundaries of chromosomal regions, which have different epigenetic statuses and transcriptional activity, as well as at the boundaries of the topologically associated domains (TADs) that spatially separate chromosomes into regions where interactions among regulatory elements occur [34−37].

CTCF is one of the few well conserved proteins that contain a cluster of C2H2 domains. A CTCF homolog in *Drosophila*, dCTCF, is also often found at the TAD boundaries and in insulators [38, 39]. In model transgenic systems, dCTCF maintains long-distance interactions between the reporter gene promoter and the GAL4 activator [40, 41]. There is a homodimerization domain at the N-terminus of dCTCF (*Fig. 3A*); probably, this domain is necessary for maintaining the long-distance interactions between remote dCTCF binding sites [42]. Attempts to find a similar dimerization domain in vertebrate CTCFs have not been successful. *In vitro* experiments have demonstrated that the C-terminal part of one CTCF molecule binds to the cluster of the C2H2 zinc finger domains of another CTCF molecule [43]. However, the specificity of this interaction has not been proven.

According to a generally accepted model, the cohesin complex required for homologous chromosome pairing during cell division [44] binds to CTCF and participates in the maintenance of the specific long-distance interactions between its sites in interphase chromosomes (*Fig. 3B*). The region interacting with one of the cohesin complex proteins was mapped to the C-terminal domain of human CTCF [44].

The binding of CTCF to DNA, which is conserved even between insects and mammals, has been thoroughly studied in many higher eukaryotes [15, 45]. The C2H2 domains 4 to 7 of CTCF (*Fig. 3A*) participate in the binding to a core consensus site [46, 47]. Approximately 20% of the sites contain a second 10 bp motif that associates with the C2H2 domains 9 to 11 [47, 48]. This second motif, separated by 5 or 6 bp from the first, is supposed to increase the stability of CTCF binding to DNA.

The transcription factor CTCF is involved in many processes, such as embryonic development, the X chromosome inactivation in females, the regulation of the gene cluster recombination during the maturation of immunoglobulin genes, and the regulation of alternative splicing [34−37, 49−51]. CTCF was shown to interact with a large number of proteins (*Fig. 3A*), such as Smad [52], the core transcription factors TFII-I [53] and TAF-3 [54], the helicase p68 containing a DEAD-box domain [55], nucleophosmin, Kaiso [56], TFs YB1, YY1, and Oct4 [57−59], the CHD8 helicase [60], Su(z)12 (polycomb repressive complex 2 (PRC2) component) [61], the deacetylase complex component Sin3A [62], CENP-E [48], and many other proteins [49]. In most cases, individual C2H2 domains of CTCF participate in protein-protein interactions [49]. Probably, CTCF involvement in various processes (the activation and repression of transcription, the long-distance interactions, and TAD formation) is largely reliant on the formation of alternative complexes with partner proteins.

There is experimental evidence demonstrating that CTCF binds to numerous RNAs that modulate its activity. The RNA-binding domain of CTCF combines a
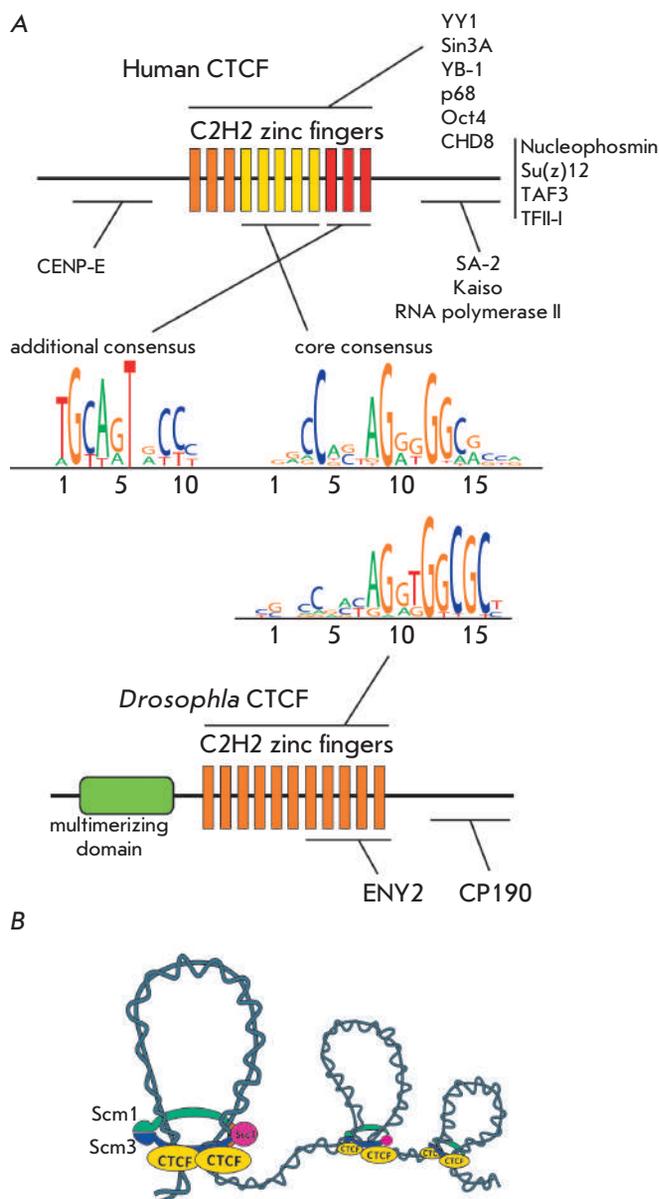
Fig. 3. Comparison of the structures and properties of the *Drosophila* and human CTCF proteins. (A) The domain structures of the *Drosophila* and human CTCF proteins. The domains involved in the site-specific DNA recognition and the protein-protein interactions are represented by thin horizontal lines. *Drosophila* [135] and human [46] CTCFs have similar consensus recognition sites. (B) The mechanism of the long-distance genomic interactions mediated by CTCF and cohesins.

portion of the C-terminal domain and two C2H2 domains (10 and 11), non-specifically recognizing RNA *in vitro* [63, 64]. It was suggested that interaction with some RNAs can increase the CTCF ability to form multimeric complexes [63] or to reduce the stability of

CTCF binding to DNA [11]. The CTCF activity is also regulated by various posttranslational modifications: poly(ADP)-ribosylation [65], phosphorylation [66], and sumoylation [67].

CTCF has been thoroughly studied and is an example of a TF containing a cluster of C2H2 domains and the unstructured N- and C-terminal regions. The majority of other C2H2 proteins have a similar structure, but their functions and properties have not yet been investigated. It may be assumed that some C2H2 proteins perform functions that are similar to those of CTCF. Interestingly, *Drosophila* mutants in the *ctcf* gene survive to the adult stage, which suggests that insect genomes contain other transcription factors that substitute for CTCF functions [42].

TRANSCRIPTION FACTORS WITH A CLUSTER OF C2H2 DOMAINS AND AN N-TERMINAL KRAB DOMAIN

About one-third of the human proteins with a cluster of C2H2 domains contain the Krüppel-associated box (KRAB) domain at the N-terminus (*Fig. 4A*) [68]. In total, there are 742 different human C2H2 proteins with the KRAB domain, which are encoded by 423 genes [69]. In this case, 384 genes are grouped into 25 chromosomal clusters and only 39 KRAB C2H2 proteins are encoded by single genes. KRAB domain proteins have been found only in tetrapods. The clustering on chromosomes and expansion within large taxa suggest that this family of genes originated through duplications that were preserved by evolutionary selection [70]. The KRAB domain consists of approximately 75 aa and may be structurally divided into two subdomains, A and B, that fold, as predicted, into two amphipathic α-helices (*Fig. 4B*). The KRAB A and KRAB B subdomains are always encoded by separate exons. Alternative splicing produces mRNAs that encode either only the KRAB A subdomain or simultaneously both subdomains, KRAB A and KRAB B, separated by a variable length spacer. Human KRAB proteins can contain from 2 up to 40 C2H2 domains. Unlike genes of other families, the C2H2 domains of KRAB proteins are most often encoded by one exon [71].

The KRAB C2H2 proteins are widely represented in the genomes of tetrapods, and many proteins are involved in the repression of transcription [70]. The versatile and well-studied mechanism of repression is associated with the recruitment of the KRAB-associated protein 1 (KAP-1), which is the only described cofactor of all studied KRAB proteins that represses transcription. The KRAB A domain directly interacts with KAP-1 that, in turn, serves as a platform for recruitment of the repressive complexes (*Fig. 4B*). The five amino acids (*Fig. 4B*) conserved in all the mammalian KRAB A domains (DV, at positions 6 and 7, and
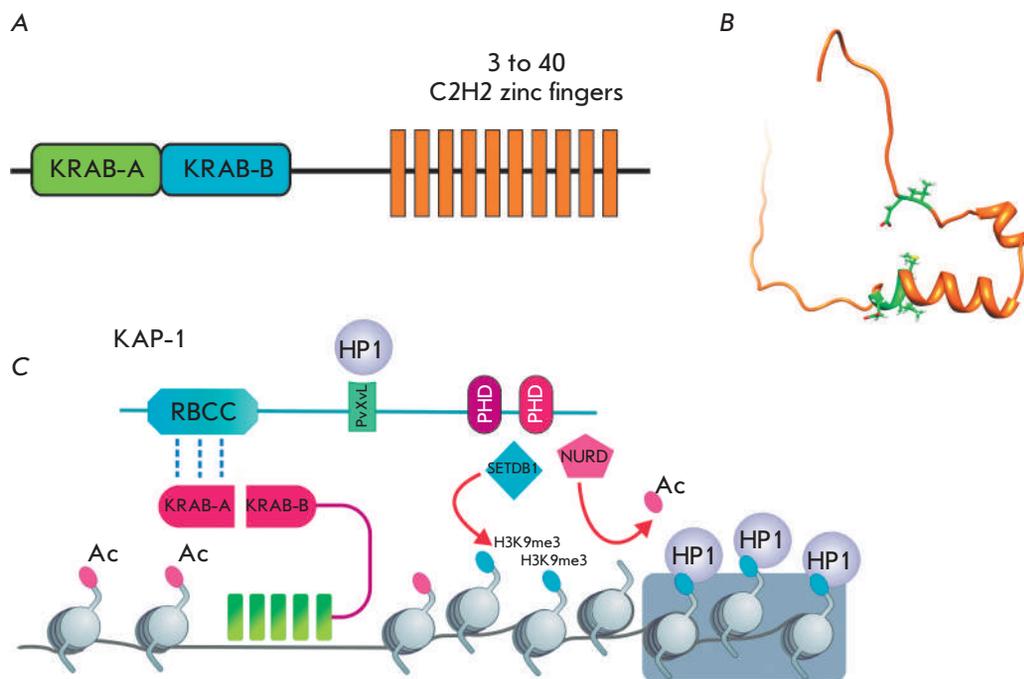
A

3 to 40
C2H2 zinc fingers

B

C

KAP-1

MLE, at positions 36−38) are needed for KAP-1 binding [72, 73]. The functional role of the KRAB B subdomain remains unexplored. It has been suggested that this domain increases the efficiency of recruitment of the KAP-1-dependent repressive complex [74]. At the N-terminus of KAP-1, there is the Ring finger/B box/coiled-coil (RBCC) domain that enables binding of KAP-1 to the KRAB domain. The central part of KAP-1 contains a hydrophobic pentapeptide that interacts with the Chromo-Shadow (CS) domain of the HP1 protein. At the C-terminus of the protein, there are two PHD domains that recruit the complexes involved in the deacetylation (NURD) and methylation (SETDB1) of histones [70, 75−77]. The repression initiated at the KRAB C2H2 protein binding site can spread tens of thousands of nucleotides across the surrounding regions of the genome through the successive introduction of the H3K9me3 modification and the subsequent binding to it of the HP1 repressor [78−80]. The KAP-1 expression peak is at the early embryonic stages, and the transcriptional repression by KRAB C2H2 proteins is critical for early embryonic development. At later stages, the somatic cell repression can be maintained by epigenetic mechanisms, with no direct involvement of the KRAB C2H2 proteins [81, 82].

The majority of the KRAB C2H2 proteins are species- and genus-specific. In some vertebrates, such as birds, lizards, and frogs, the KRAB A domain has the multiple amino acid substitutions required for the interaction with KAP-1 [31, 72]. This may be explained either by the fact that the KRAB domain in these vertebrates performs other functions or by the fact that these classes of vertebrates have a modified KAP-1 that retains its ability to interact with the KRAB domain. In general, the evolutionary analysis of the conservatism of KRAB C2H2 proteins has demonstrated that KRAB C2H2 gene families formed independently in each class of vertebrates, which confirms the high evolutionary emergence rate of new genes of this class.

There are only three known KRAB C2H2 genes that belong to a single cluster and are common to all studied vertebrate species. These genes encode C2H2 proteins containing a structurally modified KRAB domain that no longer binds to the KAP-1 repressor and is involved in transcriptional stimulation [31, 83, 84]. Of particular interest is the highly conserved gene *Meisetz* (PRDM9) that codes for not only a modified KRAB domain, but also the SET domain [85, 86]. The SET and KRAB domains jointly recruit the H3K4-methyltransferase that is responsible for trimethylation of histone H3 on lysine 4 (H3K4me3). The H3K4me3 modification at the promoter region usually correlates with an active transcription. A bioinformatic analysis demonstrated that a portion of the KRAB domain encoded by the *Meisetz* gene is homologous to the KRI motif present in the genomes of all well-studied eukaryotes, including arabidopsis, rice, fungi, and yeast [31, 86]. The widespread occurrence of the KRI motif suggests that the KRAB domain of the Meisetz protein might have originated from this motif by addition of several amino acid resi-

dues. The KRAB domain might have acquired its repressor functions through random mutations that allowed the repressor to bind to KAP-1, which was preserved during evolution.

It was experimentally demonstrated that TFs of the KRAB C2H2 class in vertebrates play an important role in various processes of embryonic development, cell differentiation and proliferation, and the regulation of the cell cycle and apoptosis [70, 73]. The binding sites for KRAB C2H2 proteins correlate with the open (nucleosome free) chromatin regions, something that is explained by the binding with the active regulatory regions of the genes [87]. Whole-genome studies have demonstrated that KRAB C2H2 proteins bind to the enhancers and promoters of genes and can activate transcription in some cases [88−90]. The ability of KRAB C2H2 proteins to activate transcription should correlate with the suppression of interaction between the KRAB domain and KAP-1. The mechanism of this suppression still remains unexplored but is probably associated with reversible modifications of the amino acid residues of the KRAB domain. An important role may be played by the C2H2 domains that are potentially capable of recruiting individual TFs and whole complexes that positively/negatively regulate transcription.

The above-mentioned Meisetz protein (PRDM9), which is expressed only in mammalian gonads, plays an interesting role [91]. Most mammalian recombination hotspots were found to contain a potential PRDM9 binding site [92]. Rapid evolutionary changes in the number and primary structure of C2H2 domains led to the binding of PRDM9 to different nucleotide DNA sequences in different mammals [91, 93−96]. The binding of PRDM9 results in the formation of a nucleosome-depleted region and H3K4me3 modification of the surrounding nucleosomes [97]. The SPO11 complex inducing double-strand breaks is supposed to simultaneously recognize the histone H3K4me3 mark and directly bind to PRDM9 [98].

Recently, a new functional role played by KRAB C2H2 proteins in the repression of foreign DNA transcription, primarily, of endogenous retroviruses and the mobile elements LINE and SINE, was discovered [79, 87, 99, 100]. Mobile elements constitute a significant part of the mammalian genome, and repression of their transcription is essential [101]. Different KRAB C2H2 proteins bind to the regulatory regions of mobile elements and those of certain retroviruses, and they induce their epigenetic repression. There is a hypothesis that holds that the newly appeared KRAB C2H2 proteins have been preserved by evolutionary selection, because they play a critical role in the suppression of the expression of new mobile elements, while

the more conserved KRAB C2H2 proteins participate in the regulation of the expression of cellular genes [79].

Another explanation for the rapid evolution of the genes that encode KRAB C2H2 proteins may be their key role in the control of the expression of the genes that determine the development of the nervous [102] and circulatory [103] systems of mammals. For example, many genes encoding the KRAB C2H2 proteins specific to humans and primates are actively transcribed in the brain [102]. However, there is no direct correlation between the number of KRAB C2H2 genes and the level of organism complexity. For example, the number of KRAB C2H2 genes in opossum is double that in humans [31]. It is hoped that the emergence of new technologies for specific antibody generation, whole-genome analysis, and single gene mutations using the CRISPR/Cas9 system will soon clarify the functional role played by KRAB C2H2 proteins.

TRANSCRIPTION FACTORS WITH A CLUSTER OF C2H2 DOMAINS AND AN N-TERMINAL SCAN DOMAIN

The SCAN domain was first described in human ZNF174 TF [114] (*Fig. 5A*). Subsequently, proteins with these domains were found in some other classes of vertebrates [104]. In humans, mice, and cows, 71, 38, and 28 SCAN C2H2 proteins were found, respectively [94]. Genes encoding the SCAN proteins usually occur in the genome as small (two to seven) clusters [104]. The SCAN domains in clusters have a higher degree of homology among themselves, which suggests that they emerged through gene duplication and a subsequent adaptive evolution. Approximately half of the genes encode simultaneously the SCAN and KRAB domains and usually occur in large clusters, along with the genes encoding only the KRAB C2H2 proteins (*Fig. 5A*) [27, 94].

The SCAN domain structure is highly similar to that of the C-terminal domains of the human immunodeficiency virus capsid protein [105] and the Gag protein from the family of *Ty3/gypsy* retrotransposons [27]. Based on such data, it was suggested that SCAN domains initially originated from the retrovirus capsid proteins in the lower vertebrates; then, during evolution, this domain acquired a new function in TFs-containing clusters of the C2H2 domains [106]. The KRAB domain in combination with the SCAN domain is present in mammals and lizards, but it is absent in chicken and frog.

To date, the spatial structures of the SCAN domains of the proteins Zfp206 [107], PEG3 [108], ZNF24, ZNF174 [105], and MZF-1 [109], which have a high degree of homology, have been resolved (*Fig. 5B*). The features of the spatial structure may be illustrated by the example of the SCAN domain of Zfp206 [107],
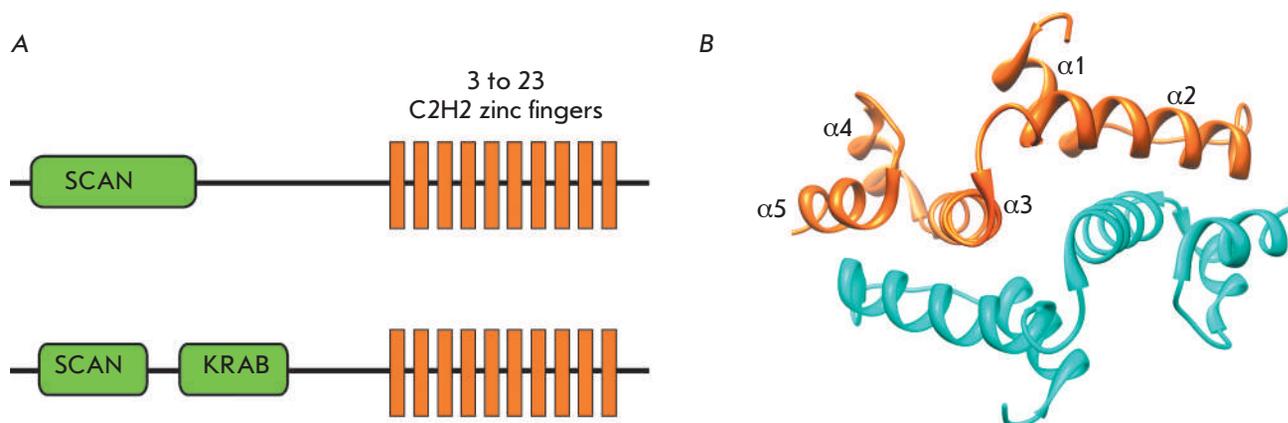
Fig. 5. The structure and properties of the SCAN domain. (A) A typical domain structure of the SCAN C2H2 and SCAN KRAB C2H2 proteins. (B) The crystal structure of a SCAN domain dimer from the Zfp206 protein [110].

which exists as an antiparallel homodimer. Each monomer in the homodimer consists of five α-helices. The core of the inner homodimer surface is formed by packing of the second α-helix of one monomer against the third and fifth α-helices of the opposing monomer and *vice versa*. The N-terminal first α-helix provides additional contacts of one monomer with the third α-helix of the opposing monomer (*Fig. 5B*). All SCAN domains can form homodimers, but only some SCAN domains are able to form heterodimers [104, 111–113]. The first α-helix of the SCAN domain has the greatest variability in the hydrophobic amino acid sequence and is considered as a potential candidate that determines the formation of heterodimers from different SCAN domains. For example, the SCAN domain of Zfp206 was shown to be able to form a heterodimer with a similar domain in Zfp110 [107]. The replacement of the first α-helix with an α-helix of a heterologous SCAN domain of ZNF174 or its removal results in a loss of the ability of these SCAN domains to form heterodimers.

The CRAB A domain is known to recruit the repressive complex, whereas there is no evidence of SCAN domain effect on transcription [104, 111]. There is only fragmentary data on the functional role of SCAN C2H2 TFs. For example, human ZNF263 TF containing the SCAN and KRAB domains and 9 C2H2 domains predominantly binds to the promoter regions and is able to participate in both the activation and repression of transcription [89]. Another member of the family, the ZNF658 protein, also contains the SCAN and KRAB domains and is involved in the activation of the expression of the rRNA genes that are transcribed by RNA polymerase I [115]. Probably, the main function of SCAN proteins may be related to their ability to form homo- and heterodimers between the SCAN domains.

TRANSCRIPTION FACTORS WITH A CLUSTER OF C2H2 DOMAINS AND AN N-TERMINAL ZINC FINGER-ASSOCIATED DOMAIN

The zinc finger-associated domain (ZAD) (*Fig. 6A*) is found at the N-terminus of the C2H2 proteins of many arthropods [28]. In vertebrates, only one protein containing an N-terminal structure similar to the ZAD has been found [116]. In the genomes of *Anopheles gambiae*, *Drosophila melanogaster*, and *Apis mellifera* (honey bee), the 147, 98, and 29 ZAD C2H2 proteins, respectively, were found [116], whereas only four genes encoding ZAD-like domains were found in crustaceans (*Daphnia pulex*). Usually, genes encoding highly homologous ZADs form small clusters. It is suggested that these genes originated from multiple duplications of original copies and then were preserved through positive selection [28, 116]. Probably, the evolutionary process was very fast, since obvious homologs were found only for a few ZAD C2H2 proteins in distant *Drosophila* species [116].

The ZAD size varies between 71 and 97 aa. The multiple alignments of the sequences of 32 family members demonstrate that this domain consists of four conserved blocks linked by regions of varying lengths [28]. A distinctive feature of ZADs is the presence of two invariant cysteine pairs coordinating a zinc ion.

To date, the crystal structure of only one ZAD from the Grauzone protein (Grau) has been resolved [117], which can serve as a prototype for all ZAD structures. The N-terminal ZAD portion forms a globule around the zinc ion, and the C-terminal stem is formed by a long α-helix 2 (α2) that comprises almost one-third of all the amino acids in the ZAD. The ZAD folding largely depends on the coordination of two cysteine pairs (separated by about 50 aa) by the zinc ion, which results in
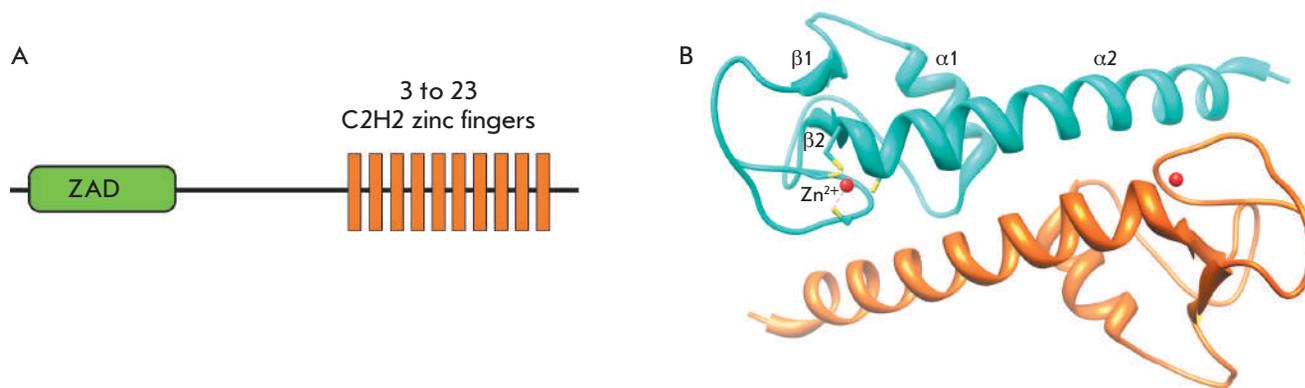
Fig. 6. The structure and properties of the ZAD. (A) A typical domain structure of ZAD C2H2 proteins. (B) The crystal structure of a ZAD dimer from the Grau protein [117].

drawing of the β2-α2 regions and the N-terminus to the domain center.

In a crystal, two ZAD molecules are associated as an antiparallel dimer. Most of the amino acid residues that are conserved in the ZAD family [28] form contacts between the two subunits. The ZAD of Grau has a negatively charged surface, suggesting that the domain is unable to bind to DNA [117]. It has been suggested that the main function of the ZAD is to form homodimers of the C2H2 proteins [118]. ZADs also participate in the regulation of the nuclear localization of some of the ZAD C2H2 proteins [119].

Proteins with ZADs account for approximately one-third of the total number of proteins with C2H2 clusters and one-tenth of all TFs in the *D. melanogaster* genome [28]. To date, the functions of only a small fraction of ZAD C2H2 TFs have been studied. The majority of ZAD C2H2 proteins are expressed during oogenesis and at early embryogenesis [116]. The results of several studies point to an important functional role for ZAD C2H2 proteins in *Drosophila* development.

The Motif 1 Binding Protein (M1BP) is expressed at a high level in all tissues and at all stages of *Drosophila* development and is a key factor in the organization of the architecture of more than 2,000 *Drosophila* promoters with a characteristic motif (T/C)GG(T/C)CA-CACTG [120].

In transgenic *Drosophila* lines, three ZAD C2H2 proteins (Pita, ZIPIC, and Zw5) exhibit the properties of insulator/architectural proteins: they block the interaction between an enhancer and a promoter and maintain long-distance interactions [118, 121−124]. The ZADs of these proteins form only stable homodimers [118]. Interestingly, the DNA fragments containing binding sites for different ZAD C2H2 proteins cannot maintain long-distance interactions, which suggests a

key role for the ZAD dimerization in the formation of specific contacts between distant chromatin regions. Indeed, the ZAD of ZIPIC is required for the maintenance of long-distance interactions between the GAL4 activator and the reporter gene promoter in yeast *Saccharomyces cerevisiae* [118]. As in the case of M1BP, binding sites for the proteins ZIPIC, Pita, and Zw5 are predominantly located close to the transcription starts [118, 125], which suggests an architectural function for these proteins in the promoter organization. Null mutations in the genes encoding the Pita and Zw5 proteins lead to late embryonic and early larval lethality, which indicates that there is an important role for these proteins in early *Drosophila* development [121, 126].

The Grau protein is expressed at all stages of *Drosophila* development; it is found in the nuclei of the nurse and follicular cells surrounding the oocyte [127, 128]. Mutations in this gene lead to oogenesis arrest at the meiosis II stage, which is related to the role of Grau in the activation of the promoter of the *cortex* gene that regulates meiosis in oocytes [127, 128]. The Serendipity delta protein (Sry δ) binds to the promoter of the *bicoid* gene, which plays a key role in early embryogenesis, and stimulates transcription of the gene [129]. Null mutations in the *sry δ* gene manifest themselves as embryonic lethals, which indicates the significance of Sry δ in the early development of *Drosophila* [130].

The Trade Embargo protein (Trem) is expressed mainly in *Drosophila* germ cells and probably performs a function similar to that of the PRDM9 protein in mammals [95, 97, 131]. Trem specifies the binding sites for the Mei-P22 protein that is involved in the induction of meiotic chromosomal breaks [131]. Mei-P22 and its partner, Mei-W68, participate in the formation of the double-strand breaks that initiate crossing-over in meiosis [132−134]. According to the model, Trem, to-

gether with its partners, creates the open chromatin regions that recruit, through specific protein-protein interactions, the Mei-P22/Mei-W68 complex, inducing double-strand breaks [131].

In general, the available data demonstrate that ZAD C2H2 TFs play an important role in the organization of the structure and functional activity of promoters, the recruitment of protein complexes, and the formation of the chromosomal architecture.

CONCLUSION

At present, there are many unresolved issues related to the regulation of gene transcription, the organization of the structure of regulatory elements, and the mechanisms of long-distance interactions. It is also quite obvious that the vertebrate CTCF cannot be the sole and key DNA-binding protein that determines the architecture of vertebrate chromosomes [7].

Unlike the well-studied TFs of other classes, C2H2 proteins specifically bind to long DNA sequences reaching several tens of base pairs. The C2H2 proteins can effectively bind to DNA as monomers, unlike most other TFs that bind to short palindromic sequences as homo- or heterodimers. Some C2H2 domains in combination with the unstructured regions of C2H2 proteins can enable a variety of interactions with protein complexes and individual TFs and RNAs. Therefore, C2H2 proteins may be considered as promising candidates for the role of organizers of the architecture of regulatory elements, such as promoters, enhancers, insulators, and silencers. Unfortunately, the available experimental evidence is insufficient in order to confirm the validity of this assumption for vertebrates. On the other hand, the well-studied CTCF protein of vertebrates has a number of properties (the specific binding to a DNA site, the formation of open chromatin regions, the re-cruitment of protein complexes, and the organization of long-distance interactions) that may be extrapolated to other C2H2 proteins.

Finally, many C2H2 proteins have domains that are capable of homodimerization. Interestingly, in arthropods and vertebrates, there was an expansion of different domains: ZAD and SCAN, respectively. The main common property of the ZAD and SCAN domains is their ability to preferentially form homodimers. Homodimerizing ZADs of the three ZAD C2H2 proteins (Pita, ZIPIC, and Zw5) were demonstrated to determine the specificity of long-distance interactions [118]. Probably, other ZAD C2H2 proteins possess similar properties. So far, only some data on the role of SCAN C2H2 proteins in the organization of active promoters in vertebrates has been obtained. Apart from the ZAD and SCAN domains, C2H2 proteins may have other domains capable of multimerization: e.g., an N-terminal domain of the *Drosophila* dCTCF protein [42].

Therefore, the available fragmentary data already allow us to suggest a model where the C2H2 proteins act as the messengers in the transfer of information from the nucleotide sequence of the regulatory elements (promoters, enhancers, and silencers) to the protein complexes that determine the properties of regulatory elements. It is assumed that investigation of individual members of this extensive class of TFs, the elucidation of the functional roles of the ZAD, SCAN, and KRAB domains, and the identification of new partner proteins and new dimerization domains will allow us to evaluate the real contribution of C2H2 proteins to the formation of the chromosomal architecture and the structure of regulatory elements. ●

REFERENCES

1. Dixon J.R., Selvaraj S., Yue F., Kim A., Li Y., Shen Y., Hu M., Liu J.S., Ren B. // Nature. 2012. V. 485. P. 376−380.
2. 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. P. 381−385.
3. Razin S.V., Gavrilov A.A., Vassetzky Y.S., Ulianov S.V. // Transcription. 2016. V. 7. P. 84−90.
4. Dekker J., Rippe K., Dekker M., Kleckner N. // Science. 2002. V. 295. P. 1306−1311.
5. 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. P. 1665−1680.
6. Lupianez D.G., Kraft K., Heinrich V., Krawitz P., Brancati F., Klopocki E., Horn D., Kayserili H., Opitz J.M., Laxova R., et al. // Cell. 2015. V. 161. P. 1012−1025.
7. Maksimenko O., Georgiev P. // Front. Genet. 2014. V. 5. P. 28.
8. Bouwman B.A., de Laat W. // Genome Biol. 2015. V. 16. P. 154.
9. Dekker J., Mirny L. // Cell. 2016. V. 164. P. 1110−1121.
10. Merkenschlager M., Nora E.P. // Ann. Rev. Genom. Human Genet. 2016. V. 17. P. 17−43.
11. Ghirlando R., Felsenfeld G. // Genes Dev. 2016. V. 30. P. 881−891.
12. Handoko L., Xu H., Li G., Ngan C.Y., Chew E., Schnapp M., Lee C.W., Ye C., Ping J.L., Mulawadi F., et al. // Nat. Genet. 2011. V. 43. P. 630−638.
13. Gibcus J.H., Dekker J. // Mol Cell. 2013. V. 49. P. 773−782.
14. Zuin J., Dixon J.R., van der Reijden M.I., Ye Z., Kolovos P., Brouwer R.W., van de Corput M.P., van de Werken H.J., Knoch T.A., van IJcken W.F., et al. // Proc. Natl. Acad. Sci. USA. 2014. V. 111. P. 996−1001.
15. Heger P., Marin B., Bartkuhn M., Schierenberg E., Wiehe T. // Proc. Natl. Acad. Sci. USA. 2012. V. 109. P. 17507−17512.

16. Razin S.V., Borunova V.V., Maksimenko O.G., Kantidze O.L. // Biochemistry (Mosc.). 2012. V. 77. P. 277−288.

17. Iuchi S. // Cell. Mol. Life Sci. 2001. V. 58. P. 625−635.

18. Pearson R.C., Funnell A.P., Crossley M. // IUBMB Life. 2011. V. 63. P. 86−93.

19. Swamynathan S.K. // Hum. Genomics. 2010. V. 4. P. 263−270.

20. Pavletich N.P., Pabo C.O. // Science. 1991. V. 252. P. 809−817.

21. Wolfe S.A., Nekludova L., Pabo C.O. // Annu. Rev. Biophys. Biomol. Struct. 2000. V. 29. P. 183−212.

22. Garton M., Najafabadi H.S., Schmitges F.W., Radovani E., Hughes T.R., Kim P.M. // Nucl. Acids Res. 2015. V. 43. P. 9147−9157.

23. Persikov A.V., Singh M. // Nucl. Acids Res. 2014. V. 42. P. 97−108.

24. Vandevenne M., Jacques D.A., Artuz C., Nguyen C.D., Kwan A.H., Segal D.J., Matthews J.M., Crossley M., Guss J.M., Mackay J.P. // J. Biol. Chem. 2013. V. 288. P. 10616−10627.

25. Brayer K.J., Segal D.J. // Cell Biochem. Biophys. 2008. V. 50. P. 111−131.

26. Brayer K.J., Kulshreshtha S., Segal D.J. // Cell Biochem. Biophys. 2008. V. 51. P. 9−19.

27. Emerson R.O., Thomas J.H. // J. Virol. 2011. V. 85. P. 12043−12052.

28. Chung H.R., Schafer U., Jackle H., Bohm S. // EMBO Repts. 2002. V. 3. P. 1158−1162.

29. Perez-Torrado R., Yamada D., Defossez P.A. // Bioessays. 2006. V. 28. P. 1194−1202.

30. Stogios P.J., Chen L., Prive G.G. // Protein Sci. 2007. V. 16. P. 336−342.

31. Liu H., Chang L.H., Sun Y., Lu X., Stubbs L. // Genome Biol. Evol. 2014. V. 6. P. 510−525.

32. Lobanenkov V.V., Nicolas R.H., Adler V.V., Paterson H., Klenova E.M., Polotskaja A.V., Goodwin G.H. // Oncogene. 1990. V. 5. P. 1743−1753.

33. Bell A.C., West A.G., Felsenfeld G. // Cell. 1999. V. 98. P. 387−396.

34. Chaumeil J., Skok J.A. // Curr. Opin. Immunol. 2012. V. 24. P. 153−159.

35. Herold M., Bartkuhn M., Renkawitz R. // Development. 2012. V. 139. P. 1045−1057.

36. Holwerda S., de Laat W. // Front. Genet. 2012. V. 3. P. 217.

37. Merkenschlager M., Odom D.T. // Cell. 2013. V. 152. P. 1285−1297.

38. Sexton T., Yaffe E., Kenigsberg E., Bantignies F., Leblanc B., Hoichman M., Parrinello H., Tanay A., Cavalli G. // Cell. 2012. V. 148. P. 458−472.

39. Pirrotta V., Li H.B. // Curr. Opin. Genet. Dev. 2012. V. 22. P. 101−109.

40. Kyrchanova O., Toshchakov S., Podstreshnaya Y., Parshikov A., Georgiev P. // Mol. Cell. Biol. 2008. V. 28. P. 4188−4195.

41. Kyrchanova O., Ivlieva T., Toshchakov S., Parshikov A., Maksimenko O., Georgiev P. // Nucl. Acids Res. 2011. V. 39. P. 3042−3052.

42. 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.

43. Pant V., Kurukuti S., Pugacheva E., Shamsuddin S., Mariano P., Renkawitz R., Klenova E., Lobanenkov V., Ohlsson R. // Mol. Cell Biol. 2004. V. 24. P. 3497−3504.

44. Xiao T., Wallace J., Felsenfeld G. // Mol. Cell. Biol. 2011. V. 31. P. 2174−2183.

45. Vietri Rudan M., Barrington C., Henderson S., Ernst C., Odom D. T., Tanay A., Hadjur S. // Cell Rep. 2015. V. 10. P. 1297−1309.

46. 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. P. 335−348.

47. Nakahashi H., Kwon K.R., Resch W., Vian L., Dose M., Stavreva D., Hakim O., Pruett N., Nelson S., Yamane A., et al. // Cell Rep. 2013. V. 3. P. 1678−1689.

48. Xiao T., Wongtrakoongate P., Trainor C., Felsenfeld G. // Cell Rep. 2015. V. 12. P. 1704−1714.

49. Ohlsson R., Bartkuhn M., Renkawitz R. // Chromosoma. 2010. V. 119. P. 351−360.

50. 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. P. 796−801.

51. Proudhon C., Hao B., Raviram R., Chaumeil J., Skok J.A. // Adv. Immunol. 2015. V. 128. P. 123−182.

52. Bergstrom R., Savary K., Moren A., Guibert S., Heldin C. H., Ohlsson R., Moustakas A. // J. Biol. Chem. 2010. V. 285. P. 19727−19737.

53. 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. P. E677−686.

54. Liu Z., Scannell D.R., Eisen M.B., Tjian R. // Cell. 2011. V. 146. P. 720−731.

55. Yao H., Brick K., Evrard Y., Xiao T., Camerini-Otero R.D., Felsenfeld G. // Genes Dev. 2010. V. 24. P. 2543−2555.

56. Defossez P.A., Kelly K.F., Filion G.J., Perez-Torrado R., Magdinier F., Menoni H., Nordgaard C.L., Daniel J.M., Gilson E. // J. Biol. Chem. 2005. V. 280. P. 43017−43023.

57. Klenova E., Scott A.C., Roberts J., Shamsuddin S., Lovejoy E.A., Bergmann S., Bubb V.J., Royer H.D., Quinn J.P. // J. Neurosci. 2004. V. 24. P. 5966−5973.

58. Donohoe M.E., Zhang L.F., Xu N., Shi Y., Lee J.T. // Mol. Cell. 2007. V. 25. P. 43−56.

59. Donohoe M.E., Silva S.S., Pinter S.F., Xu N., Lee J.T. // Nature. 2009. V. 460. P. 128−132.

60. Ishihara K., Oshimura M., Nakao M. // Mol. Cell. 2006. V. 23. P. 733−742.

61. Li T., Hu J.F., Qiu X., Ling J., Chen H., Wang S., Hou A., Vu T.H., Hoffman A.R. // Mol. Cell. Biol. 2008. V. 28. P. 6473−6482.

62. Lutz M., Burke L.J., Barreto G., Goeman F., Greb H., Arnold R., Schultheiss H., Brehm A., Kouzarides T., Lobanenkov V., et al. // Nucl. Acids Res. 2000. V. 28. P. 1707−1713.

63. Saldana-Meyer R., Gonzalez-Buendia E., Guerrero G., Narendra V., Bonasio R., Recillas-Targa F., Reinberg D. // Genes Dev. 2014. V. 28. P. 723−734.

64. 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. P. 361−375.

65. Guastafierro T., Catizone A., Calabrese R., Zampieri M., Martella O., Bacalini M.G., Reale A., Di Girolamo M., Miccheli M., Farrar D., et al. // Biochem. J. 2013. V. 449. P. 623−630.

66. Klenova E.M., Chernukhin I.V., El-Kady A., Lee R.E., Pugacheva E.M., Loukinov D.I., Goodwin G.H., Delgado D., Filippova G.N., Leon J., et al. // Mol. Cell. Biol. 2001. V. 21. P. 2221−2234.

67. MacPherson M.J., Beatty L.G., Zhou W., Du M., Sadowski P.D. // Mol. Cell. Biol. 2009. V. 29. P. 714−725.

68. Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O., Martial J.A. // Proc. Natl. Acad. Sci. USA. 1991. V. 88. P. 3608−3612.

69. Huntley S., Baggott D.M., Hamilton A.T., Tran-Gyamfi M., Yang S., Kim J., Gordon L., Branscomb E., Stubbs L. // Genome Res. 2006. V. 16. P. 669−677.

70. Lupo A., Cesaro E., Montano G., Zurlo D., Izzo P., Costanzo P. // Curr. Genomics. 2013. V. 14. P. 268−278.

71. Shannon M., Kim J., Ashworth L., Branscomb E., Stubbs L. // DNA Seq. 1998. V. 8. P. 303−315.

72. Margolin J.F., Friedman J.R., Meyer W.K., Vissing H., Thiesen H. J., Rauscher F. J., 3rd // Proc. Natl. Acad. Sci. USA. 1994. V. 91. P. 4509−4513.

73. Urrutia R. // Genome Biol. 2003. V. 4. P. 231.

74. Vissing H., Meyer W.K., Aagaard L., Tommerup N., Thiesen H.J. // FEBS Lett. 1995. V. 369. P. 153−157.

75. Lechner M.S., Begg G.E., Speicher D.W., Rauscher F.J., 3rd // Mol. Cell. Biol. 2000. V. 20. P. 6449−6465.

76. Schultz D.C., Ayyanathan K., Negorev D., Maul G.G., Rauscher F.J., 3rd // Genes Dev. 2002. V. 16. P. 919−932.

77. Sripathy S.P., Stevens J., Schultz D.C. // Mol. Cell. Biol. 2006. V. 26. P. 8623−8638.

78. Groner A.C., Meylan S., Ciuffi A., Zangger N., Ambrosini G., Denervaud N., Bucher P., Trono D. // PLoS Genet. 2010. V. 6. P. e1000869.

79. Wolf G., Macfarlan T.S. // Cell. 2015. V. 163. P. 30−32.

80. Ying L., Lin J., Qiu F., Cao M., Chen H., Liu Z., Huang Y. // FEBS J. 2015. V. 282. P. 174−182.

81. Wiznerowicz M., Jakobsson J., Szulc J., Liao S., Quazzola A., Beermann F., Aebischer P., Trono D. // J. Biol. Chem. 2007. V. 282. P. 34535−34541.

82. Rowe H.M., Friedli M., Offner S., Verp S., Mesnard D., Marquis J., Aktas T., Trono D. // Development. 2013. V. 140. P. 519−529.

83. Okumura K., Sakaguchi G., Naito K., Tamura T., Igarashi H. // Nucl. Acids Res. 1997. V. 25. P. 5025−5032.

84. Conroy A.T., Sharma M., Holtz A.E., Wu C., Sun Z., Weigel R.J. // J. Biol. Chem. 2002. V. 277. P. 9326−9334.

85. Hayashi K., Yoshida K., Matsui Y. // Nature. 2005. V. 438. P. 374−378.

86. Birtle Z., Ponting C.P. // Bioinformatics. 2006. V. 22. P. 2841−2845.

87. Najafabadi H.S., Albu M., Hughes T.R. // Bioinformatics. 2015. V. 31. P. 2879−2881.

88. Losson R., Nielsen A.L. // Biochim. Biophys. Acta. 2010. V. 1799. P. 463−468.

89. Frietze S., Lan X., Jin V.X., Farnham P.J. // J. Biol. Chem. 2010. V. 285. P. 1393−1403.

90. Yang L., Wang H., Kornblau S.M., Graber D.A., Zhang N., Matthews J.A., Wang M., Weber D.M., Thomas S.K., Shah J.J., et al. // Oncogene. 2011. V. 30. P. 1329−1340.

91. Parvanov E.D., Petkov P.M., Paigen K. // Science. 2010. V. 327. P. 835.

92. Smagulova F., Gregoretti I.V., Brick K., Khil P., Camerini-Otero R.D., Petukhova G.V. // Nature. 2011. V. 472. P. 375−378.

93. Oliver P.L., Goodstadt L., Bayes J.J., Birtle Z., Roach K.C., Phadnis N., Beatson S.A., Lunter G., Malik H.S., Ponting C.P. // PLoS Genet. 2009. V. 5. P. e1000753.

94. Thomas J.H., Emerson R.O., Shendure J. // PLoS One. 2009. V. 4. P. e8505.

95. Baudat F., Buard J., Grey C., Fledel-Alon A., Ober C., Przeworski M., Coop G., de Massy B. // Science. 2010. V. 327. P. 836−840.

96. Berg I.L., Neumann R., Lam K.W., Sarbajna S., Odenthal-Hesse L., May C.A., Jeffreys A.J. // Nat. Genet. 2010. V. 42. P. 859−863.

97. Baker C.L., Walker M., Kajita S., Petkov P.M., Paigen K. // Genome Res. 2014. V. 24. P. 724−732.

98. Brick K., Smagulova F., Khil P., Camerini-Otero R.D., Petukhova G.V. // Nature. 2012. V. 485. P. 642−645.

99. Castro-Diaz N., Ecco G., Coluccio A., Kapopoulou A., Yazdanpanah B., Friedli M., Duc J., Jang S.M., Turelli P., Trono D. // Genes Dev. 2014. V. 28. P. 1397−1409.

100. Turelli P., Castro-Diaz N., Marzetta F., Kapopoulou A., Raclot C., Duc J., Tieng V., Quenneville S., Trono D. // Genome Res. 2014. V. 24. P. 1260−1270.

101. Richardson S.R., Doucet A.J., Kopera H.C., Moldovan J.B., Garcia-Perez J.L., Moran J.V. // Microbiol Spectr. 2015. V. 3. P. MDNA3-0061-2014.

102. Nowick K., Gernat T., Almaas E., Stubbs L. // Proc. Natl. Acad. Sci. USA. 2009. V. 106. P. 22358−22363.

103. Barde I., Rauwel B., Marin-Florez R.M., Corsinotti A., Laurenti E., Verp S., Offner S., Marquis J., Kapopoulou A., Vanicek J., et al. // Science. 2013. V. 340. P. 350−353.

104. Sander T.L., Haas A.L., Peterson M.J., Morris J.F. // J. Biol. Chem. 2000. V. 275. P. 12857−12867.

105. Ivanov D., Stone J.R., Maki J.L., Collins T., Wagner G. // Mol. Cell. 2005. V. 17. P. 137−143.

106. Sander T.L., Stringer K.F., Maki J.L., Szauter P., Stone J.R., Collins T. // Gene. 2003. V. 310. P. 29−38.

107. Liang Y., Huimei Hong F., Ganesan P., Jiang S., Jauch R., Stanton L.W., Kolatkar P.R. // Nucl. Acids Res. 2012. V. 40. P. 8721−8732.

108. Rimsa V., Eadsforth T.C., Hunter W.N. // PLoS One. 2013. V. 8. P. e69538.

109. 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. P. 137−147.

110. Liang Y., Choo S.H., Rossbach M., Baburajendran N., Palasingam P., Kolatkar P.R. // Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 2012. V. 68. P. 443−447.

111. Williams A.J., Blacklow S.C., Collins T. // Mol. Cell. Biol. 1999. V. 19. P. 8526−8535.

112. 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. P. 17173−17179.

113. Porsch-Ozcurumez M., Langmann T., Heimerl S., Borsukova H., Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G. // J. Biol. Chem. 2001. V. 276. P. 12427−12433.

114. Stone J.R., Maki J.L., Blacklow S.C., Collins T. // J. Biol. Chem. 2002. V. 277. P. 5448−5452.

115. Ogo O.A., Tyson J., Cockell S.J., Howard A., Valentine R.A., Ford D. // Mol. Cell. Biol. 2015. V. 35. P. 977−987.

116. Chung H.R., Lohr U., Jackle H. // Mol. Biol. Evol. 2007. V. 24. P. 1934−1943.

117. Jauch R., Bourenkov G.P., Chung H.R., Urlaub H., Reidt U., Jackle H., Wahl M.C. // Structure. 2003. V. 11. P. 1393−1402.

118. Zolotarev N., Fedotova A., Kyrchanova O., Bonchuk A., Penin A.A., Lando A.S., Eliseeva I.A., Kulakovskiy I.V., Maksimenko O., Georgiev P. // Nucl. Acids Res. 2016. V. 44. P. 7228−7241.

119. Zolotarev N.A., Maksimenko O.G., Georgiev P.G., Bonchuk A.N. // Acta Naturae. 2016. V. 8. №3. P. 97−102.

120. Li J., Gilmour D.S. // EMBO J. 2013. V. 32. P. 1829−1841.

121. Gaszner M., Vazquez J., Schedl P. // Genes Dev. 1999. V. 13. P. 2098−2107.

122. Blanton J., Gaszner M., Schedl P. // Genes Dev. 2003. V. 17. P. 664−675.

123. 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. P. 2188−2198.

124. Kyrchanova O., Leman D., Parshikov A., Fedotova A., Studitsky V., Maksimenko O., Georgiev P. // PLoS One. 2013. V. 8. P. e62690.

125. 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. P. 89–99.

126. Laundrie B., Peterson J.S., Baum J.S., Chang J.C., Fileppo D., Thompson S.R., McCall K. // Genetics. 2003. V. 165. P. 1881–1888.

127. Chen B., Harms E., Chu T., Henrion G., Strickland S. // Development. 2000. V. 127. P. 1243–1251.

128. Chu T., Henrion G., Haegeli V., Strickland S. // Genesis. 2001. V. 29. P. 141–152.

129. Payre F., Noselli S., Lefrere V., Vincent A. // Development. 1990. V. 110. P. 141–149.

130. Crozatier M., Kongsuwan K., Ferrer P., Merriam J.R., Lengyel J.A., Vincent A. // Genetics. 1992. V. 131. P. 905–916.

131. Lake C.M., Nielsen R.J., Hawley R.S. // PLoS Genet. 2011. V. 7. P. e1002005.

132. McKim K.S., Hayashi-Hagihara A. // Genes Dev. 1998. V. 12. P. 2932–2942.

133. Jang J.K., Sherizen D.E., Bhagat R., Manheim E.A., McKim K.S. // J. Cell. Sci. 2003. V. 116 (Pt 15). P. 3069–3077.

134. Liu H., Jang J. K., Kato N., McKim K.S. // Genetics. 2002. V. 162. P. 245–258.

Structure Modeling of Human Tyrosyl-DNA Phosphodiesterase 1 and Screening for Its Inhibitors

I.V. Gushchina[1≠], D.K. Nilov[2≠], A.L. Zakharenko[3], O.I. Lavrik[3,4], V.K. Švedas[1,2*]

[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Lenin Hills 1, bldg. 73, Moscow, 119991, Russia

[2]Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Lenin Hills 1, bldg. 40, Moscow, 119991, Russia

[3]Institute of Chemical Biology and Fundamental Medicine, Russian Academy of Sciences, Siberian Branch, Lavrentiev avenue 8, Novosibirsk, 630090, Russia

[4]Altai State University, Lenin avenue 61, Barnaul, 656049, Russia

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

≠Authors, who contributed equally to this work

ABSTRACT The DNA repair enzyme tyrosyl-DNA phosphodiesterase 1 (Tdp1) represents a potential molecular target for anticancer therapy. A human Tdp1 model has been constructed using the methods of quantum and molecular mechanics, taking into account the ionization states of the amino acid residues in the active site and their interactions with the substrate and competitive inhibitors. The oligonucleotide- and phosphotyrosine-binding cavities important for the inhibitor design have been identified in the enzyme's active site. The developed molecular model allowed us to uncover new Tdp1 inhibitors whose sulfo group is capable of occupying the position of the 3'-phosphate group of the substrate and forming hydrogen bonds with Lys265, Lys495, and other amino acid residues in the phosphotyrosine binding site.

KEYWORDS inhibitor, docking, molecular modeling, tyrosyl-DNA phosphodiesterase 1.

ABBREVIATIONS Tdp1 − tyrosyl-DNA phosphodiesterase 1, Top1 − topoisomerase I, QM − quantum mechanics, MM − molecular mechanics.

INTRODUCTION

During DNA replication or transcription, single-strand breaks are usually introduced by topoisomerase I (Top1) in order to remove local helical tensions [1, 2]. However, various DNA damages (strand breaks, nucleobase lesions), as well as Top1 inhibition, lead to the accumulation of covalent Top1-DNA complexes with a catalytic tyrosine that is linked to the 3'-terminal phosphate [3, 4]. To maintain the native DNA structure and enable the replication process to proceed, such complexes are hydrolyzed by tyrosyl-DNA phosphodiesterase 1 (Tdp1), an important DNA repair enzyme found in humans and other eukaryotic organisms [5−8].

The Tdp1 substrate is a Top1-DNA complex in which Top1 is preliminarily proteolyzed to a short peptide fragment [9]. Tdp1 exhibits broad substrate specificity, because Top1 creates nicks at various sites in the DNA backbone (although it shows preference for the thymidine 3'-phosphodiester bond) [10]. The Tdp1 active site is centrally located in a substrate-binding groove. The narrow part of the groove on one side of the active site is positively charged and involved in the binding of the DNA strand. The wider part of the groove on the other side binds a peptide fragment of the substrate. The position of the substrate's 3'-phosphate group in the Tdp1 active center is stabilized by hydrogen bonds with the Lys265 and Lys495 residues. It is considered that carboxamide groups of Asn283 and Asn516 are also involved in the phosphate binding [11, 12]. The phosphodiester bond between the 3'-phosphate and tyrosine residue is cleaved via an S_N2 mechanism, with the participation of the His263 and His493 side chains, and a transition state is formed in a trigonal bipyramidal configuration when the $N^{\varepsilon2}$ atom of His263 and tyrosyl oxygen occupy apical positions at the nucleophilic attack by His263, whereas the His493 residue donates a proton to the tyrosine residue in the leaving group (*Fig. 1*) [13, 14]. The protonated state of the $N^{\delta1}$ atoms of His263 and His493 is stabilized by hydrogen bonds with the Glu538 and Gln294 side chains, respectively. The deprotonation of the $N^{\varepsilon2}$ atom of His263 may be forced by the close proximity of the charged amino groups of Lys265 and Lys495; and the charged state of His493, by the proximity of the Asp288 side chain.

Camptothecin and its derivatives (irinotecan, topotecan) cause the formation of irreversible covalent Top1-DNA complexes and are, therefore, used to inflict DNA damage on cancer cells [3]. The suppression of the elimination of such complexes by Tdp1 inhibitors is a promising way with which to enhance the antitumor effect of camptothecins, which is confirmed by the fact that *TDP1*-deficient cells are sensitive to chemotherapy [15−17]. While there are several compounds known to suppress enzyme activity, drug development based on Tdp1 inhibitors remains far from a preclinical or clinical stage. For instance, the vanadate ion VO_4^{3-}, forming a coordinate bond with His263 and resembling the transition state of the reaction, was used to study the catalytic mechanism and to obtain crystal Tdp1 complexes with various oligonucleotides and peptide fragments [10, 13]. Tdp1 inhibitors were detected by *in vitro* screening of low-molecular-weight compounds, including steroid derivatives [18], indenoisoquinolines [19, 20], phosphotyrosine mimetics [21], thioxothiazolidinones [22], benzopentathiepines [23], and diazaadamantanes [24]. The above-mentioned compounds presumably compete for the substrate binding site, though the structures of the enzyme-inhibitor complexes are unknown, and the specific interactions between these molecules and active site residues are still to be uncovered. A molecular docking investigation of the interactions between several inhibitors and Tdp1 led to contradictory results which poorly correlated with experimental data on the inhibitory effect of the compounds [25, 26]. This suggests that protein models built on the basis of crystal structures need to be elaborated and optimized. In some studies, the reaction mechanism and molecular environment were not taken into account when estimating the ionization states of the histidine [22] and lysine [18, 27] side chains in the active site: that questions the reliability of the modeling. Obviously, a high-quality model of human Tdp1 which takes into account the structural features of the active site is needed to simulate the binding of potential inhibitors. The goal of the present study was to build a molecular model of Tdp1 using hybrid methods of quantum and molecular mechanics, as well as to verify its validity for virtual screening for competitive inhibitors.

EXPERIMENTAL SECTION

Protein structure modeling
The molecular model of human Tdp1 was built on the basis of the 1nop crystal structure (chains A, C, D) [14]. The coordinates of the missing loops in the protein structure were predicted with the Swiss-PDBViewer 4.1 program (which implements structure superimposition) [28] and ModLoop web server (predicts the position of the missing heavy atoms) [29]. The coordinates of the loop 425-434 were transferred from the 1qzq structure following its superimposition onto 1nop, and the coordinates of the loop 560-567 missing in all Tdp1 crystal structures were predicted from the amino acid sequence.

Next, the enzyme-substrate complex of Tdp1 was modeled using the AmberTools 1.2 (http://ambermd.org) and Amber 12 [30, 31] packages installed on the MSU supercomputer [32]. The substrate molecule was constructed based on a structural analogue from 1nop (covalent complex vanadate-oligonucleotide-peptide), by replacing the vanadium atom with phosphorus. Parameters from the AMBER parameter database [33] were used to provide a molecular mechanical description for the phosphotyrosine moiety of the substrate molecule. The remaining portion of the substrate and the protein were described by the *ff99SB* force field [34]. Hydrogen atoms were added to the structure of the enzyme-substrate complex, and, then, it was placed in a water box (TIP3P solvent model, minimum distance of 12 Å between the protein and the box's edge). Chlorine ions were added into the box to neutralize the positive net charge caused by the ionogenic groups of the protein and the substrate. The energy minimization of the obtained system was performed in two stages. At the first stage (2,500 steps of the steepest descent algorithm followed by 2,500 conjugate gradient steps), the protein and substrate coordinates were kept fixed by positional constraints of 2 kcal/(mol·Å²) on heavy atoms. At the second stage (5,000 steepest descent steps followed by 5,000 conjugate gradient steps), the system was partitioned into quantum mechanics (QM) and molecular mechanics (MM) regions. The QM region consisting of a fragment of the substrate and the side chains of His263 and His493 (see *Fig. 1*) was described using the semi-empirical Hamiltonian RM1 [35, 36] and linker atoms at the region boundaries. A PME (Particle Mesh Ewald) approach and periodic boundary conditions were chosen in computing long-range electrostatic interactions.

A search for binding pockets in the obtained Tdp1 structure was performed using the fpocket 2.0 and pocketZebra software [37, 38], with cavities identified as clusters of alpha spheres (spheres that are in contact with four atoms and do not contain internal atoms). To identify small cavities, the minimum number of alpha spheres in a cavity was reduced from 35 to 30, and the maximum distance between alpha spheres at a clustering step was also reduced from 2.5 to 2.4 Å. Hydrogen atoms were not taken into account during the search for cavities.
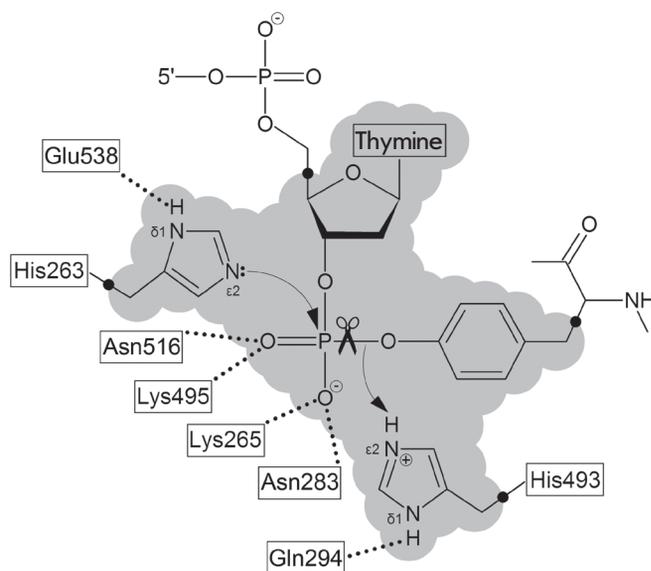
Fig. 1. The structure of the Tdp1 active site. The Lys265, Asn283, Lys495, and Asn516 residues are involved in the binding of the substrate's phosphate group. In the reaction mechanism, nucleophilic attack by His263 residue occurs and a proton is transferred from His493 to the leaving group. The shaded area corresponds to the QM region defined in the performed modeling of the enzyme-substrate complex.

Virtual screening

Virtual screening for Tdp1 inhibitors was carried out among low-molecular-weight compounds from the Vitas-M commercial library (http://www.vitasmlab.com). The protonation and structure optimization of compounds was performed as described previously [39]. Compounds containing a sulfo group and in conformity with the rule of three (molecular weight < 300, log $P \leq 3$, hydrogen bond donors ≤ 3, hydrogen bond acceptors ≤ 3, rotatable bonds ≤ 3) [40, 41] were selected from the library using the ACD/SpectrusDB 14.0 program (http://www.acdlabs.com). The substrate and water molecules were removed from the obtained model of Tdp1 enzyme-substrate complex, and an energy grid box (map of interaction potential) overlapping the active site was generated through the Lead Finder 1.1.15 program [42, 43]. Next, the molecular docking of the compounds into the Tdp1 active site was performed using a genetic algorithm in "extra precision" mode. The resulting structures of the complexes with inhibitors were optimized according to the procedure applied to the Tdp1 enzyme-substrate complex. The QM-region included an inhibitor molecule and the side chains of His263 and His493, and the molecular-mechanical parameters of the inhibitors were taken from the *GAFF*

force field [44]. The visualization of predicted poses was performed using the VMD 1.9.2 software [45].

Enzyme activity assay

The recombinant human Tdp1 protein was expressed in *Escherichia coli* and extracted according to the earlier described procedure [46]. The plasmid pET 16B Tdp1 was kindly provided by Dr. K.W. Caldecott (University of Sussex, United Kingdom). The enzyme was purified by chromatography with nickel sorbent NTA-Ni^{2+}-Sepharose CL-6B, and, then, the final purification was done with phosphocellulose P-11. A previously constructed biosensor 5'-(5,6 FAM-aac gtc agg gtc ttc c-BHQ1)-3', where FAM is a fluorophore and BHQ1 is a fluorescence quencher, was used for measurements of enzyme activity [23, 47]. The Tdp1 activity was monitored by detecting the release of 3'-terminal substituent BHQ1 under the following conditions: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 7 mM β-mercaptoethanol, 50 nM biosensor, 1.3 nM Tdp1, 26°C. The reaction rate at different concentrations of compounds STK370528 (Sigma-Aldrich) and STK376552 (Vitas-M Laboratory, Ltd) was measured using a POLARstar OPTIMA fluorimeter (BMG LABTECH, Germany). The measurements were conducted in two independent experiments. The IC$_{50}$ values (the inhibitor concentration required to reduce the enzyme activity by 50% [48]) were determined using the MARS Data Analysis 2.0 program (BMG LABTECH).

RESULTS AND DISCUSSION

Protein model

To construct a molecular model of human Tdp1 that could be used to screen for its competitive inhibitors, it was necessary to select an appropriate crystal structure of the enzyme, take into account the ionization of catalytically important amino acid residues, and reproduce the conformations of these residues that allow for an optimal interaction with the substrate. The Protein Data Bank contains structures of the Tdp1 apo form (PDB ID 1jy1, 1qzq), as well as complexes with various transition state analogues (1mu7, 1mu9, 1nop, 1rff, 1rfi, 1rg1, 1rg2, 1rh0, 1rgt, 1rgu). A complex with the closest substrate analogue − 1nop − in which vanadate is covalently bound to the catalytic His263 residue was selected as the initial structure for modeling. Through the replacement of the vanadium atom with phosphorus, the starting substrate structure was obtained: the oligonucleotide 5'-GTT-3' linked to the peptide KLNYL via a tyrosine side chain.

An important modeling step was the reconstruction of the missing loops 425-434 and 560-567, as a protein structure without chain breaks was required for fur-

ther optimization. Hydrogen atoms were added to the Tdp1 structure with reconstructed loops; a hydrogen atom was attached to the $N^{\delta 1}$ atom of the His263 side chain, and the His493, Lys265, and Lys495 side chains were taken to be charged. The optimization of the coordinates of the substrate and those of the added hydrogen atoms was done in two stages. At the first stage, a molecular-mechanical minimization was performed to remove the largest strains in the system. At the second stage, the semi-empirical Hamiltonian RM1, whose efficiency was demonstrated in simulations of biomolecules [49, 50], was used to describe the interactions between the substrate and the catalytic residues His263 and His493 more precisely. The most important interatomic distances in the active site of the starting and optimized Tdp1 models are listed in *Table 1*. The initial position of phosphate atoms in the starting model corresponds to the coordinates of vanadate in a complex with the enzyme, resembling the transition state. Through the structure optimization, the phosphate adopts a tetrahedral configuration and the distance between phosphorus and His263 increases from 2.0 to 2.7 Å, which corresponds to their disposition in the ground state of the active site. Hydrogen bonding of the phosphate group with other residues does not undergo significant changes. This demonstrates that in both

Table 1. Interactions of the 3′-terminal phosphate group of the substrate with the active site residues in the starting and optimized models of human Tdp1.

Interaction	Distance (Å)	
	Starting model	Optimized model
PO_4^-:P $\cdots$ His263:NE2	2.0	2.7
PO_4^-:$O_{bridging}$ $\cdots$ His493:NE2	2.6	2.6
PO_4^-:O $\cdots$ Lys265:NZ	2.8	2.7
PO_4^-:O $\cdots$ Lys495:NZ	2.8	2.7
PO_4^-:O $\cdots$ Asn283:ND2	3.0	2.8
PO_4^-:O $\cdots$ Asn516:ND2	3.2	3.0

the ground state and transition state the Asn283 and Asn516 side chains participate in the substrate binding and, together with the charged amino groups of Lys265 and Lys495, form a hydrogen-bonding network with the 3'-phosphate group. A Tdp1 model for the docking of small molecules was obtained by removing the substrate from the optimized structure, where the orienta-
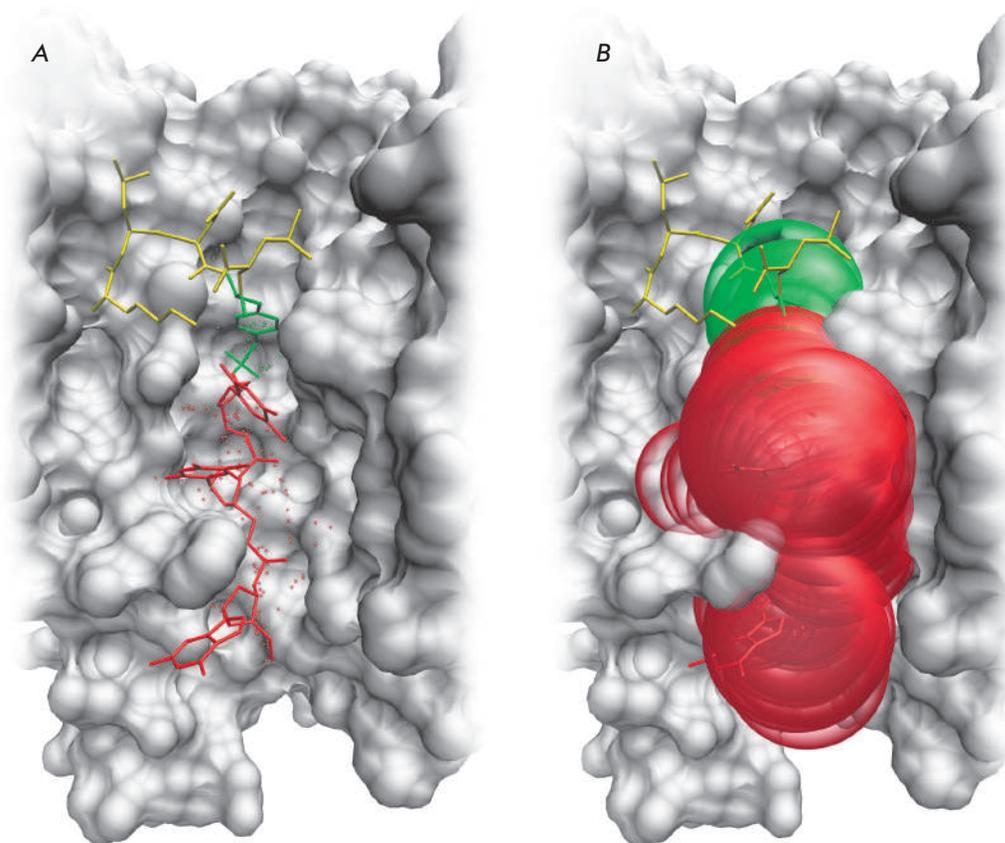
Fig. 2. The substrate-binding groove in the human Tdp1 model. (*A*) The interaction of the substrate molecule with the oligonucleotide and phosphotyrosine binding sites. The oligonucleotide is shown in red, phosphotyrosine is shown in green, and the rest of the peptide is shown in yellow. Cavities are labeled with points corresponding to the centers of alpha spheres. (*B*) The localization of alpha spheres in the oligonucleotide and phosphotyrosine binding sites.

tions of the active site residues could provide multiple interactions with competitive inhibitors.

An analysis of the substrate-binding groove surface in the Tdp1 model allowed us to identify binding sites for potential inhibitors. There are two distinct binding cavities; one for the phosphotyrosine and a second for the oligonucleotide, with the Asn516 and His263 side chains located at the boundary between them (*Fig. 2*). The oligonucleotide-binding cavity is a large region which has a total surface area of 666 Å². Among the amino acid residues positioned in this region are the residue pairs Ser400-Ser518 and Ser403-Ala520 involved in the binding of the second and third phosphate groups from the 3'-terminus. The phosphotyrosine-binding cavity is substantially smaller (206 Å²), but all the key active site residues participate in its formation: His263, His493, Lys265, Lys495, Asn283, Asn516, as well as the Tyr204, Pro461 and Trp590 residues involved in hydrophobic contacts.

Most of the known Tdp1 inhibitors are deprived of negatively charged moieties. Therefore, it is quite possible that the phosphotyrosine cavity, adapted to accommodate the 3'-terminal phosphate, does not participate in the binding of these compounds. This assumption is confirmed by a modeling of inhibitor binding using molecular docking. So, diazaadamantane derivatives, whose inhibitory properties were recently reported [24], are localized in the oligonucleotide region of the active site upon simulation of their binding using the Tdp1 model. The tricyclic moiety of these inhibitors occupies the site of the third ribose residue from the

3'-terminus, while an extended hydrophobic substituent is oriented towards the phosphotyrosine binding site, but does not interact with it (*Fig. 3A*).

Inhibitor screening

The presence of a cluster of the conserved Lys265, Lys495, Asn283, and Asn516 residues in the phosphotyrosine binding site makes possible an effective electrostatic interaction between the enzyme and substrate and may be an important structural factor in the binding of competitive inhibitors containing an appropriate charged group. A sulfo group, SO_3^-, might serve as a functional group of that type, being a structural analogue of phosphate. To verify this assumption, sulfonic acids and their salts (71 compounds) were retrieved from a library of low-molecular-weight compounds conforming to the rule of three that specifies the ranges of physicochemical parameters of the molecular fragments (small molecules used in primary screening and subsequent structure optimization). The compounds were docked into the Tdp1 model active site and examined for their ability to form hydrogen bonds with Lys265, Lys495, Asn283, Asn516, as well as other interactions with the DNA and peptide binding sites.

As a result of the screening, we selected the most promising inhibitors, STK370528 and STK376552, in which the sulfo group was attached to a heterocyclic moiety via a thioether linker (*Table 2, Fig. 3B*). The conformations of amino acid residues that interact with STK370528 and STK376552 in the obtained enzyme-

Table 2. Compounds selected by virtual screening as human Tdp1 inhibitors.

	Chemical structure	ΔG^{calc} (kcal/mol)	ΔG^{recalc} (kcal/mol)	IC_{50} (μM)
STK370528		−7.5	−8.7	83±24
STK376552		−8.4	−8.0	686±14
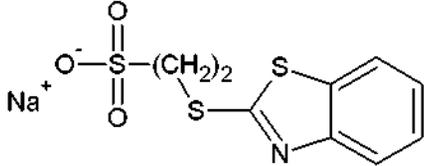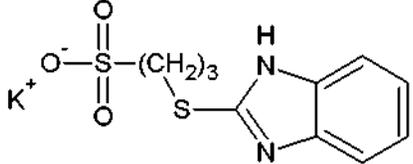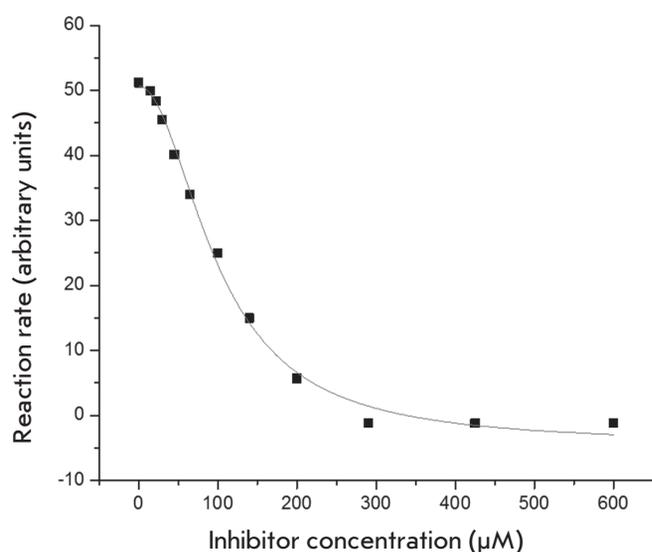
Fig. 4. The dependence of the Tdp1-catalyzed reaction rate on the concentration of the inhibitor STK370528.

Fig. 5. The position of the inhibitor, STK370528, in the active site of the molecular model of human Tdp1. Dotted lines indicate hydrogen bonds important for the stabilization of the sulfo group position. The gray color denotes the substrate coordinates in the model of the enzyme-substrate complex.

inhibitor complexes were subsequently optimized using the RM1 Hamiltonian. Re-docking into the refined protein models revealed that STK370528 was a more effective inhibitor and had higher binding energy ΔG^{recalc} (see data in *Table 2*).

For the experimental testing of the inhibitory properties of the compounds against the recombinant form of human Tdp1, we applied a biosensor (an oligonucleotide containing a fluorophore at the 5'-end and a fluorescence quencher at the 3'-end) that enables measurement of enzyme activity in real time. The method is based on the ability of Tdp1 to remove various large adducts from the 3'-end of DNA [17], including the fluorescence quencher BHQ1 (Black Hole Quencher 1) [51]. Upon BHQ1 removal by the enzyme, the intensity of the 5'-terminal fluorophore emission depends

on the amount of cleaved substrate. *Figure 4* shows a typical plot of the reaction rate as a function of the STK370528 concentration. The IC_{50} values were 83 μM for STK370528 and 686 μM for STK376552. Thus, the experimental study confirmed the conclusions of molecular modeling and showed that the selected com-

pounds were Tdp1 inhibitors that suppress enzyme activity in the micromolar concentration range.

The sulfo group of the inhibitors is capable of occupying the position of the 3'-phosphate group of the substrate and can form hydrogen bonds with the amino acid residues Lys265, Lys495, Asn283, Asn516, and His493, which constitute the phosphotyrosine binding site (*Fig. 5*). The location of the heterocyclic moiety in the oligonucleotide binding site leads to additional interactions. In the case of STK370528, a benzothiazole group forms a hydrogen bond with Asn516 and hydrophobic contacts with the Ala520 and Ala521 side chains. A flexible linker in the inhibitor structure provides a connection between groups located in different regions of the Tdp1 active site. The linker in STK376552 is elongated by one methylene unit, which disrupts interactions with Ala521 and Asn516 and decreases the inhibitory activity of this compound compared to STK370528.

Electrostatic interactions with the charged residues Lys265, Lys495, and His493 play an important role in the binding and orientation of inhibitors in the Tdp1 active site. In the case of uncharged sulfonates (phenyl and methyl esters of STK370528), the efficiency of interaction with the active site residues is reduced as confirmed by a number of different inhibitors' orientations upon simulation of their binding in the enzyme active site. Modeling of the binding of indenoizoquinoline

sulfonates which had been previously considered to be potential inhibitors but exhibited no activity against Tdp1 [25], has also shown that an esterified sulfo group cannot mediate interactions with the phosphotyrosine binding site.

CONCLUSIONS

The present study shows that the constructed molecular model of the DNA repair enzyme Tdp1, taking into account the structural features of the active site, adequately describes the binding of small molecules and makes possible a selection of substrate-competitive inhibitors through virtual screening. Based on a detailed analysis of intermolecular interactions, we selected from the computer library of potential inhibitors the sulfonates STK370528 and STK376552, which are capable of suppressing enzyme activity in the micromolar concentration range. The structural organization and localization of the oligonucleotide- and phosphotyrosine-binding sites in the substrate-binding groove were shown to be important factors to be considered when developing new Tdp1 inhibitors. ●

REFERENCES
1. Champoux J.J. //Annu. Rev. Biochem. 2001. V. 70. P. 369−413.
2. Wang J.C. // Nat. Rev. Mol. Cell Biol. 2002. V. 3. P. 430−440.
3. Pommier Y. // Nat. Rev. Cancer. 2006. V. 6. P. 789−802.
4. Lebedeva N., Rechkunova N., Boiteux S., Lavrik O. // IUBMB Life. 2008. V. 60. P. 130−134.
5. Pommier Y., Redon C., Rao V.A., Seiler J.A., Sordet O., Takemura H., Antony S., Meng L., Liao Z., Kohlhagen G., et al. // Mutat. Res. 2003. V. 532. P. 173−203.
6. Murai J., Huang S.Y., Das B.B., Dexheimer T.S., Takeda S., Pommier Y. // J. Biol. Chem. 2012. V. 287. P. 12848−12857.
7. Jakobsen A.K., Lauridsen K.L., Samuel E.B., Proszek J., Knudsen B.R., Hager H., Stougaard M. // Exp. Mol. Pathol. 2015.V. 99. P. 56−64
8. Interthal H., Pouliot J.J., Champoux J.J. // Proc. Natl. Acad. Sci. USA. 2001. V. 98, P. 12009−12014.
9. Debéthune L., Kohlhagen G., Grandas A., Pommier Y. // Nucleic Acids Res. 2002. V. 30, P. 1198−1204.
10. Davies D.R., Interthal H., Champoux J.J., Hol W.G. // J. Med. Chem. 2004. V. 47. P. 829−837.
11. Davies D.R., Interthal H., Champoux J.J., Hol W.G. // Structure. 2002. V. 10. P. 237−248.
12. Raymond A.C., Rideout M.C., Staker B., Hjerrild K., Burgin A.B., Jr. // J. Mol. Biol. 2004. V. 338. P. 895−906.
13. Davies D.R., Interthal H., Champoux J.J., Hol W.G.J. // J. Mol. Biol. 2002. V. 324. P. 917−932.
14. Davies D.R., Interthal H., Champoux J.J., Hol W.G.J. //

Chem. Biol. 2003. V. 10. P. 139−147.
15. Dexheimer T.S., Antony S., Marchand C., Pommier Y. // Anticancer Agents Med. Chem., 2008. V. 8. P. 381−389.
16. Beretta G.L., Cossa G., Gatti L., Zunino F., Perego, P. // Curr. Med. Chem. 2010. V. 17. P. 1500−1508.
17. Comeaux E.Q., van Waardenburg R.C. // Drug Metab. Rev. 2014. V. 46. P. 494−507.
18. Dexheimer T.S., Gediya L.K., Stephen A.G., Weidlich I., Antony S., Marchand C., Interthal H., Nicklaus M., Fisher R.J., Njar V.C., et al. // J. Med. Chem. 2009. V. 52. P. 7122−7131.
19. Conda-Sheridan M., Reddy P.V., Morrell A., Cobb B.T., Marchand C., Agama K., Chergui A., Renaud A., Stephen A.G., Bindu L.K., et al. // J. Med. Chem. 2013. V. 56. P. 182−200.
20. Nguyen T.X., Abdelmalak M., Marchand C., Agama K., Pommier Y., Cushman M. // J. Med. Chem. 2015. V. 58. P. 3188−3208.
21. Marchand C., Lea W.A., Jadhav A., Dexheimer T.S., Austin C.P., Inglese J., Pommier Y., Simeonov A. // Mol. Cancer Ther. 2009. V. 8. P. 240−248.
22. Sirivolu V.R., Vernekar S.K., Marchand C., Naumova A., Chergui A., Renaud A., Stephen A.G., Chen F., Sham Y.Y., Pommier Y., et al. // J. Med. Chem. 2012. V. 55. P. 8671−8684.
23. Zakharenko A., Khomenko T., Zhukova S., Koval O., Zakharova O., Anarbaev R., Lebedeva N., Korchagina D., Komarova N., Vasiliev V., et al. // Bioorg. Med. Chem. 2015. V. 23. P. 2044−2052.

24. Zakharenko A.L., Ponomarev K.U., Suslov E.V., Korchagina D.V., Volcho K.P., Vasil'eva I.A., Salakhutdinov N.F., Lavrik O.I. // Bioorg. Khim. 2015. V. 41. P. 657–662.

25. Nguyen T.X., Morrell A., Conda-Sheridan M., Marchand C., Agama K., Bermingham A., Stephen A.G., Chergui A., Naumova A., Fisher R., et al. // J. Med. Chem. 2012. V. 55. P. 4457–4478.

26. Dean R.A., Fam H.K., An J., Choi K., Shimizu Y., Jones S.J., Boerkoel C.F., Interthal H., Pfeifer T.A. // J. Biomol. Screen. 2014. V. 19. P. 1372–1382.

27. Weidlich I.E., Dexheimer T., Marchand C., Antony S., Pommier Y., Nicklaus M.C. // Bioorg. Med. Chem. 2010. V. 18. P. 182–189.

28. Guex N., Peitsch M.C. // Electrophoresis. 1997. V. 18. P. 2714–2723.

29. Fiser A., Sali A. // Bioinformatics. 2003. V. 19. P. 2500–2501.

30. Case D.A., Darden T.A., Cheatham T.E., 3rd, Simmerling C.L., Wang J., Duke R.E., Luo R., Walker R.C., Zhang W., Merz K.M., et al. // AMBER 12. University of California, San Francisco. 2012.

31. Case D.A., Cheatham T.E., 3rd, Darden T., Gohlke H., Luo R., Merz K.M., Jr., Onufriev A., Simmerling C., Wang B., Woods R.J. // J. Comput. Chem. 2005. V. 26., P. 1668–1688.

32. Voevodin Vl.V., Zhumatiy S.A., Sobolev S.I., Antonov A.S., Bryzgalov P.A., Nikitenko D.A., Stefanov K.S., Voevodin Vad.V. // Open Systems J. (Mosc.). 2012. V. 7. P. 36–39.

33. Homeyer N., Horn A.H., Lanig H., Sticht H. // J. Mol. Model. 2006. V. 12. P. 281–289.

34. Hornak V., Abel R., Okur A., Strockbine B., Roitberg A., Simmerling C. // Proteins. 2006. V. 65. P. 712–725.

35. Rocha G.B., Freire R.O., Simas A.M., Stewart J.J.P. // J. Comp. Chem. 2006. V. 27. P. 1101–1111.

36. Walker R.C., Crowley M.F., Case D.A. // J. Computat. Chem. 2008. V. 29. P. 1019–1031.

37. Le Guilloux V., Schmidtke P., Tuffery P. // BMC Bioinformatics. 2009. V. 10. P. 168.

38. Suplatov D., Kirilin E., Arbatsky M., Takhaveev V., Švedas V. // Nucleic Acids Res. 2014. V. 42. P. W344-W349.

39. Nilov D.K., Prokhorova E.A., Švedas V.K. // Acta Naturae. 2015. V. 7. №2. P. 57–63.

40. Congreve M., Carr R., Murray C., Jhoti H. // Drug Discov. Today. 2003. V. 8. P. 876–877.

41. Lipinski C.A. // Drug Discov. Today Technol. 2004. V. 1. P. 337–341.

42. Stroganov O.V., Novikov F.N., Stroylov V.S., Kulkov V., Chilov G.G. // J. Chem. Inf. Model. 2008. V. 48. P. 2371–2385.

43. Novikov F.N., Stroylov V.S., Stroganov O.V., Kulkov V., Chilov G.G. // J. Mol. Model. 2009. V. 15. P. 1337–1347.

44. Wang J., Wolf R.M., Caldwell J.W. Kollman P.A., Case D.A. // J. Comput. Chem. 2004.V. 25. P. 1157–1174.

45. Humphrey W., Dalke A., Schulten K. // J. Mol. Graph. 1996. V. 14. P. 33–38.

46. Lebedeva N.A., Rechkunova N.I., Lavrik O.I. // FEBS Lett. 2011. V. 585. P. 683–686.

47. Johansson M.K., Fidder H., Dick D., Cook R.M. // J. Am. Chem. Soc. 2002. V. 124, P. 6950–6956.

48. Neubig R.R., Spedding M., Kenakin T., Christopoulos A. Pharmacol. Rev. // 2003. V. 55, P. 597–606.

49. Seabra G. de M., Walker R.C., Roitberg A.E. // J. Phys. Chem. A. 2009. V. 113, P. 11938–11948.

50. Khaliullin I.G., Nilov D.K., Shapovalova I.V., Švedas V.K. // Acta Naturae. 2012. V. 4. №2. P. 80–86.

51. Jensen P.W., Falconi M., Kristoffersen E.L., Simonsen A.T., Cifuentes J.B., Marcussen L.B., Frohlich R., Vagner J., Harmsen C., Juul S., et al. // Biosens. Bioelectron. 2013. V. 48, P. 230–237.

Dipole Modifiers Regulate Lipid Lateral Heterogeneity in Model Membranes

S. S. Efimova, O. S. Ostroumova*

Institute of Cytology of the Russian Academy of Sciences, Tikhoretsky av. 4, St. Petersburg 194064, Russia

*E-mail: osostroumova@mail.ru

ABSTRACT In this study we report on experimental observations of giant unilamellar liposomes composed of ternary mixtures of cholesterol (Chol), phospholipids with relatively low T_{melt} (DOPC, POPC, or DPoPC) and high T_{melt} (sphingomyelin (SM), or tetramyristoyl cardiolipin (TMCL)) and their phase behaviors in the presence and absence of dipole modifiers. It was shown that the ratios of liposomes exhibiting noticeable phase separation decrease in the series POPC, DOPC, DPoPC regardless of any high-T_{melt} lipid. Substitution of SM for TMCL led to increased lipid phase segregation. Taking into account the fact that the first and second cases corresponded to a reduction in the thickness of the lipid domains enriched in low- and high-T_{melt} lipids, respectively, our findings indicate that the phase behavior depends on thickness mismatch between the ordered and disordered domains. The dipole modifiers, flavonoids and styrylpyridinium dyes, reduced the phase segregation of membranes composed of SM, Chol, and POPC (or DOPC). The other ternary lipid mixtures tested were not affected by the addition of dipole modifiers. It is suggested that dipole modifiers address the hydrophobic mismatch through fluidization of the ordered and disordered domains. The ability of a modifier to partition into the membrane and fluidize the domains was dictated by the hydrophobicity of modifier molecules, their geometric shape, and the packing density of domain-forming lipids. Phloretin, RH 421, and RH 237 proved the most potent among all the modifiers examined.

KEYWORDS chalcones, flavonoids, lateral heterogeneity, lipid bilayers, lipid domains, dipole modifiers, styryl-pyridinium dyes phase separation

ABBREVIATIONS Chol − cholesterol, DOPC − 1,2-dioleoyl-sn-glycero-3-phosphocholine, DPoPC − 1,2-dipalmito-leoyl-sn-glycero-3-phosphocholine, POPC − 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, SM − porcine brain sphingomyelin, TMCL − 1,1',2,2'-tetramyristoylcardiolipin.

INTRODUCTION

The lipid bilayer may have a domain structure determined by immiscible lipid phases coexisting in different aggregate states. Single-component lipid bilayers exist in the solid state at temperatures below the melting points (T_{melt}) of lipids. Depending on the tilt angle of lipid molecules and the packing of hydrocarbon tails, the solid bilayer is comprised of the following phases: the solid phase (crystalline), the gel phase, and the ripple phase, which is typical of saturated phosphocholines [1]. At a point above the transition temperature, the state of bilayer lipids changes into a liquid-like state. Lipid components with varying melting temperatures can show complicated phase behavior in different areas of the membrane in a temperature-dependent manner. This leads to the coexistence of solid (s_o) and liquid states (l_d) attributed to lipids with high and low melting temperatures, respectively. The presence of sterols, in

particular cholesterol, promotes phase segregation and induces the liquid-ordered state (l_o). There is evidence that the coefficient of lateral lipid diffusion in the l_o-phase is 2−3 times lower as compared to the l_d-areas [2]. The existence of lipid lateral segregation has been demonstrated in biological membranes. Although gel domains are not exclusive to model membranes (they are also present in biological membranes [3]), it has been generally assumed that phase segregation in biological membranes is mainly represented by two liquid phases (l_d + l_o) [4]. Since not only membrane lipids are sensitive to lateral segregation, but also peptides, a concept of lipid-protein nanodomains (rafts) has been proposed and received increasing attention. These rafts are enriched in high-T_{melt} lipids and cholesterol and exist in the l_o-phase. In recent years there has been growing interest in lipid rafts due to their important role in protein trafficking, signaling, immune response, etc. [5−

The main characteristics of the lipid molecules

Lipid	Chemical structure	C $n{:}m$	T_{melt}, ºC
DPoPC		16 : 1	–36
DOPC		18 : 1	–17
POPC		16 : 0−18 : 1	–2
SM		16 : 1−18 : 0	45
TMCL		14 : 0	47
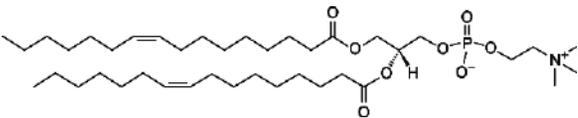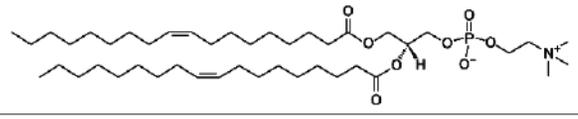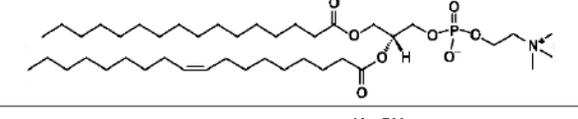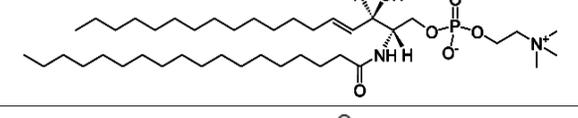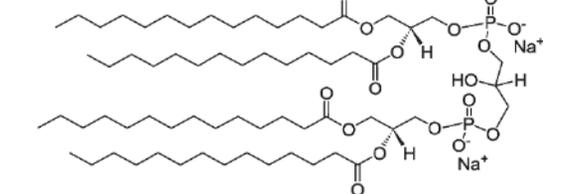
Note. Cn : m is the number of carbon atoms (n) and double bonds (m) in acyl chains; T_{melt} is the main phase-transition temperature.

16]. Importantly, the occurrence of lipid-protein rafts has not yet been agreed upon. These nanodomains are one to hundreds of nanometers in size and are extremely dynamic. In lipid bilayers, the ordered domains can be of large dimension, which allows for visualization by fluorescence microscopy using single unilamellar liposomes [17]. It is possible to observe phase segregation in liposomes loaded with fluorescently labeled lipids. Most dyes are targeted at the liquid-disordered raft fraction, leaving the ordered domains unlabeled.

Amphiphilic low-molecular-weight compounds, known as dipole modifiers, in particular some flavonoids, can influence the equilibrium between the phases. Ostroumova et al. [18] reported that flavonoid compounds such as biochanin A, phloretin, and myricetin are able to negatively affect phase separation scenarios in model membranes composed of binary lipid mixtures (DOPC : SM (80 : 20 mol.%), DOPC : DMPC (50 : 50 mol. %) or DOPC : DPPC (50 : 50 mol. %)). A similar effect was observed for phloretin, its glycoside phlorizin, quercetin, myricetin, and styrylpyridinium dyes of the RH series in a three-component bilayer mixture

of POPC, Chol, and SM [19]. Although Efmova et al. [19] examined the influence of the above-mentioned dipole modifiers on the domain structure of POPC membranes incorporating sterols and sphingolipids, the roles of these phospholipids, which constitute the disordered liquid phase, remain poorly understood. The objective of this work was to investigate the effect of low-T_{melt} lipid components on the phase separation scenario in liposomes packed with Chol and SM before and after the introduction of flavonoids or RH dyes. With a variety of phospholipids, POPC, DOPC and DPoPC, we were able to sequentially change the disordered lipid phase thickness of a fluid membrane. Lipid mixtures containing TMCL were also studied for their ability to modify the thickness of ordered lipid domains.

MATERIALS AND METHODS

Materials
The following compounds were used in the study: sorbitol, phloretin, phlorizin, quercetin, myricetin, and RH 421 (Sigma, USA); RH 237 (Molecular Probes, USA);
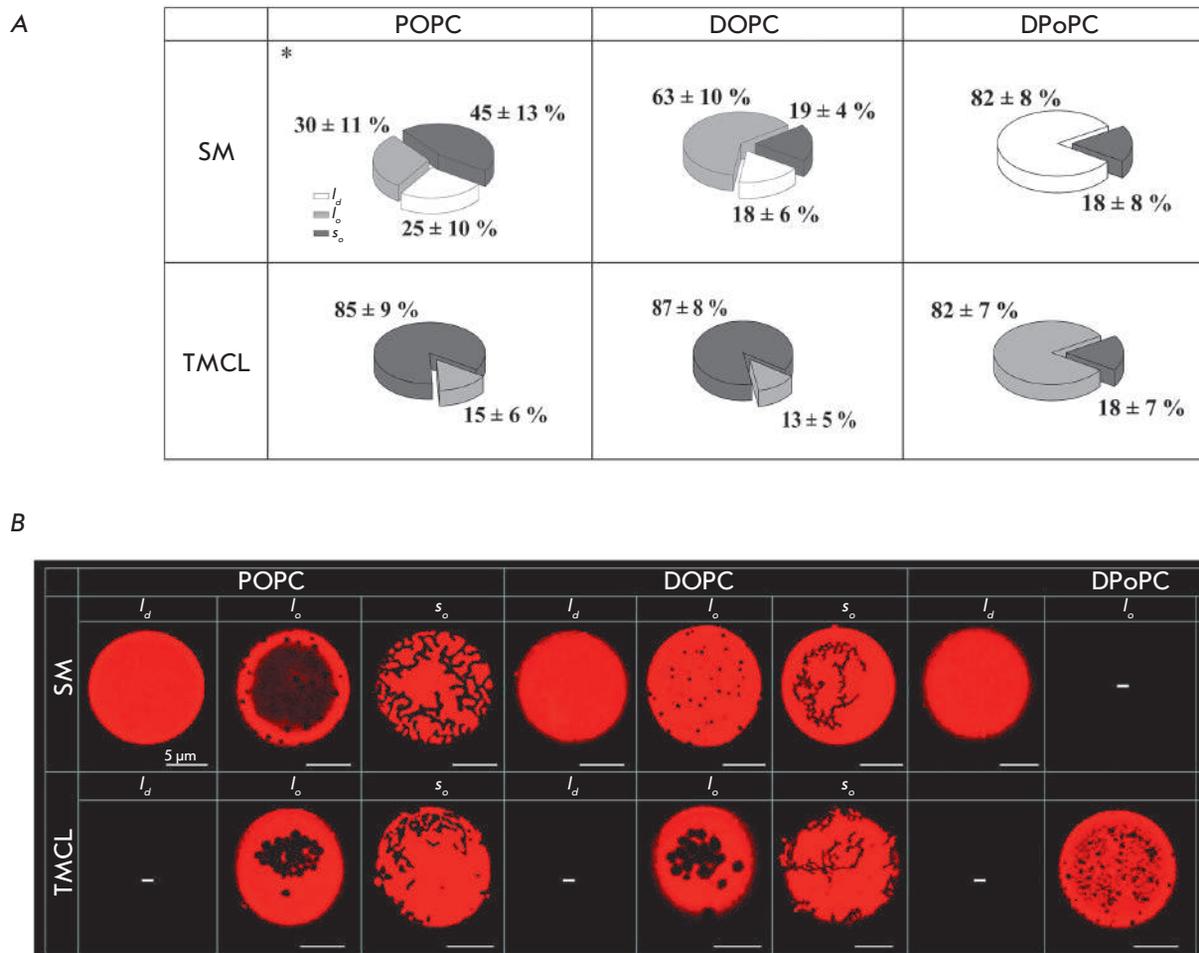
Fig. 1. Pie charts demonstrating the possible scenarios of phase separation in liposome membranes composed of sphingomyelin (SM) or tetramyristoyl cardiolipin (TMCL) (40 mol. %), cholesterol (Chol) (20 mol. %), and different phospholipids (POPC, DOPC, or DPoPC) (40 mol. %). **(A)** Microphotographs of liposomes with different lipid compositions and phase behaviors (l_d, l_o, s_o) (B). Here and in *Figs. 3* and *4*, dark gray sectors denote the percentage of vesicles with solid ordered domains (s_o); light gray sectors denote the percentage of vesicles with liquid-ordered domains (l_o); white sectors denote the relative number of homogeneously stained liposomes with liquid-disordered state without noticeable phase separation (l_d). * – data from ref. [19]. **(B)** Fluorescence micrographs of giant unilamellar vesicles demonstrating various types of membrane phase separation scenarios (l_d, l_o, s_o).

1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine 1 (POPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-dipalmitoleoyl-*sn*-glycero-3-phosphocholine (DPoPC), 1,1',2,2'-tetramyristoyl cardiolipin (TMCL), porcine brain sphingomyelin (SM), cholesterol (Chol) and lissamine rhodamine B-1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (Rh-DPPE) (Avanti Polar Lipids, USA). The table provides details for each of the lipids used.

Confocal microscopy of giant unilamellar liposomes

Giant unilamellar liposomes were prepared by the electroformation technique using the Vesicle Prep Pro machine (Nanion, Germany) on glass slides coated with titanium and indium oxides (90% indium oxide : 10% indium oxide, 29 × 68 × 0.9 mm) with a surface specific resistivity of 20−30 Ω/sq. (standard protocol, 3 V, 10 Hz, 1 h, 25°C.) Lateral phase segregation was visualized by adding the Rh-DPPE fluorescent probe into a three-component mixture that consisted of 40 mol. % low-T_{melt} phospholipid (DOPC, POPC or DPoPC), 40 mol. % high-T_{melt} phospholipid (SM or TMCL), and 20 mol. % Chol in chloroform (2 mM). The final Rh-DPPE concentration was 1 mol. %. The liposome suspension was aliquoted for storage. An aliquot without a dipole modifier was used as control. Test samples contained 400 µM flavo-

noid (phloretin, phlorizin, quercetin, or myricetin) or 10 μM styrylpyridinium dye (RH 421 or RH 237). Images were acquired with APO oil-immersion objective lens 100.0 × /1.4 HCX PL using a Leica TCS SP5 confocal laser scanning microscope (Leica Microsystems, Germany). Liposomes were examined at 25˚C. Rh-DPPE emission was excited at 543 nm (a helium-neon laser). There is evidence that in lipid bilayer systems with phase segregation, Rh-DPPE shows partitioning preference mainly for the disordered liquid phase (l_d) [20], whereas the liquid-ordered phase (l_o) and solid ordered phase (gel, s_o) remain unlabeled [21]. Ordered domains were identified morphologically: the dye-unlabeled circular domains were considered to be in the l_o-state, while the dye-unlabeled domains of irregular shape were assigned to the s_o-state. Each sample was characterized by the ratio (p_i, %) of homogeneous and heterogonous vesicles:

$$p_i = \frac{N_i}{N} \cdot 100\,\%, \qquad (1)$$

where i is liposome phase separation (homogeneous l_d-vesicles or liposomes that carry the l_o or s_o-domains); N_i is the vesicle number in a sample with a certain phase scenario (from 0 to 50); and N is the total liposome number in a sample (50 in each system). The p_i values were obtained by averaging values from four independent experiments. The data for each lipid system were presented in pie charts, along with standard deviations for liposomes with assigned phase behavior.

RESULTS AND DISCUSSION

Figure 1A (upper panel) shows findings on possible types of phase behavior in unilamellar membranes comprised of SM (40 mol. %), Chol (20 mol. %), and low-T_{melt} phospholipids (POPC, DOPC or DPoPC; 40 mol. %) (see Table for details on T_{melt}). Microphotographs with each type of phase segregation scenario (l_d, l_o, s_o) are presented in *Fig. 1B (upper panel)*. Phase behavior of ternary mixtures containing SM/Chol/ POPC had been previously examined [19]. We found that liposomes that incorporate 45 ± 13% SM/Chol/POPC contain solid domains of irregular shape (s_o), whereas 30 ± 11% SM/Chol/POPC vesicles are enriched in liquid-ordered domains with a circular morphology (l_o). The remaining liposomes are vesicles homogeneously labeled with the fluorescent probe (l_d), exhibiting no phase segregation. *Figure 1A (upper panel)* demonstrates that substitution of POPC for DOPC in the membrane mixture reduces the number of vesicles with s_o-domains (19 ± 4%) and increases the number of liposomes with the l_o-state (63 ± 10%). When DPoPC was used, 82 ± 8% vesicles were homogeneously dye-labeled without noticeable phase separation, while the remaining vesicles contained

solid domains. It is tempting to suggest that visual phase separation decreases in the series POPC, DOPC, DPoPC, and so does the thickness of the disordered phase (d_{Ld}), which includes different low-T_{melt} phospholipids, whereas the mismatch (Δd) in the hydrophobic bilayer thickness of the coexisting liquid-ordered and liquid-disordered phases increases (*Fig. 2, left part*) [22, 23]. As a result, the formation of well-defined boundaries between the ordered and disordered domains, which seemingly favors the exposure of a portion of the hydrophobic region to the aqueous environment, becomes energetically prohibitive, thus decreasing the number of liposomes with visible phase separation. A similar conclusion can be reached based on the results shown in *Fig. 1 (lower panel)*, which presents the data on the phase separation of TMCL membranes ((TMCL; 40 mol. %), Chol (20 mol. %) and other low-T_{melt} phospholipids (40 mol. %)). One can notice that TMCL-containing liposomes show phase separation regardless of any low-T_{melt} phospholipids (no homogeneously labeled liposome). The differences between the lipid systems are due to the proportion of vesicles carrying l_o- and s_o-domains. As noted above, DPoPC contributes to the lowest thickness of the l_d-phase among all the phospholipids tested, which corresponds to the highest Δd value, and consequently to the highest energy of ordered domain formation. This explains why DPoPC-containing liposomes showed poor phase separation (82 ± 7% liposome have l_o-domains) versus the POPC- and DOPC-vesicles that form l_d-phases with greater thickness and lower Δd values with phase separation in most liposomes (85 ± 9% and 87 ± 8% in the l_d/s_o ratio, respectively).

An analysis of phase behavior scenarios involving various high-T_{melt} phospholipids also suggests a role for Δd in regulating the lateral heterogeneity of ternary membrane mixtures. *Figure 1B (lower panel)* depicts microphotographs of TMCL-containing liposomes with low-T_{melt} phospholipids. *Figure 1A* demonstrates that SM to TMCL substitution in the membrane mixture leads to enhanced phase separation. In the case of POPC- (*left column*) and DOPC-containing bilayers (*middle column*), the proportion of liposomes with s_o- domains increases, whereas DPoPC-bilayers display a statistically significant increase in the numbers of vesicles with l_o-domains (*right column*). This is attributed to the fact that the presence of TMCL in place of SM lowers the thickness of the ordered phase and decreases Δd (*Fig. 2, right part*). Taken together, this substitution finally reduces the energy of ordered domain boundary formation. Overall, the findings in *Fig. 1* allow one to link the lateral heterogeneity of ternary membranes to the mismatch in the membrane thickness of the liquid-ordered and liquid-disordered
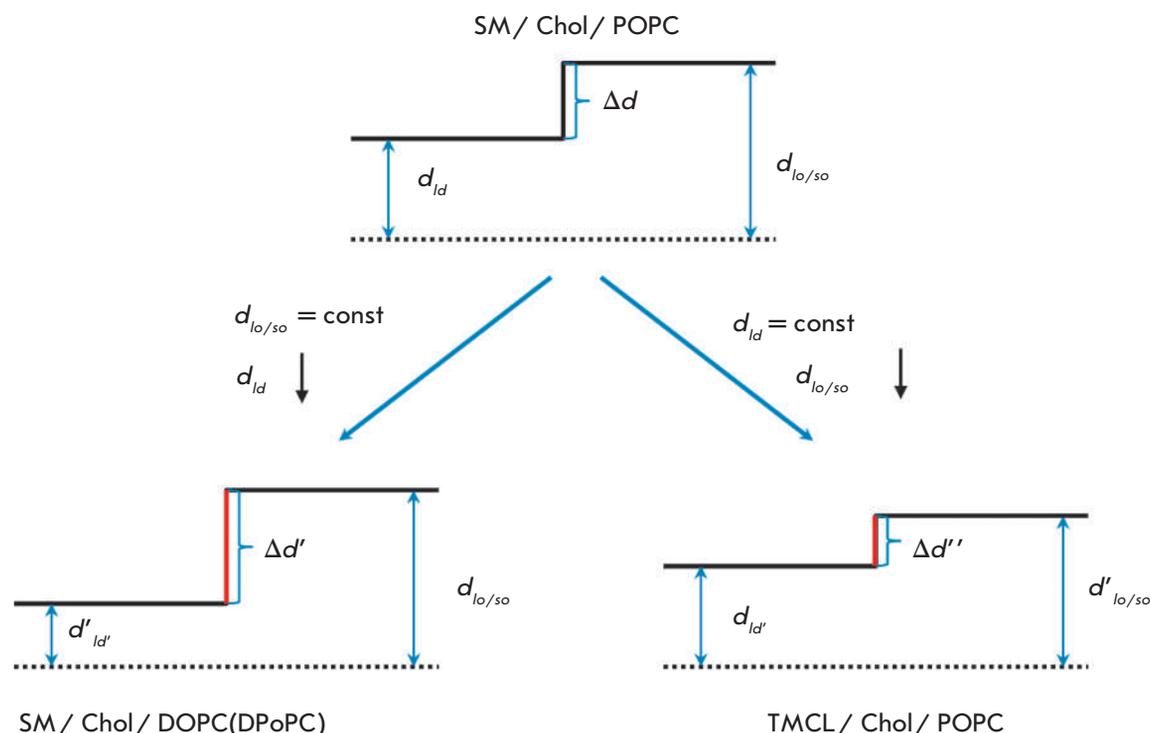
Fig. 2. Schematic representation of the correlation between the thickness mismatch (Δd) of the ordered ($d_{lo/so}$) and disordered domains (d_{ld}) and bilayer lipid composition. The dotted line marks the center of the bilayer; the solid line indicates the boundary between the polar and nonpolar regions of the membranes. For color designations, see *Fig. 1*.

phases: the degree of phase separation is inversely proportional to Δd values.

Taking into account the fact that dipole modifiers impact not only the dipole potential, but also the packing of lipid components [24−27], we suggest that these agents possess the ability to alter the phase separation scenario. Recently, we have investigated the effects of the dipole modifiers phloretin, phlorizin, quercetin, myricetin, RH 421, and RH 237 on the phase separation behavior in SM/Chol/POPC-vesicles [19]. The data are shown in *Fig. 3A (upper panel)*. It is clear that the dipole modifiers decrease membrane phase separation, which manifests itself as reduced liposome numbers with gel domains. However, upon incorporation of phloretin, quercetin, or myricetin, the decline in the number of vesicles with s_o-domains is accompanied by a corresponding 40−45% increase in the number of homogeneously stained liposomes. The presence of phlorizin, RH 421, and RH 237 induced a 30−35% increase in the ratio of vesicles with l_o-domains and a 5−10% increase in the number of homogeneous liposomes. The elevated liposome concentrations versus homogeneously labeled DOPC-liposomes in the presence of phloretin, phlorizin, RH 42,1 and RH 237 (by 10−30%) and elimination of vesicles with s_o-domains

in the presence of phloretin could be explained by decreased phase separation following the addition of dipole modifiers as in the case with POPC (*Fig. 3A, middle panel*). *Figure 3B* shows microphotographs of lipid vesicles containing DOPC, Chol, and SM and their phase behaviors (l_d, l_o, s_o) in the presence of phloretin and RH 421. No statistical significance was found regarding the effects of quercetin and myricetin on phase separation in DOPC membranes.

Changes in the phase separation scenario of SM-containing membranes in the presence of dipole modifiers could be caused by elevated Δd values under the influence of the agents tested. The most likely scenario is that the polar heads of lipids take over more space in the membrane in response to burying of the modifiers into the lipid layer and dipole-dipole interactions between them. As shown by differential scanning calorimetry, this relatively increases the mobility of carbohydrate chains and reduces the T_{melt} of lipids [18, 25, 28]. The more "fluid-like" state of the membrane correlates with the decreased bilayer thickness. In this case, the extent of the effect of a modifier will depend on the backbone and overall hydrophobicity, which govern the degree to which the modifier is buried into the bilayer. That is why the hydrophobic phloretin exerts

A

B

Fig. 3. **(A)** Pie charts demonstrating the possible scenarios of phase separation in liposome membranes composed of sphingomyelin (SM) (40 mol. %), cholesterol (Chol) (20 mol. %), and different phospholipids (POPC, DOPC or DPoPC) (40 mol. %) in the presence of dipole modifiers (400 µM phloretin, 400 µM phlorizin, 400 µM quercetin, 400 µM myricetin, 10 µM RH 421, and 10 µM RH 237). For color designations, see *Fig.*1. * — data from ref. [19] **(B)** Fluorescence microphotographs of SM/Chol/DOPC-liposomes demonstrating various types of membrane phase separation scenarios (l_d, l_o, s_o) in the presence of phloretin and RH 421.

the strongest effect on membrane lateral heterogeneity, whereas its hydrophilic analog, phlorizin, and the highly hydroxylated flavonoids quercetin and myricetin exhibit weaker effects. The length of the styrylpyridinium dyes RH 421 and RH 237 is sufficient to transverse the lipid monolayer, but the increase in the space occupied by a single lipid in the membrane is largely due to electric repulsion among the sulfonate groups located in the polar bilayer region [27].

In addition to the modifier type, the geometric characteristics of lipid molecules that form the phase into which a modifier partitions also play a regulating role. In the case of lipids with a cylindrical geometry, such as DOPC, POPC, and SM [29−31], l_d-domains become sensitive to fluidization as compared to ordered domains, since partitioning of modifiers into the l_d-domains seems to be impeded in the context of tightly packed lipids. This scenario is schematically illustrated in *Fig.* 2 (*left part*).

As shown by the lower panel in *Fig. 3A*, DPoPC-containing membranes exhibited no statistically significant differences between phase behavior scenarios

before and after the modifiers had been added. Bearing in mind that no phase separation is observed in most DPoPC-vesicles even in the absence of dipole modifiers due to the greatest mismatch in the membrane thickness of the liquid-ordered and liquid-disordered phases, further elevation of Δd does not lead to significant changes in bilayer phase separation.

In contrast to SM, TMCL has an inverted cone shape that triggers inverted spontaneous curvatures of the monolayers formed by it [32]. It is highly likely that this favors partitioning of dipole modifiers having a cone shape into the ordered TMCL-enriched phase. Simultaneous fluidization of disordered l_d-domains and ordered domains will not dramatically alter the thickness mismatch between the phases, thus preventing changes in phase behavior scenarios. *Figure 4A* shows that regardless of the type of low-T_{melt} lipid within the model membranes, the presence of a dipole modifier neither significantly increases the relative number of TMCL-containing liposomes with l_o-domains nor induces the emergence of liposomes with noticeable phase separation. *Figure 4B* shows mi-

A

vesicle lipid composition (40/20/40 mol. %)	dipole modifier		
	Phloretin	RH 421	RH 237
TMCL/Chol/POPC	I_o s_o $80 \pm 11\,\%$ $20 \pm 9\,\%$	$86 \pm 12\,\%$ $14 \pm 9\,\%$	$87 \pm 7\,\%$ $13 \pm 5\,\%$
TMCL/Chol/DOPC	$82 \pm 10\,\%$ $18 \pm 8\,\%$	$90 \pm 12\,\%$ $10 \pm 9\,\%$	$91 \pm 9\,\%$ $9 \pm 5\,\%$
TMCL/Chol/DPoPC	$92 \pm 5\,\%$ $8 \pm 4\,\%$	$95 \pm 7\,\%$ $5 \pm 4\,\%$	$95 \pm 5\,\%$ $5 \pm 3\,\%$
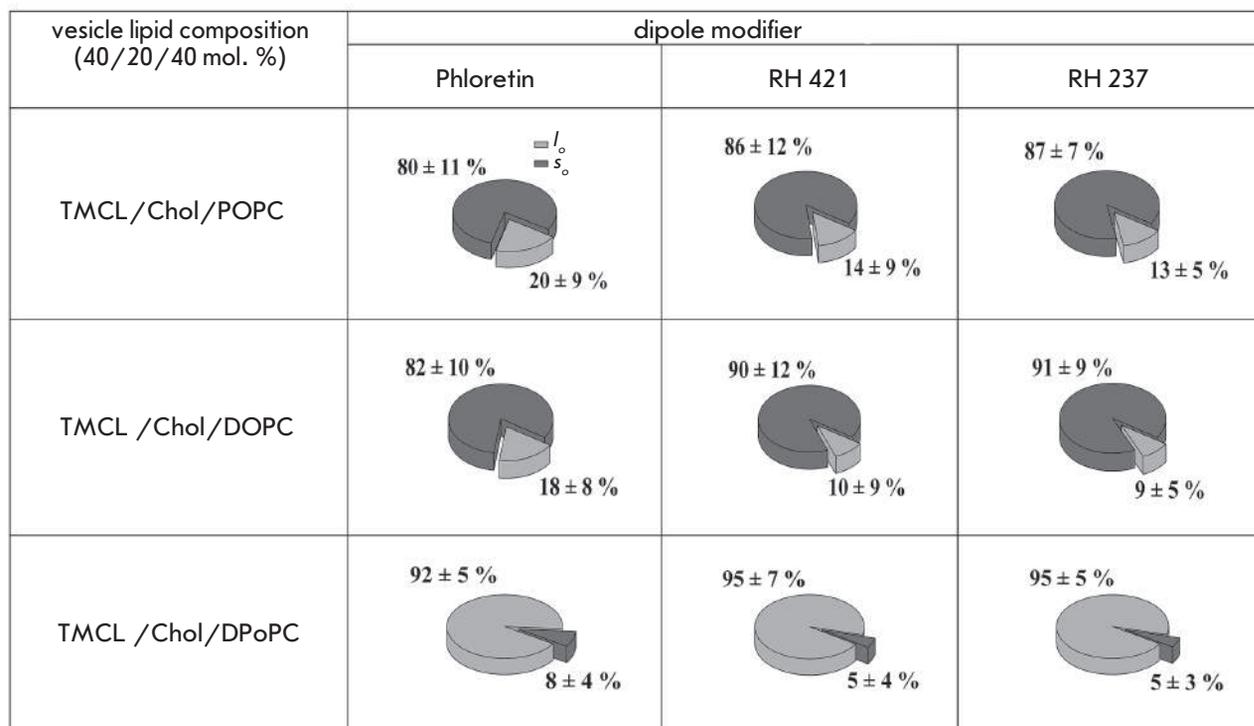
B
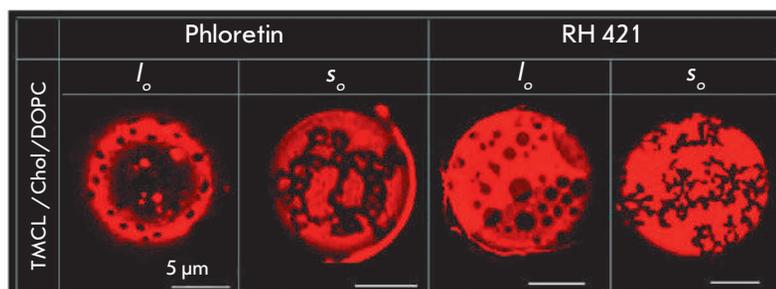
Fig. 4. **(A)** Pie charts demonstrating the possible scenarios of phase separation in liposomes membranes composed of tetramyristoyl cardiolipin (TMCL) (40 mol. %), cholesterol (Chol) (20 mol. %), and various phospholipids (POPC, DOPC, or DPoPC) (40 mol. %) in the presence of dipole modifiers (400 µM phloretin, 10 µM RH 421, and 10 µM RH 237). For color designations, see *Fig*.1. **(B)** Fluorescence micrographs of TMCL/Chol/DOPC-vesicles demonstrating l_d/l_o- or l_d/s_o-phase separation in the presence of phloretin and RH 421.

crophotographs of lipid vesicles incorporating DOPC, Chol, and TMCL and liposomes modified with phloretin or RH 421.

In conclusion, our findings suggest a key role for the mismatch thickness between the ordered and disordered phases in modulating phase behavior scenarios in ternary model membranes. It is believed that our work will open up new a venues for research into the use of dipole modifiers for the regulation of lipid lateral heterogeneity in bilayers. ●

REFERENCES

1. Cevc G. // Chem. Phys. Lipids. 1991. V. 57. P. 293−307.
2. Almeida P.F., Vaz W.L., Thompson T.E. // Biophys. J. 1993. V. 64. P. 399−412.
3. Aresta-Branco F., Cordeiro A.M., Marinho H.S., Cyrne L., Antunes F., de Almeida R.F. // J. Biol. Chem. 2011. V. 286. P. 5043−5054.
4. Brown D. // Int. J. Med. Microbiol. 2002. V. 291. P. 433−437.
5. Simons K., Ikonen E. // Nature. 1997. V. 387. P. 569−572.
6. Simons K., Toomre D. // Nat. Rev. Mol. Cell Biol. 2000. V. 1. P. 31−39.
7. Tsui-Pierchala B.A., Encinas M., Milbrandt J., Johnson E.M. // Trends Neurosci. 2002. V. 25. P. 412−417.
8. Pierce S.K. // Nat. Rev. Immunol. 2002. V. 2. P. 96−105.
9. Van Laethem F., Leo O. // Curr. Mol. Med. 2002. V. 2. P. 557−570.
10. Morgan M.J., Kim Y.S., Liu Z. // Antioxid. Redox. Signal. 2007. V. 9. P. 1471−1483.
11. Jury E.C., Flores-Borja F., Kabouridis P.S. // Semin. Cell Dev. Biol. 2007. V. 18. P. 608−615.
12. Yoshizaki F., Nakayama H., Iwahara C., Takamori K., Ogawa H., Iwabuchi K. // Biochim. Biophys. Acta. 2008. V. 1780. P. 383−392.
13. Fulop T., Le Page A., Garneau H., Azimi N., Baehl S., Dupuis G., Pawelec G., Larbi A. // Longev. Healthspan. 2012. V. 1. P. 6.
14. Head B.P., Patel H.H., Insel P.A. // Biochim. Biophys. Acta. 2014. V. 1838. P. 532−545.
15. Ratajczak M.Z., Adamiak M. // Leukemia. 2015. V. 29. P. 1452−1457.
16. Farnoud A.M., Toledo A.M., Konopka J.B., Del Poeta M., London E. // Curr. Top. Membr. 2015. V. 75. P. 233−268.
17. Wesołowska O., Michalak K., Jadwiga Maniewska J., Hendrich A.B. // Acta Biochim. Pol. 2009. V. 56. P. 33−39.
18. Ostroumova O.S., Chulkov E.G., Stepanenko O.V., Schagina L.V. // Chem. Phys. Lipids. 2014. V. 178. P. 77−83.
19. Efimova S.S., Malev V.V., Ostroumova O.S. // J. Membr. Biol. 2016. V. 279. P. 97−106.
20. Juhasz J., Davis J.H., Sharom F.J. // Biochem. J. 2010. V. 430. P. 415−423.
21. Muddana H.S., Chiang H.H., Butler P.J. // Biophys. J. 2012. V. 102. P. 489−497.
22. García-Sáez A.J., Chiantia S., Schwille P. // J. Biol. Chem. 2007. V. 282. P. 33537−33544.
23. Heberle F.A., Petruzielo R.S., Pan J., Drazba P., Kucerka N., Standaert R.F., Feigenson G.W., Katsaras J. // J. Am. Chem. Soc. 2013. V. 135. P. 6853−6859.
24. Cseh R., Hetzer M., Wolf K., Kraus J., Bringmann G., Benz R. // Eur. Biophys. J. 2000. V. 29. P. 172−183.
25. Tarahovsky Y.S., Muzafarov E.N., Kim Y.A. // Mol. Cell Biochem. 2008. V. 314. P. 65−71.
26. Ollila F., Halling K., Vuorela P., Vuorela H., Slotte J.P. // Arch. Biochem. Biophys. 2002. V. 399. P. 103−108.
27. Apetrei A., Mereuta L., Luchian T. // Biochim. Biophys. Acta. 2009. V. 1790. P. 809−816.
28. Cseh R., Benz R. // Biophys. J. 1999. V. 77. P. 1477−1488.
29. Sakuma Y., Taniguchi T., Imai M. // Biophys. J. 2010. V. 99. P. 472−479.
30. Bezrukov S.M. // Curr. Opin. Colloid Interface Sci. 2000. V. 5. P. 237−243.
31. Byström T., Lindblom G. // Spectrochim. Acta A Mol. Biomol. Spectrosc. 2003. V. 59. P. 2191−2195.
32. Powell G.L., Hui S.W. // Biophys. J. 1996. V. 70. P. 1402−1406.

Role of the Inserted α-Helical Domain in *E. coli* ATP-Dependent Lon Protease Function

A. M. Kudzhaev, A. G. Andrianova, E. S. Dubovtseva, O. V. Serova, T. V. Rotanova*

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Str., 16/10, Moscow, 117997, Russia

*E-mail: tatyana.rotanova@ibch.ru

ABSTRACT Multidomain ATP-dependent Lon protease of *E. coli* (**Ec-Lon**) is one of the key enzymes of the quality control system of the cellular proteome. A recombinant form of Ec-Lon with deletion of the inserted characteristic α-helical HI(CC) domain (Lon-dHI(CC)) has been prepared and investigated to understand the role of this domain. A comparative study of the ATPase, proteolytic, and peptidase activities of the intact Lon protease and Lon-dHI(CC) has been carried out. The ability of the enzymes to undergo autolysis and their ability to bind DNA have been studied as well. It has been shown that the HI(CC) domain of Ec-Lon protease is required for the formation of a functionally active enzyme structure and for the implementation of protein-protein interactions.
KEYWORDS AAA$^+$ proteins, ATP-dependent proteolysis, DNA binding, inserted α-helical domain, LonA proteases.
ABBREVIATIONS AMPPNP − adenosine 5'-(β,γ-imido)triphosphate, DTDP − 4,4'-dithiodipyridine, Glt − glutaryl, Nu − nucleotide, PepTBE − Suc-Phe-Leu-Phe-SBzl, Suc − succinyl.

INTRODUCTION

ATP-dependent Lon protease of *Escherichia coli* (Ec-Lon [EC 3.4.21.53], MEROPS: clan SJ, family S16, ID S16.001) is a member of the Lon protease family which plays a key role in the quality control system of the cellular proteome that functions in all domains of life [1−4]. The Lon family consists of two subfamilies: LonA, which includes bacterial and eukaryotic enzymes, and LonB, which combines the archaea enzymes. Proteases of the subfamilies A and B differ in the domain organization of their subunits, as well as in the environment of the catalytic residues of the proteolytic center [5]. Ec-Lon belongs to subfamily A and degrades abnormal and defective polypeptides, as well as a number of regulatory cellular proteins by a processive mechanism under conditions of a coupling of proteolysis to ATP hydrolysis [4−7]. The distinctive characteristic of Ec-Lon, as well as that of other LonA proteases, is their ability to bind DNA [8−10].

The Ec-Lon subunit (784 amino acid residues) consists of five domains: N−HI(CC)−NB−H−P (*Fig. 1A*), where the nucleotide-binding (NB) and α-helical (H) domains form a ATPase module that belongs to the superfamily of AAA$^+$ proteins (**A**TPases **a**ssociated with various cellular **a**ctivities) [11, 12]; the C-terminal P domain is serine-lysine peptide hydrolase; and the N-terminal and subsequent "inserted" α-helical domains form a non-catalytic region (N–HI(CC)) which includes a sequence fragment with a specific coiled-coil (CC) conformation [13, 14]. The crystal structures of the individual domains (except for the HI(CC) domain) of Ec-Lon and some other LonA proteases have been determined. The spatial structure of the full-length enzymes of the LonA subfamily remains unknown.

The two-domain organization of the N-terminal region is a unique characteristic of Ec-Lon and the entire pool of LonA proteases. LonA proteases differ from other AAA$^+$ proteins of the protein quality control system, such as the set of ATP-dependent proteases (ClpAP, ClpXP, FtsH, HslUV) and chaperone-disaggregases (ClpB, Hsp104), by the presence of the inserted HI(CC) domain. We have shown that the HI(CC) domain of Ec-Lon exhibits a marked similarity to both the H domain of its own AAA$^+$ module and to the α-helical domain (H1(M)) of the first of the two AAA$^+$ modules of ClpB chaperones [13, 14]. At the same time, the role of the HI(CC) domain in the functioning of Ec-Lon protease, its interaction with nucleic acids and/or the structural organization of the enzyme, has not been characterized to date.

In order to study the role of the inserted HI(CC) domain in the manifestation of Ec-Lon functional properties, we performed a comparative study of the enzymatic characteristics and ability to bind DNA of the intact enzyme (*Fig. 1A*) and its deletion form Lon-dHI(CC) without its HI(CC) domain (*Fig. 1B*).

MATERIALS AND METHODS

Materials
Commercial reagents from Sigma, Bio-Rad, Thermo Scientific (USA), Fluka (Switzerland), Boehringer Mannheim (Germany), Pharmacia (Sweden), Difco (England), Panreac (Spain) and Reakhim (Russia) were used in the study.

Preparation of Ec-Lon (Lon-H_6) and its deletion form Lon-dHI(CC)
A recombinant form of Ec-Lon containing a hexahistidine fragment (in LEHHHHHH octapeptide) at the C-terminus of the protein (Lon-H_6) was prepared according to the previously described procedure [15].

Deletion form Lon-dHI(CC) was obtained on the base of Lon-H_6 protease. Lon_d_124-304, Lon_HindIII and Lon_BamHI_rev primers (5'-TTTTTTGACCTTGCT-GCGCGCATCAATGGTCGGCGACTCCAG-3', 5'-CG-CAGAAAGAAGCTTCAACGG-3' and 5'-GTTCT-GCTCTGGATCCAGCAC-3', respectively) were constructed using the megaprimer method. Amplification of the gene fragment was carried out in two steps using plasmid DNA pET28-lon-H_6 as the template. In the first step, a PCR fragment was obtained using the Lon_d_124-304 and Lon_HindIII primers, and the fragment was subsequently used as the primer in the second step, together with a Lon_BamHI_rev primer. The resulting DNA fragment was about 625 bp in length and was cloned into the pET28_lon vector at the unique HindIII and BamHI restriction sites.

Sequencing of the cloned DNA and synthesis of the primers were carried out by ZAO EVROGEN (www.evrogen.ru). The restriction and ligation procedures were carried out according to the protocols of the manufacturers of the corresponding enzymes.

Isolation and purification of Lon-H_6 and Lon-dHI(CC) were performed in two steps by Ni^{2+} chelate affinity chromatography using HisTrap FF columns (tandem 2×5 mL, GE Healthcare, USA) and anion exchange chromatography on a HiTrap™ Q FF column (5 mL, GE Healthcare) according to the previously described procedure [15].

The protein concentrations were determined by the Bradford method [16].

The homogeneity of the proteins in the preparations was tested electrophoretically [17] using a commercial set of markers (M, kDa): β-galactosidase (116.0), bovine serum albumin (66.2), ovalbumin (45.0), lactate dehydrogenase (35.0), Bsp98I restriction enzyme (25.0), β-lactalbumin (18.4), and lysozyme (14.4).

DNA PURIFICATION
The DNA was purified according to the protocol presented in the manual [18].

Determination of the enzymatic properties of Lon-H_6 protease and its deletion form Lon-dHI(CC)

ATPase activity was tested by the accumulation of inorganic phosphate over time in the ATP hydrolysis reaction in 50 mM Tris-HCl buffer, pH 8.1, containing 150 mM NaCl, 5 mM ATP, 20 mM $MgCl_2$ and 1 µM enzyme at 37°C [19]. In the control experiment, the enzyme was replaced with a buffer. The initial reaction rates were determined from the optical absorption of a mixture of 200 µL of the reaction medium and 600 µL of the reagent (100 mM $Zn(AcO)_2$, 15 mM $(NH_4)_6Mo_7O_{24}$, 1% SDS, pH 4.5−5.0) at a wavelength of 350 nm ($\varepsilon_{350} = 7,800$ M^{-1} cm^{-1}).

The thioesterase activity. The hydrolysis of thiobenzyl ester of N-substituted tripeptide Suc-Phe-Leu-Phe-SBzl (PepTBE) was monitored spectrophotometrically at a wavelength of 324 nm from the optical absorption of 4-thiopyridone ($\varepsilon_{324} = 16,500$ M^{-1} cm^{-1}), which is the product of the reaction between the hydrolysis product (benzylthiolate, $BzlS^-$) and 4,4'-dithiodipyridine (DTDP) [20]. PepTBE hydrolysis was carried out at 37°C in 50 mM Tris-HCl buffer, pH 8.1, containing 150 mM NaCl, 10% DMSO, 0.2 mM DTDP, 0.1 mM PepTBE, and 0.2 µM enzyme. When studying the influence of effectors, a nucleotide up to 2.5 mM and $MgCl_2$ up to 20 mM were added to the mixture.

The proteolytic activity of the enzymes was tested electrophoretically [17]. The reaction was carried out at a temperature of 37°C in 50 mM Tris-HCl buffer, pH 8.1, containing 150 mM NaCl, 20 µM β-casein and 2−6 µM enzyme, in the absence or presence of 5 mM Nu and 20 mM $MgCl_2$. An aliquot of the reaction or control mixture (20 µL) was mixed with 7 µL lysis buffer (0.2 M Tris-HCl, pH 8.9, 4% SDS, 20% glycerol, 0.5 mM EDTA, 0.8% bromophenol blue, 3% mercaptoethanol), refluxed for 10 min, and was applied to a 12% polyacrylamide gel (PAGE) for electrophoresis.

The autolytic activity of the enzymes was tested electrophoretically [17] under conditions analogous to the conditions for determining the proteolytic activity, but in the absence of β-casein.

Testing of the Lon-H_6 protease and Lon-dHI(CC) protease complexes with plasmid DNA
The formation of enzyme-DNA complexes was monitored by a deceleration of DNA in an agarose gel (GMSA method) [21]. 20−25 µg of Lon-H_6 or Lon-dHI(CC) were incubated for 30 min at 25°C with 500 ng of plasmid DNA (pET28a) in 25 µL of 20 mM Tris-HCl buffer, pH 7.5, containing 60 mM NaCl. The protein−
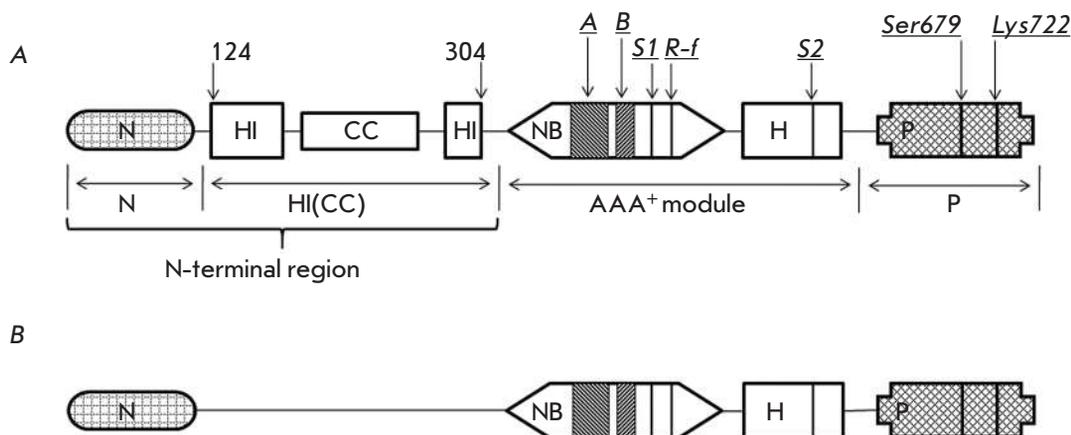
Fig. 1. Domain organization of the *E. coli* LonA protease (*A*) and its deletion form Lon-dHI(CC) (*B*). Domain designations: N – N-terminal; HI(CC) – inserted α-helical with a coiled-coil (CC) region; NB – nucleotide-binding; H – α-helical; P – proteolytic. ATPase center components: *A* and *B* – Walker motifs, *S1* and *S2* – sensor residues, *R-f* –"arginine finger" residue; proteolytic center components: *Ser679* and *Lys722* – catalytic residues.

DNA complexes were analyzed by gel electrophoresis in a standard 1.0% agarose gel. DNA bands were visualized by staining with ethidium bromide.

RESULTS AND DISCUSSION

The recombinant Ec-Lon protease used in the study, which contained an additional C-terminal octapeptide bearing a hexahistidine fragment (Lon-H_6), had been produced and characterized previously [15]. The recombinant deletion form Lon-dHI(CC), without the inserted HI(CC) domain (residues Glu124−Asn304, *Fig. 1B*), was obtained on the base of Lon-H_6. Preparative amounts of Lon-H_6 (M 88.5 kDa) and its deletion form Lon-dHI(CC) (M 67.5 kDa) were isolated using affinity chromatography on Ni-Sepharose and anion exchange chromatography on Q Sepharose. A comparative study of the enzymatic activity of intact Lon-H_6 protease and its deletion form was carried out. Three types of activity were characterized: ATPase, proteolytic (substrate: β-casein), and peptidase (substrate: Suc-Phe-Leu-Phe-SBzl, PepTBE), and the possibility of autolysis of the enzyme preparations was studied. In addition, the presence of nucleic acid in various protein preparations was tested by the phenol extraction method.

ATPase activity of the deletion form of Ec-Lon protease

The following standard conditions were selected for testing ATPase activity, as well as other types of activities of Lon-H_6 protease and its deletion form: 37 °C and 50 mM Tris-HCl buffer, pH 8.1, containing 150 mM NaCl.

It is known that native wt-Ec-Lon exhibits a maximum level of ATPase activity at equal concentrations of ATP and Mg^{2+}, and that excess of magnesium ions has an inhibitory effect on the hydrolysis of ATP, which is leveled by binding of the protein substrate [22].

The same trends are typical for intact Lon-H_6 protease (*Fig. 2A*): the efficiency of hydrolysis of ATP by the enzyme under conditions close to physiological ones (concentration ratio Nu:Mg^{2+} = 1:4) is significantly lower than at equimolar concentrations of Nu and Mg^{2+}. Addition of a protein substrate (β-casein) in both cases results in a significant increase in ATPase activity.

The Lon protease almost completely loses its ability to hydrolyze ATP with a loss of the HI(CC) domain: ATPase activity of Lon-dHI(CC) is reduced by more than 10 times compared to the activity of intact Lon-H_6 protease and by all means does not depend on either the ratio of nucleotide and Mg^{2+} ions concentrations or the addition of a substrate protein (*Fig. 2B*).

The obtained results indicate that the inserted α-helical HI(CC) domain is necessary for the formation of the ATPase center of the Ec-Lon protease and its correct functioning.

Activity of the peptidase center of the deletion form of Ec-Lon protease

Similarly to the Lon-H_6 protease, Lon-dHI(CC) is capable of hydrolyzing a model peptide substrate, PepTBE, but the basic peptidase activity of the deletion form is about 30% that of the activity of the intact enzyme (*Table*). The data in the *Table* demonstrate that only Mg^{2+} ions activate the peptidase centers of both Lon-H_6 and Lon-dHI(CC). The influence of nucleotide effectors on the intact and modified enzymes is radically different. Free nucleotides (except ADP) and Nu-Mg complexes

A Lon-H$_6$

B Lon-dHI(CC)
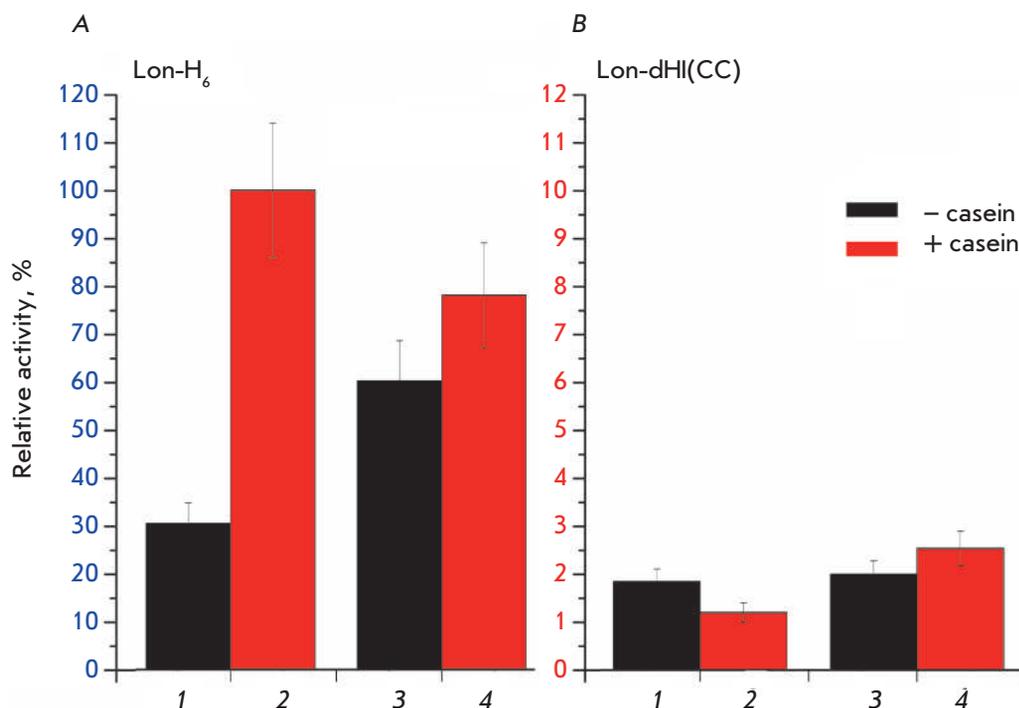
■ — casein
■ — + casein

Fig. 2. ATPase activity of the intact Lon-H$_6$ protease (*A*) and its deletion form Lon-dHI(CC) (*B*) in the absence (black columns) or presence (red columns) of the protein substrate, β-casein. Experimental conditions: 50 mM Tris-HCl buffer, pH 8.1; 0.15 M NaCl; 37°C; concentrations: 5 mM ATP; 20 (*1, 2*) or 5 mM (*3, 4*) MgCl$_2$; 0 (*1, 3*) or 0.5 mg/ml (*2, 4*) β-casein; 0.5-1.0 μM enzyme.

activate the Lon-H$_6$ protease to varying degrees (2−11 times) and ADP inhibits it, but none of the nucleotides has any effect on the hydrolysis of the peptide by Lon-dHI(CC). The Nu-Mg complexes exert a similar but relatively low activating effect on the enzyme peptidase center, comparable to the effect of Mg^{2+} ions. These data show that Lon-dHI(CC) is incapable of binding free nucleotides and weakly interacts with their complexes with magnesium ions. The most powerful effectors affecting the activity of the peptidase center are magnesium ions.

Thus, removal of the HI(CC) domain results in a decrease in the activity of the peptidase center of the Ec-Lon protease and a loss of the regulatory effect of the ATPase center on the peptidase one, which is defined by the nature of the bound nucleotide in the intact enzyme.

Proteolytic and autolytic activity of the deletion form of Ec-Lon protease

The proteolytic activity of Lon-H$_6$ and its deletion form Lon-dHI(CC) was tested using the hydrolysis of a model protein substrate, β-casein, in the absence and presence of Mg^{2+} ions, free nucleotides, and their complexes. The efficiency of hydrolysis of the target protein and the accumulation of degradation products were detected by gel electrophoresis.

The intact Lon-H$_6$ protease is capable of hydrolyzing β-casein in two cases: by the processive mechanism (without the formation of large intermediate frag-

Table. Influence of the effectors on the activity of Lon-H$_6$ and Lon-dHI(CC) peptidase centers

Effector	Lon-H$_6$		Lon-dHI(CC)	
	v	n	v	n
No effector	5.88	1	1.64	1
Mg	**33.1**	**5.62**	**5.19**	**3.16**
ATP	**47.1**	**8.01**	*1.33*	*0.81*
ADP	*0.49*	*0.08*	**1.79**	**1.09**
AMPPNP*	**14.2**	**2.41**	**1.82**	**1.11**
ATP-Mg	**63.5**	**10.8**	**4.62**	**2.82**
ADP-Mg	**10.1**	**1.73**	**4.89**	**2.98**
AMPPNP-Mg	**58.0**	**9.86**	**5.6**	**3.41**

Note. The specific rates of PepTBE hydrolysis (*v*, ([S], μM)/([E], μM) min) are given; *n* is the ratio of substrate hydrolysis rates in the presence and absence of the effector (*v*$_{ef}$/*v*$_0$), where *n* < 1 corresponds to inhibition (*italicized*), and *n* > 1 corresponds to activation of hydrolysis (shown **in bold**). The error did not exceed 10%. Experimental conditions: 50 mM Tris-HCl buffer, pH 8.1; 0.15 M NaCl; 10% DMSO; 0.1 mM PepTBE; 0.2 mM DTDP; 2.5 mM Nu; 20 mM MgCl$_2$; 0.2 μM enzyme; 37°C.
* Nonhydrolysable ATP analog, adenosine-5'-(β,γ-imido)-triphosphate.
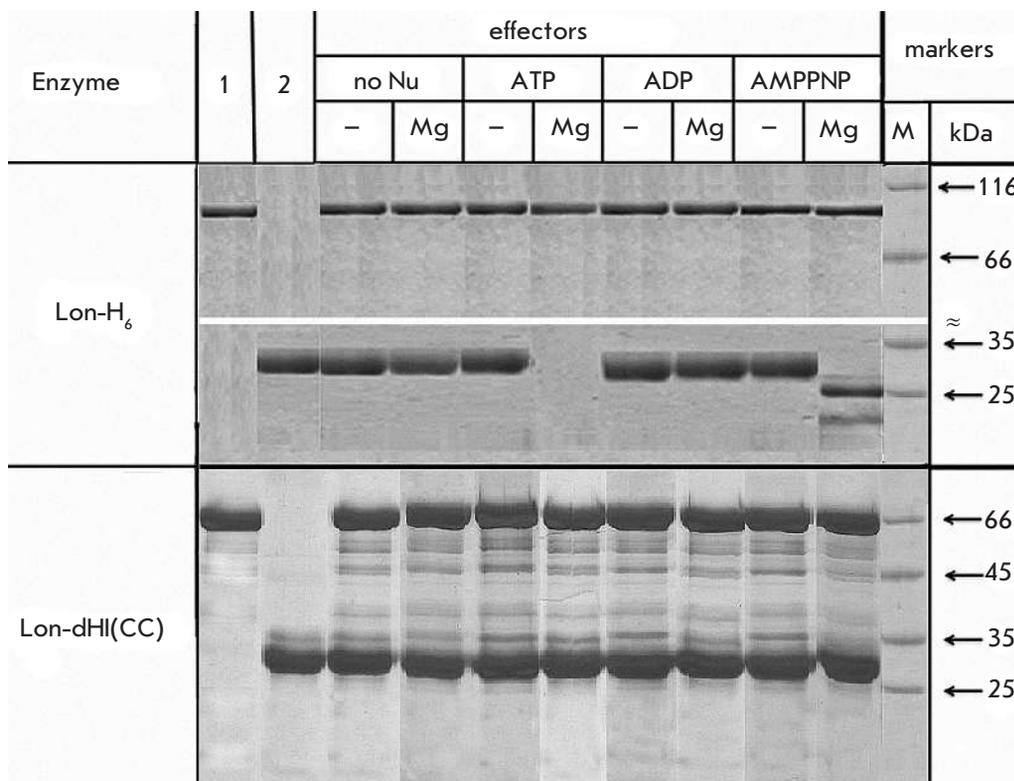
Fig. 3. Hydrolysis of β-casein by Lon-H$_6$ protease and its deletion form Lon-dHI(CC) with and without effectors (electrophoresis in 12% PAGE). Experimental conditions: 50 mM Tris-HCl buffer, pH 8.1; 0.15 M NaCl; 37°C; reaction time 2 h. Concentrations: Nu – 5 mM; MgCl$_2$ – 20 mM; β-casein – 0.5 mg/ml; Lon-H$_6$ – 2.5 μM; Lon-dHI(CC) – 6 μM. 1 – enzyme (control), 2 – β-casein (control), "–" – in the absence of Mg^{2+}, Mg – in the presence of Mg^{2+}, M – markers.
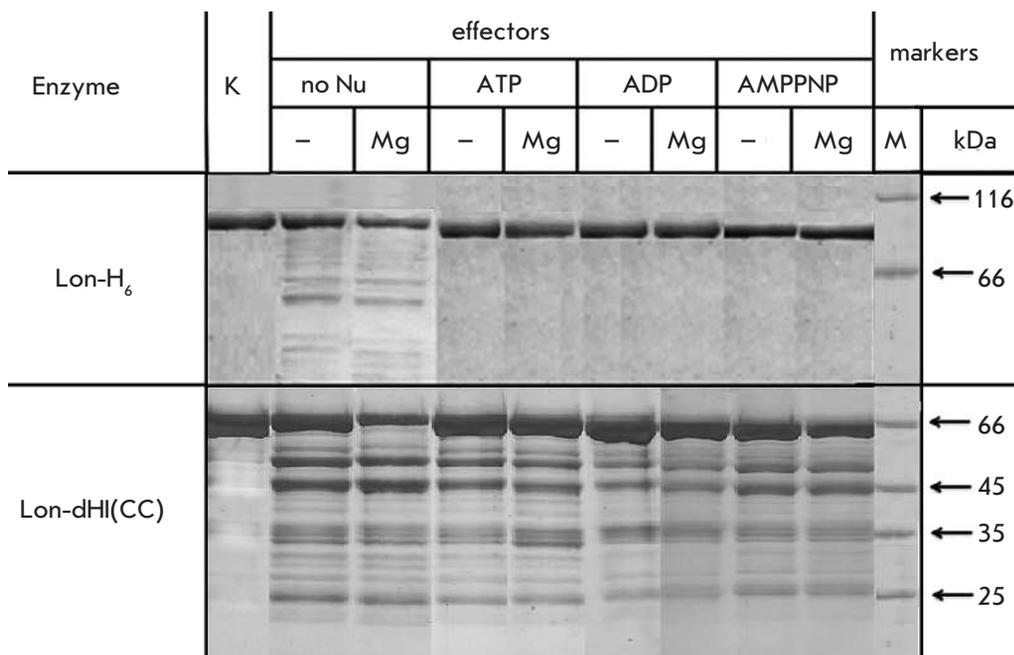
Fig. 4. Autolysis of Lon-H$_6$ protease and its deletion form Lon-dHI(CC) with and without effectors. The experimental conditions and designations follow Fig. 3 with the following modifications: Lon-H$_6$ – 3.4 μM, reaction time 24 h. K – the original enzyme (control, reaction time 0 h).

ments) under conditions of a coupling of proteolysis to ATP hydrolysis or by a nonprocessive mechanism in the presence of a complex of a nonhydrolyzable analogue of ATP with magnesium (*Fig. 3*).

Deletion of the HI(CC) domain leads to a complete loss of the proteolytic activity towards β-casein by the deletion form, which indicates the importance of this domain for binding and hydrolyzing the protein substrate (*Fig. 3*). The appearance of bands corresponding to polypeptides with molecular weights ranging from 40 to 60 kDa on the electrophoretic image of the incubated reaction mixture indicates the possibility of
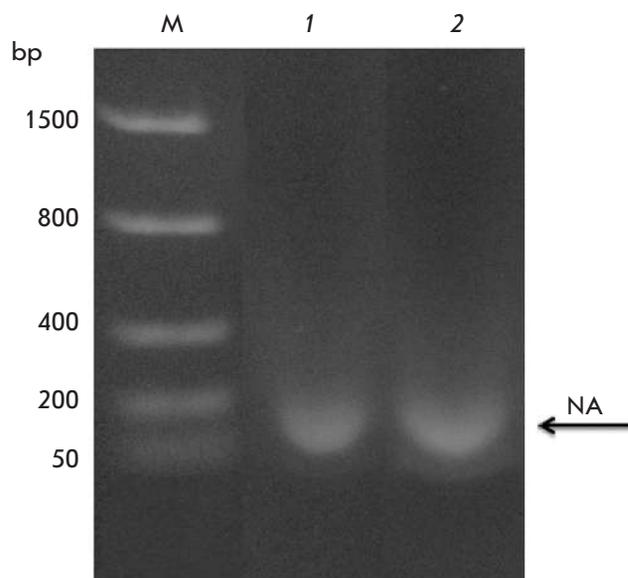
Fig. 5. Phenolic extracts of Lon-H$_6$ (1) and Lon-dHI(CC) (2) samples. M — markers, NA — nucleic acid.

self-degradation of Lon-dHI(CC) under the conditions used for the monitoring of the hydrolysis of the target protein.

Identification of an autolytic activity of Lon-dHI(CC), which accompanies the potential hydrolysis of the protein substrate, required a study of the autolysis process itself. The intact Lon-H$_6$ protease was shown to be resistant to self-degradation in the presence of any nucleotide effector (*Fig. 4*). However, during a prolonged incubation (24 hours or more), weak autolysis of Lon-H$_6$ is detected in the absence of effectors or in the presence of magnesium ions (*Fig. 4*), which agrees with the previously obtained results [23].

In contrast to Lon-H$_6$, the deletion form Lon-dHI(CC) is unstable and it undergoes autolysis both in the absence and presence of nucleotide effectors: moreover, the autolysis of Lon-dHI(CC) is most pronounced in the presence of Mg ions (*Fig. 4*).

Thus, the loss of the HI(CC) domain leads to a complete loss of the ability of Lon-H$_6$ protease to hydrolyze the protein substrate and destabilizes the structure of the enzyme.

Binding of the nucleic acid by Lon-H$_6$ protease and its deletion form Lon-dHI(CC)

An important characteristic of Ec-Lon is its ability to bind DNA [8−10], but the site of the interaction between the enzyme and nucleic acid has not been localized to date. Since other ATP-dependent proteases of the quality control system of cellular proteins do not have DNA-binding properties and do not contain the characteristic inserted HI(CC) domain typical of LonA
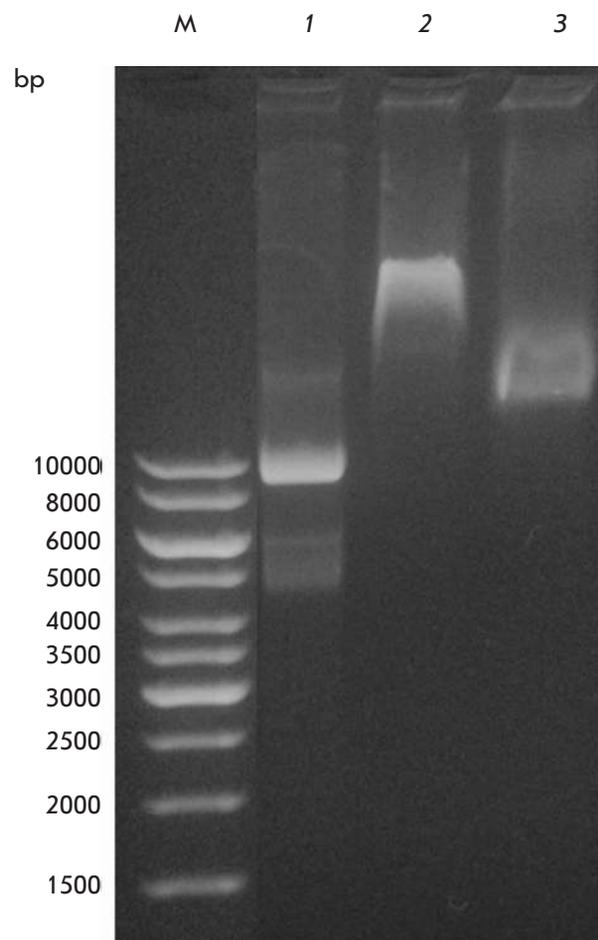
Fig. 6. DNA-binding ability of Lon-H$_6$ and Lon-dHI(CC). Experimental conditions: 20 mM Tris-HCl buffer, pH 7.5; 60 mM NaCl; 25°C; DNA (pET28a) − 28 nM (*1 — 3*); Lon-H$_6$ − 33.9 μM (*2*), Lon-dHI(CC) − 22.2 μM (*3*); M − markers.

proteases, the HI(CC) domain can be expected to be involved in nucleic acid binding. Therefore, we examined the content of nucleic acid in the preparations of Lon-H$_6$ protease and its deletion form obtained in the present study.

The DNA content in the preparations of both enzymes, determined from the ratio of optical absorption (A_{260}/A_{280}) in solutions of Lon-H$_6$ and Lon-dHI(CC) (1.09 and 1.06, respectively), did not exceed 5%. The enzyme-bound nucleic acid was isolated from the preparations by the phenol-chloroform extraction method. Treatment of the extracts with benzonase (nonspecific nuclease, Sigma) resulted in exhaustive hydrolysis of the targets, which confirms their classification as nucleic acids. At the same time, both extracts were resistant to treatment with RNase A. These results indicate that both the full-length and deletion forms of Ec-Lon are isolated from *E. coli* cells as complexes with DNA.

Phenol-chloroform extracts were analyzed by electrophoresis in 1% agarose gel, followed by staining with ethidium bromide (*Fig. 5*). It was found that the preparations of both intact Lon-H$_6$ protease and Lon-dHI(CC) contain a significant amount of bound DNA in the form of fragments of about 150 bp in size.

In addition, it turned out that both forms of Lon protease are capable of binding additional amounts of nucleic acid. It was shown that the incubation of plasmid DNA with Lon-H$_6$ or with Lon-dHI(CC) leads to the formation of DNA enzyme complexes and to a change in the mobility of nucleic acid during electrophoresis in an agarose gel (*Fig. 6*).

The presented data suggest that the HI(CC) domain of the Ec-Lon protease either does not participate in the interaction with nucleic acid or is not determinant in this interaction.

CONCLUSION

According to the obtained data, the characteristic inserted HI(CC) domain of Ec-Lon protease is necessary for the formation and correct functioning of the enzyme ATPase center. At the same time, the HI(CC) domain does not affect the formation of the peptidase center of Ec-Lon, but it is extremely important for the mutual influence of active sites. It should be emphasized that even though the activity of the peptidase center is retained, deletion of the HI(CC) domain leads to a complete loss of the proteolytic activity of the enzyme, which demonstrates the importance of this domain for the binding and hydrolysis of the protein substrate by Ec-Lon protease.

Interestingly, the deletion forms of the Lon protease from *Brevibacillus thermoruber* (Bt-Lon) [24] without the fragment (246-259) or (248-256) in the coiled-coil (CC) region lose all three types of activity. The discrepancy in the evaluation of the functioning of the peptidase center in the deletion forms of LonA proteases, revealed by comparing the results of this study and the data in [24], may be due to the use of different substrates in the testing of the peptidase center: thiobenzyl ester of the N-protected tripeptide (Suc-Phe-Leu-Phe-SBzl) in our work and 4-methoxy-β-naphthylamide of a less specific tripeptide (Glt-Ala-Ala-Phe-MNA) in [24].

We believe that the identified intensive autolysis of Lon-dHI(CC) is caused by the loss of its ability to efficiently bind nucleotides, a property that is a stabilizing factor for a full-length enzyme. The suggestion that the HI(CC) domain plays the role of a nucleic acid binding site in the Ec-Lon protease has not been experimentally confirmed.

Therefore, it can be concluded that the inserted HI(CC) domain of Ec-Lon-protease is necessary for the formation of a functionally active structure of the enzyme and the implementation of protein-protein interactions. ●

REFERENCES
1. Gottesman S., Wickner S., Maurizi M.R. // Genes Dev. 1997. V. 11. P. 815−823.
2. Tyedmers J., Mogk A., Bukau B. // Nat. Rev. Mol. Cell. Biol. 2010. V. 11. P. 777−788.
3. Mogk A., Haslberger T., Tessarz P., Bukau B. // Biochem. Soc. Trans. 2008. V. 36. P. 120−125.
4. Lee I., Suzuki C.K. // Biochim. Biophys. Acta. 2008. V. 1784. P. 727−735.
5. Rotanova T.V., Melnikov E.E., Khalatova A.G., Makhovskaya O.V., Botos I., Wlodawer A., Gustchina A. // Eur. J. Biochem. 2004. V. 271. P. 4865−4871.
6. Goldberg A.L., Moerschell R.P., Chung C.H., Maurizi M.R. // Meth. Enzymol. 1994. V. 244. P. 350−375.
7. Charette M.F., Henderson G.W., Markovitz A. // Proc. Natl. Acad. Sci. USA. 1981. V. 78. P. 4728−4732.
8. Fu G.K., Smith M.J., Markovitz D.M. // J. Biol. Chem. 1997. V. 272. P. 534−538.
9. Lee A.Y.L., Hsu C.H., Wu S.H. // J. Biol. Chem. 2004. V. 279. P. 34903−34912.
10. Liu T., Lu B., Lee I., Ondrovicova G., Kutejova E., Suzuki C.K. // J. Biol. Chem. 2004. V. 279. P. 13902−13910.
11. Lupas A.N., Martin J. // Curr. Opin. Struct. Biol. 2002. V. 12. P. 746−753.
12. Iyer L.M., Leipe D.D., Koonin E.V., Aravind L. // J. Struct. Biol. 2004. V. 146. P. 11−31.
13. Rotanova T.V., Melnikov E.E. // Biochemistry (Moscow) Suppl. Series B: Biomed. Chem. 2010. V. 4. P. 404−408.
14. Rotanova T.V., Dergousova N.I., Morozkin A.D. // Russ. J. Bioorgan. Chem. 2013. V. 39. P. 268−282.
15. Andrianova A.G., Kudzhaev A.M., Serova O.V., Dergousova N.I., Rotanova T.V. // Russ. J. Bioorgan. Chem. 2014. V. 40. P. 620−627.
16. Bradford M.M. // Anal. Biochem. 1976. V. 72. P. 248−254.
17. Laemmli U.K. // Nature. 1970. V. 227. P. 680−685.
18. Maniatis T., Fritsch E.F., Sambrook J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1982.
19. Bencini D.A., Wild J.R., O'Donovan G.A. // Anal. Biochem. 1983. V. 132. P. 254−258.
20. Castillo M.J., Nakajima K., Zimmerman M., Powers J.C. // Anal. Biochem. 1979. V. 99. P. 53−64.
21. Lee A.Y.L., Tsay S.S., Chen M.Y., Wu S.H. // Eur. J. Biochem. 2004. V. 271. P. 834−844.
22. Melnikov E.E., Tsirulnikov K.B., Rotanova T.V. // Russ. J. Bioorgan. Chem. 2000. V. 26. P. 474−481.
23. Kudzhaev A.M., Andrianova A.G., Serova O.V., Arkhipova V.A., Dubovtseva E.S., Rotanova T.V. // Russ. J. Bioorgan. Chem. 2015. V. 41. P. 518−524.
24. Chir J.L., Liao J.H., Lin Y.C., Wu S.H. // Biochem. Biophys. Res. Commun. 2009. V. 382. P. 762−765.

Bacteriolytic Activity Of Human Interleukin-2, Chicken Egg Lysozyme In The Presence Of Potential Effectors

P. A. Levashov [1*], D. A. Matolygina[1], E. D. Ovchinnikova[1], D. L. Atroshenko[1], S. S. Savin[1], N. G. Belogurova[1], S. A. Smirnov[1], V. I. Tishkov[1,2]*, A. V. Levashov[1]

[1]Department of Chemical Enzymology, Faculty of Chemistry, M.V. Lomonosov Moscow State University, Leninskie Gory, 1–3, Moscow, 119991, Russia

[2]Bach Institute of Biochemistry, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky Ave. 33, bldg. 2, Moscow, 119071, Russia

*E-mail: levashov@yahoo.com, vitishkov@gmail.com

Received: December 14, 2016; in final form February 22, 2017

ABSTRACT The bacteriolytic activity of interleukin-2 and chicken egg lysozyme in the presence of various substances has been studied. Glycine and lysine do not affect the activity of interleukin-2 but increase that of lysozyme, showing a bell-shape concentration dependence peaking at 1.5 mM glycine and 18 mM lysine. Arginine and glutamate activate both interleukin-2 and lysozyme with a concentration dependence of the saturation type. Aromatic amino acids have almost no effect on the activity of both interleukin-2 and lysozyme. Aromatic amines, tryptamine, and tyramine activate interleukin-2 but inhibit lysozyme. Peptide antibiotics affect interleukin and lysozyme similarly and exhibit maximum activity in the micromolar range of antibiotics. Taurine has no effect on the activity of interleukin-2 and lysozyme. Mildronate showed no influence on lysozyme, but it activated interleukin-2 with the activity maximum at 3 mM. EDTA activates both interleukin-2 and lysozyme at concentrations above 0.15 mM.

KEYWORDS bacteriolytic activity, chicken egg lysozyme, interleukin-2.

INTRODUCTION

Interleukin-2 plays a key role in the regulation of the immune system and is used as medication for various oncological diseases [1, 2]. This cytokine was recently shown to exhibit bacteriolytic acitivity [3–6]. The physiological significance of the recently identified bacteriolytic activity for this important cytokine is unclear. Interleukin-2 shows a substrate specificity distinct from that of chicken egg lysozyme [3–6]. However, there are microorganisms that are affected by both interleukin-2 and lysozyme. This work has aimed at identifying the potential effectors of interleukin and lysozyme activity by a direct comparison under identical experimental conditions. A series of amino acids of various types, biogenic amines, peptide antibiotics, EDTA, and mildronate were selected as model compounds, since biological systems may contain these compounds or their analogs. *Escherichia coli* cells were taken as the model substrate, because they undergo lysis with both interleukin-2 and lysozyme [3–5]. This study on the character of the effect of various additives may help in future elucidation of the mechanism of interleukin-2 bacteriolytic activity. In addition, an understanding of the peculiarities of the effects of various compounds on interleukin-2 and lysozyme activity may provide a clue in future efforts directed towards enhancing the efficiency of existing medication, as well as designing new ones.

EXPERIMENTAL

The following materials were used: glycine (Fluka, Germany); EDTA (Panreac, Spain); L-lysine (Serva, Germany); tyramine, triptamine, taurine (Acros Organics, USA), Tris, MES (Amresco, USA); bacitracin (MP Biomedicals, Germany); polymyxin B, L-tryptophane, L-tyrosine, L-phenylalanine, chicken egg lysozyme (Sigma-Aldrich, USA); NaOH (Merck, Germany); acetic acid (ChemMed, Russia); hydrochloric acid (Laverna, Russia); mildronate (2- (2-carboxylatoethyl)-1,1,1-trimethylhydrazinium) (Cridex, Latvia); sodium L-glutamate (HongMei (红梅), China); Roncoleukin®, the 0.25 mg/mL solution of purified recombinant interleukin-2 for intravenous and subcutaneous injections (Biotech, Russia).

The *E. coli* JM109 strain used in this work was provided by Dr. J.Messing (Waksman Institute, New Jersey, USA). The cells were grown in accordance with the standard protocol [7]. The 10^9 CFU/mL cell sus-

pension in 0.15 M NaCl was frozen by immersing 1 mL aliquots into liquid nitrogen. The cells were stored at −70°C for no longer than for 2−3 weeks. The cells were thawed right before the experiment. The thawed cell suspension was centrifuged at 4,500 rpm for 5 min in a Minispin centrifuge (Eppendorf, Germany) and then re-suspended in the assay buffer.

Bacteriolytic activity (as the rate of cell lysis) was measured turbidimetrically by following the decrease in the suspension absorbance, $-dA/dt$, min^{-1} [5, 8] at 650 nm, which is linearly dependent on the rate of cell count changes, $dCFU/dt$, under these conditions. The measurements were taken in a cuvette with a 1-cm light path and 0.5mL volume; the absorbance was measured on a UV-1800 spectrophotometer (Shimadzu, Japan). A lysozyme solution was prepared right before the experiment by dissolving in the assay buffer. The commercial solution of interleukin-2 was used without additional purification, and an ampoule was opened just before the experiment. The bacteriolytic activity measurements were assayed at 37°C in a 10 mM MES-Tris-acetate buffer, pH 8.8 for interleukin-2, and pH 8.5 for lysozyme. The final concentrations of interleukin-2 and lysozyme were equal to 15 µg/mL and 0.1 µg/mL, respectively, to ensure comparable values of cell lysis rates. The cell suspension was mixed with the buffer in the cuvette to achieve an initial absorbance (A_{650}) of 0.43−0.45. The background changes in the absorbance were recorded for 5 min to account for the cell's self-lysis or precipitation. Then, the effectors under study were added and the background absorbance changes recorded for 5 min; this was followed by the addition of the enzyme. The initial rate of cell lysis was determined from the absorbance changes in a timeframe from 5−25 s after enzyme addition. The background rate for cell self-lysis or precipitation was subtracted from the initial rate of cell lysis in the presence of the enzyme. In all experiments, the background rate value did not exceed the average value of a standard deviation for the cell lysis rates determined in the enzyme presence. All added compounds (except for the enzyme) did not change the background lysis rates within the experimental error. The pH value for the compounds under study was tested before the addition and adjusted to 8.8 (8.5) with NaOH or HCl solutions if necessary. The effects of the additives observed in this experiment did not originate from the activity changes caused by the changes in the ionic strength: within the range of ionic strength changes in this work, no significant changes in the bacteriolytic activity were observed [3, 8].

RESULTS AND DISCUSSION

The dependences of interleukin-2 and lysozyme activity on the concentration of glycine, lysine, arginine, and glutamate are shown in *Fig. 1*. As seen in *Figs. 1A and 1B*, the activity of interleukin-2 in the presence of glycine, the simplest based on structure natural amino acid, and positively charged lysine remained unchanged. For lysozyme, a maximum was observed at 2 mM glycine or 15−18 mM lysine, where lysozyme activity was significantly higher than the original. A further increase in the concentrations of glycine and lysine returned the lysozyme activity to its original level. Hence, lysozyme and interleukin-2 show completely different behaviors in the presence of these two amino acids, and this may point to the difference in their mechanisms of action. Such effect of lysozyme activity enhancement in the presence of glycine has never been reported in the literature. However, it is known that glycine, in addition to its bacteriostatic properties, may increase the efficiency of various antimicrobial agents [9]. The distinct action of glycine on lysozyme and interleukin-2 is difficult to explain. One may speculate that glycine affects one of the bacterial-type porines to ease the lysozyme interaction with the cell wall, and that at the same time it has no effect on the action of interleukin-2.

The effect of arginine on lysozyme and interleukin-2 bacteriolytic activity is shown in *Fig. 1C*. As seen, in both cases, a significant increase in cell lysis rates is observed at effector concentrations of 10 mM and higher. The activation by arginine could be of a complex nature and reflect a combination of arginine effects on the enzyme and the cell: it is well-known that arginine enhances the efficiency of lysozyme-based pharmaceuticals by diminishing protein aggregation [10]. It is also necessary to mention that the dependences of arginine and lysine on the bacteriolytic activity show stark differences. Probably, this difference is due to the various polarities and geometries of positively charged side chains.

A similar trend in the changes in lysozyme and interleukin-2 activity is observed in the presence of glutamate: a 2-fold increase for lysozyme and 3-fold increase for interleukin-2 at 15 mM glutamate. Further increase in the glutamate concentrations does not significantly change this activity, which approaches a manner of threshold. A similar effect by glutamate on the activity of lysozyme and interleukin-2 can be explained based on the hypothesis that glutamate forms a complex with positively charged groups on the protein surface, preventing various types of nonproductive enzyme adsorption on cells, which may significantly change the apparent values of bacteriolytic activity parameters [11, 12].

The dependence of lysozyme and interleukin-2 activity on the concentration of aromatic amino acids is shown in *Fig. 2*. For tyrosine, the highest concentration used was restricted to 0.6 mM because of its low solubil-

A

B

C

D
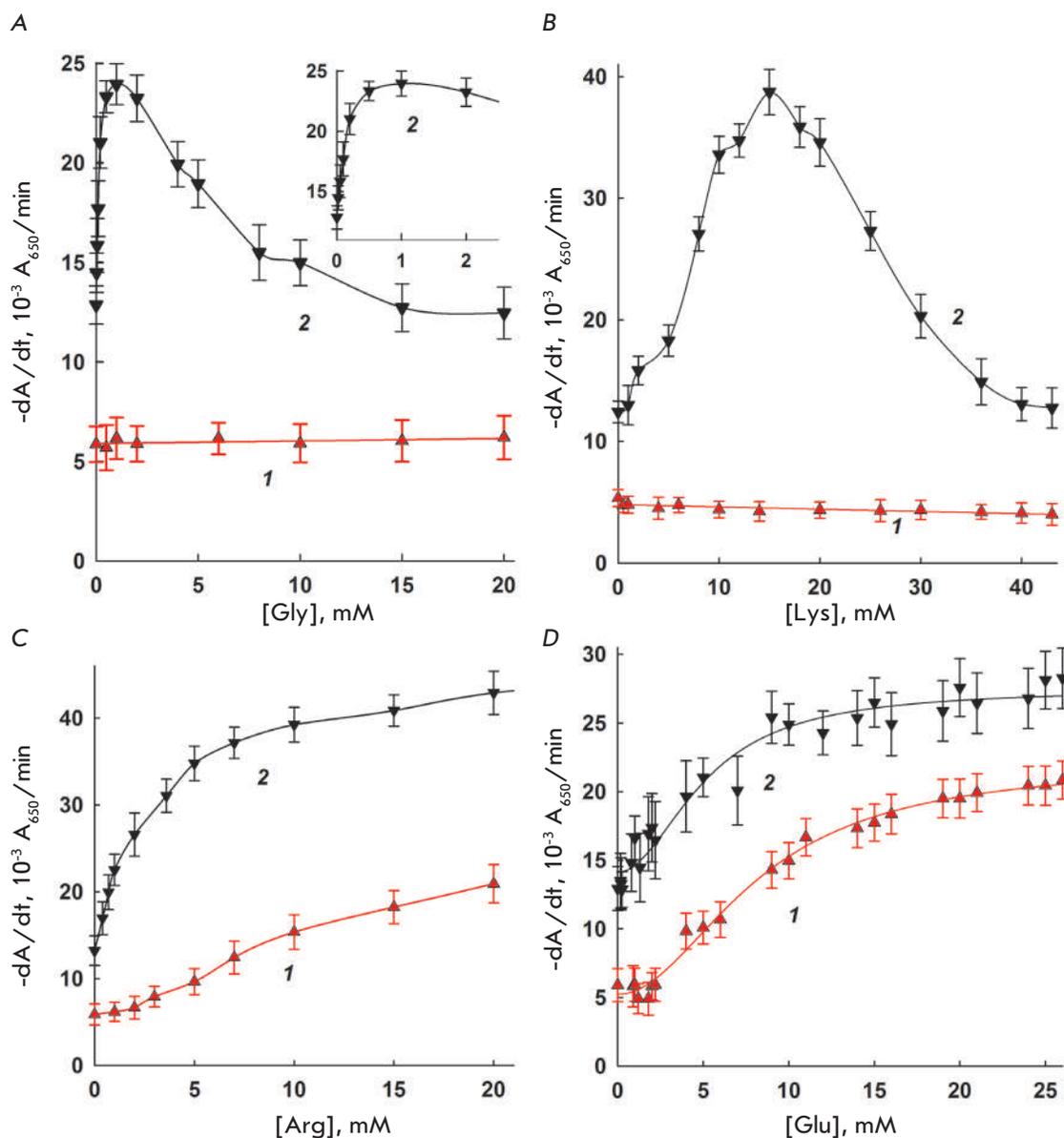
Fig. 1. The dependence of interleukin-2 (1) and lysozyme (2) activity on the concentration of added glycine (1A), lysine (1B), arginine (1C), and glutamate (1D). 37°C, 10 mM MES-Tris-acetate buffer, pH 8.8, and pH 8.5 for interleukin-2 and lysozyme, respectively.

ity in water. As seen, in the presence of phenylalanine and tryptophan, the small reduction in lysozyme activity is negligible within the experimental error. The apparent increase in lysozyme activity in the presence of tyrosine is also within the experimental error. Interleukin-2 activity in the presence of phenylalanine and tryptophan is unchanged. The dependence of interleukin-2 activity on the tyrosine concentration shows a 30% increase at 0.25−0.3 mM. The general conclusion is that aromatic amino acids have no significant effect on the activity of lysozyme, as well as on interleukin-2. A completely different picture emerges for aromatic amino acid derivatives: namely, biogenic aromatic amines − tryptamine and tyramine − as discussed below.

The dependence of interleukin-2 and lysozyme activity on the concentrations of the biogenic amines tyramine and tryptamine, which can be formally considered as derivatives of the tyrosine and tryptophan amino acids, is shown in *Fig. 3*. As can be seen, interleukin-2 is activated by either biogenic amine, whereas the activity of lysozyme is inhibited. This result may be used as proof of the substantive differences between interleukin-2 and lysozyme with respect to their mechanism of action. Interleukin-2 is prone to binding to various ligands via hydrophobic interactions [13]: hence, it is possible that tyramine and tryptamine bind to some hydrophobic loci on the interleukin-2 surface, lowering its nonproductive adsorption on cells.
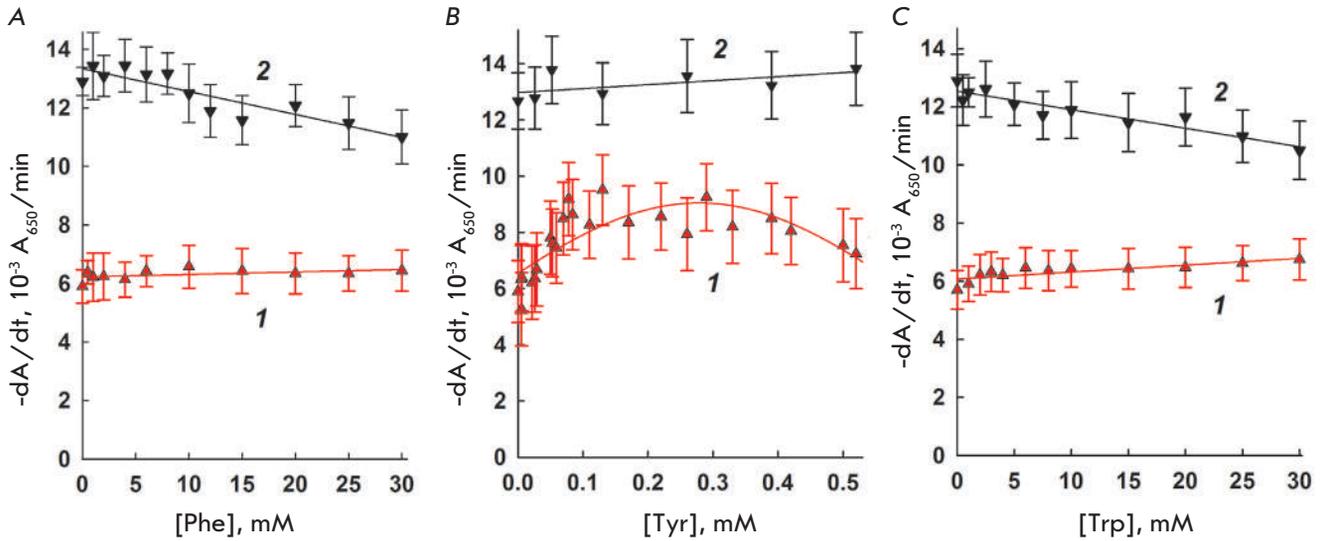
Fig.2. The dependence of interleukin-2 (1) and lysozyme (2) activity on the concentration of added phenylalanine (2A), tyrosine (2B), and tryptophan (2C). 37°C, 10 mM MES-Tris-acetate buffer, pH 8.8, and pH 8.5 for interleukin-2 and lysozyme, respectively.

The dependence of interleukin-2 and lysozyme activity on the concentrations of the peptide antibiotics polymyxin B and bacitracin is shown in *Fig. 4*. A similar picture is observed for both bacteriolytic factors and both antibiotics: an activity maximum at 5−7 μM. These peptide antibiotics are known cytostatics for *E.coli* [14, 15]: hence, the similarity in the observed effects may originate from their direct action on the cells and not from a modulation of the properties of bacteriolytic factors. The antibiotic by itself cannot cause cell lysis but renders a cell more sensitive to bacteriolytic enzymes, as was observed for endolysine from bacteriohages [16].

The dependence of interleukin-2 and lysozyme activity on the concentrations of mildronate, taurine,

and EDTA is shown in *Fig. 5*. Mildronate has no effect on the activity of lysozyme but increases the activity of interleukin-2: the maximum is observed at 3 mM. The physiological effects of mildronate are usually explained by its similarity to natural, biologically active compounds, and γ-butyrobetaine in particular or its derivatives: for example, L-carnitine [17, 18]. Mildronate binds to and inhibits γ-butyrobetaine hydroxylase (IC_{50} = 62 μM) and carnitine acetyltransferase (IC_{50} = 1.6 mM). So, it may also bind other proteins and change their conformation and properties. Taurine has no effect on the activity of interleukin-2 and lysozyme. EDTA at concentrations above 0.1 mM enhances the effect of both bacteriolytic factors, and similarly to peptide antibiotics, its effect, at least in part, can be

Fig.3. The dependence of interleukin-2 (1) and lysozyme (2) activity on the concentration of added tyramine (3A) and tryptamine (3B). 37°C, 10 mM MES-Tris-acetate buffer, pH 8.8, and pH 8.5 for interleukin-2 and lysozyme, respectively.
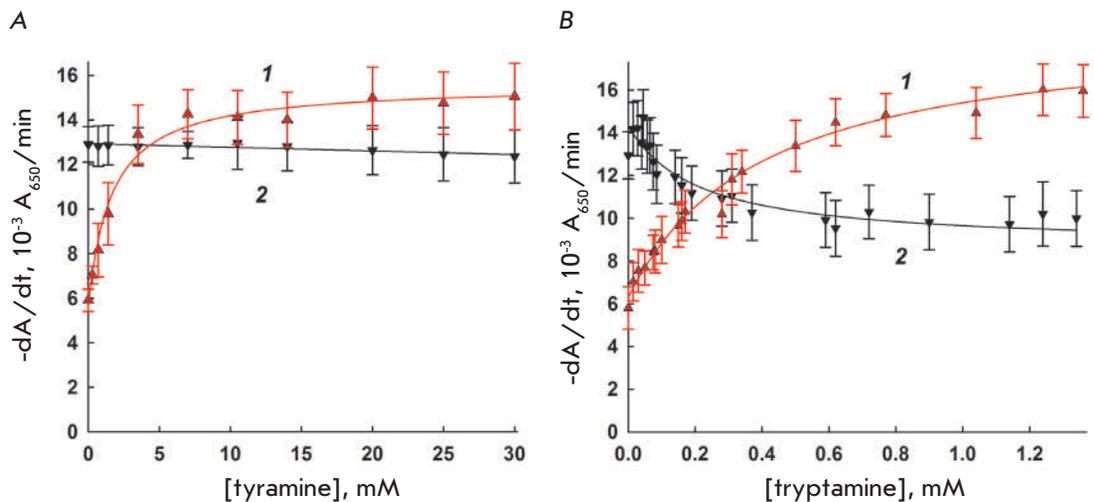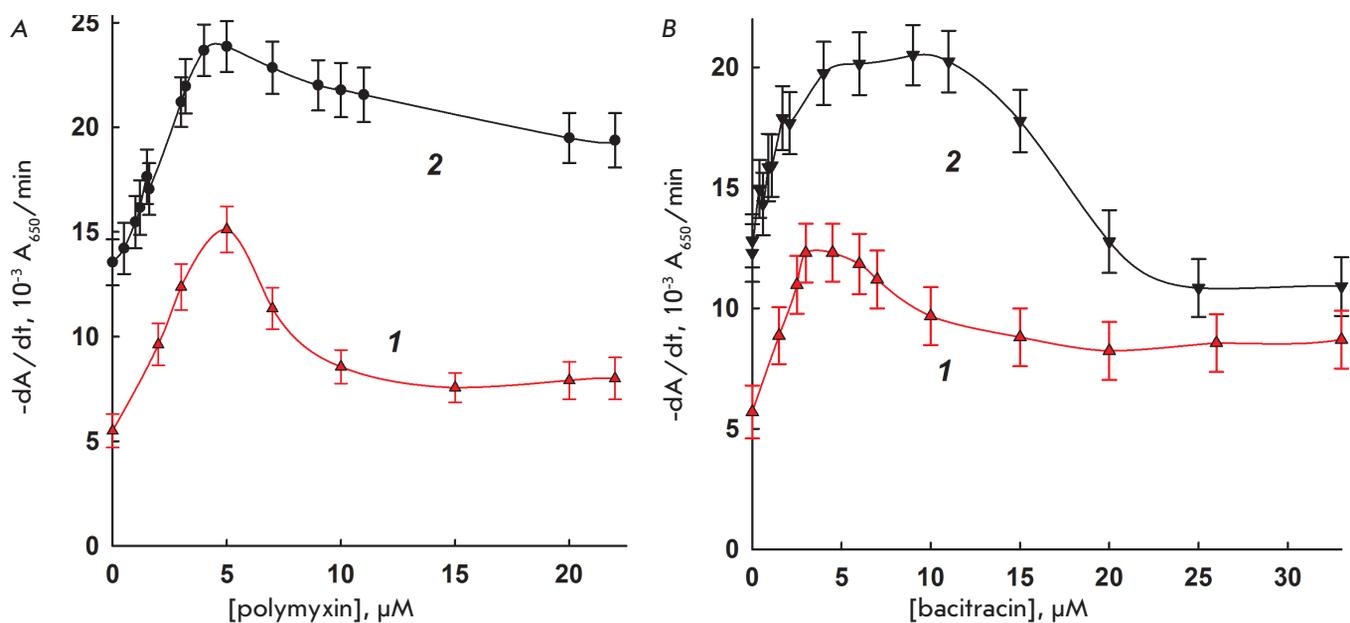
Fig.4. The dependence of interleukin-2 (1) and lysozyme (2) activity on the concentration of added polymyxin B (4A) and bacitracin (4B). 37°C, 10 mM MES-Tris-acetate buffer, pH 8.8, and pH 8.5 for interleukin-2 and lysozyme, respectively.
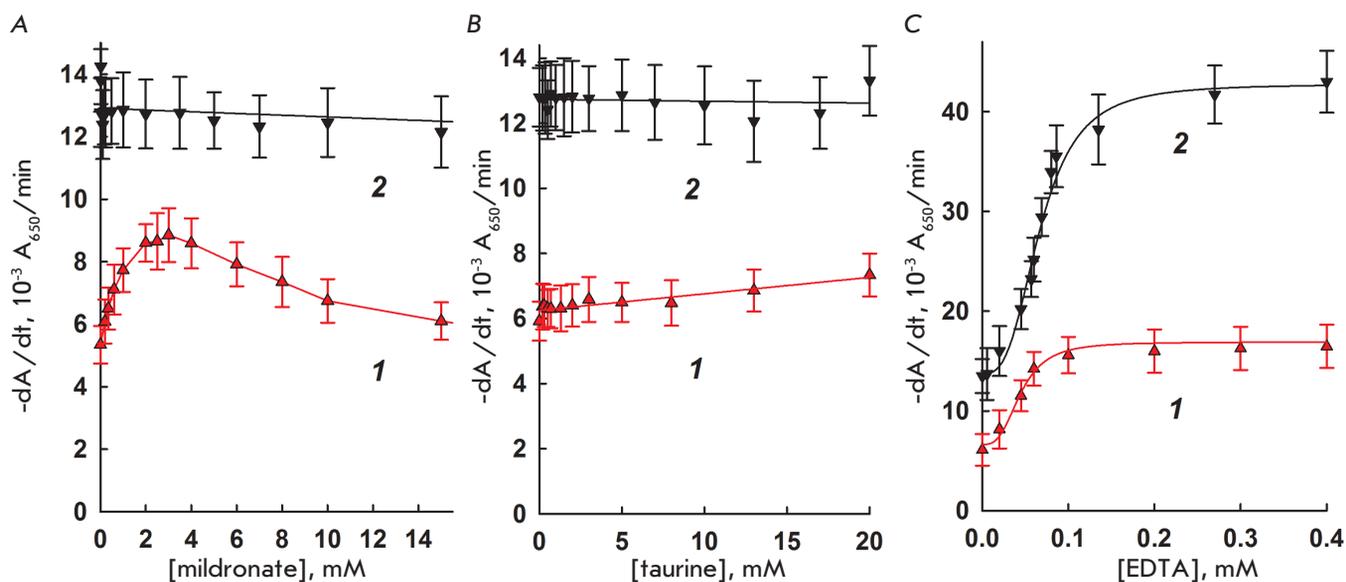
Fig.5. The dependence of interleukin-2 (1) and lysozyme (2) activity on the concentration of added mildronate (5A), taurine (5B), and EDTA (5C). 37°C, 10 mM MES-Tris-acetate buffer, pH 8.8, and pH 8.5 for interleukin-2 and lysozyme, respectively.

explained by the effect of EDTA on cells, and not on the enzyme.

CONCLUSION

Thus, the effect of additives on interleukin-2 and lysozyme depends on the chemical nature of the additives. This can be indicative of different mechanisms of action. We have identified substances which activate these bacteriolytic factors. This can be of practical importance. Effectors can be used to improve the effectiveness of existing medication, as well as to create new medicinal compositions. For example, our research

shows that glycine, lysine, and glutamate enhance the bacteriolytic activity of lysozyme. Glycine, lysine, and lysozyme are widely used as drugs, but their combined action has not been studied. The effect of glutamate and arginine on the activity of lysozyme had also not been investigated previously. In current medical practice, interleukin-2 is used as a regulator of the immune system but not as a bacteriolytic factor, since its bacteriolytic properties had not been previously known. However, it is possible that antimicrobial properties also play an important role in some cases when the effectiveness of interleukin-2 is confirmed. Interleukin-2 is used both in the case of sepsis, where the role of bacteria is obvious, and in the treatment of cancer, where the role of bacteria is less obvious but there may be a combination of bacterial tissue damage and the underlying disease. The mechanism of bacteriolytic action of interleukin-2 has not yet been established, and the mechanism of action of effectors on interleukin-2 activity also requires further investigation. It has become clear that special attention should be focused on the activation of interleukin-2 in the presence of additives: for example, mildronate, arginine, and glutamate. Combined use of these drugs could open new possibilities in the treatment of serious diseases. ●

REFERENCES
1. Mizui M., Tsokos G.C. //Curr. Rheumatol. Rep. 2016. V. 18. № 11. P. 68.
2. Gill D.M., Stenehjem D.D., Parikh K., Merriman J., Sendilnathan A., Agarwal A.M., Hahn A.W., Gupta S., Tantravahi S.K., Samlowski W.E., Agarwal N. // Ecancermedicalscience. 2016. V. 10. P. 676.
3. Sedov S.A., Belogurova N.G., Shipovskov S.V., Semenova M.V., Gitinov M.M., Levashov A.V., Levashov P.A. // Rus. J. Bioorg. Chem. 2012. V. 38. № 3. P. 274–281.
4. Levashov P.A., Sedov S.A., Belogurova N.G., Shipovskov S.V., Levashov A.V. // Biochemistry (Moscow). 2012. V. 77. № 11. P. 1312–1314.
5. Levashov P.A., Matolygina D.A., Osipova H.E., Savin S.S., Zaharova G.S., Gasanova D.A., Belogurova N.G., Ovchinnikova E.D., Smirnov S.A., Tishkov V.I., Levashov A.V. // Moscow Univ. Chem. Bull. 2015. V. 70. № 6. P. 257–261.
6. Levashov P.A., Ovchinnikova E.D., Morozova O.A., Matolygina D.A., Osipova H.E., Cherdyntseva T.A., Savin S.S., Zakharova G.S., Alekseeva A.A., Belogurova N.G., Smirnov S.A., Tishkov V.I., Levashov A.V. // Acta Naturae. 2016. V. 8. № 1(28). P. 98–102.
7. Maniatis T., Fritsch E.F., Sambrook J. // Molecular Cloning: A Laboratory Manual, New York: Cold Spring Lab. Press. 1982.
8. Levashov P.A., Sedov S.A., Shipovskov S.V., Belogurova N.G., Levashov A.V. // Anal. Chem. 2010. V. 82. P. 2161–2163.
9. Minami M., Ando T., Hashikawa S.,Torii K., Hasegawa T., Israel D.A., Ina K., Kusugami K., Goto H., Ohta M. //Antimicrob Agents Chemother. 2004. V. 48. № 10. P. 3782–3788.
10. Shah D, Shaikh A.R. // J. Biomol. Struct. Dyn. 2016. V. 34. № 1. P. 104–114
11. Matolygina D.A., Osipova H.E., Smirnov S.A., Belogurova N.G., Eremeev N.L., Tishkov V.I., Levashov A.V., Levashov P.A. // Moscow Univ. Chem. Bull.. 2015. V. 70. № 6. P. 262–267.
12. Sedov S.A., Belogurova N.G., Shipovskov S., Levashov A.V., Levashov P.A. // Colloids Surf. B. 2011. V. 88. P. 131–133.
13. Afonin P.V., Fokin A.V., Tsygannik I.N., Mikhailova I.Yu., Onoprienko L.V., Mikhaleva I.I., Ivanov V.T., Mareeva T.Yu, Nesmeyanov V.A., Li N., Pangborn, Duax W.L., Pletnev V.Z. //
Protein Sci. 2001. V. 10. № 8. P. 1514–1521.
14. Zong L., Teng D., Wang X., Mao R., Yang N., Hao Y., Wang J. // Appl. Microbiol. Biotechnol. 2016. V. 100. № 11. P. 5045–5057.
15. Yu Zh., Qin W., Lin J., Fang Sh., Qiu J. // Biomed. Res. Int. 2015. P. 679109.
16. Levashov P.A., Belogurova N.G., Sedov S.A., Klyachko N.L., Levashov A.V., Popov D.V., Popova V.M., Zhilenkov E.L., Dyatlov I.A., Morozova O.A. // Biochemistry (Moscow). 2010. V. 75. № 9. P. 1160–1164.
17. Liepinsh E., Konrade I., Skapare E., Pugovics O., Grinberga S., Kuka J., Kalvinsh I., Dambrova M.J. // Pharm. Pharmacol. 2011. V. 63. № 9. P. 1195–1201.
18. Beitnere U., Dzirkale Z., Isajevs S., Rumaks J., Svirskis S., Klusa V. // Eur. J. Pharmacol. 2014. V. 745. P. 76–83.

Comparing New-Generation Candidate Vaccines against Human Orthopoxvirus Infections

R. A. Maksyutov[1]*, S. N. Yakubitskyi[1], I. V. Kolosova[1], S. N. Shchelkunov[1,2]*

[1]State Research Center of Virology and Biotechnology «Vector», Koltsovo, Novosibirsk region, 630559, Russia

[2]Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, 630090, Russia

E-mail: maksrinat@yandex.ru; snshchel@rambler.ru

Received: July 05, 2016; in final form January 23, 2017

ABSTRACT The lack of immunity to the variola virus in the population, increasingly more frequent cases of human orthopoxvirus infection, and increased risk of the use of the variola virus (VARV) as a bioterrorism agent call for the development of modern, safe vaccines against orthopoxvirus infections. We previously developed a polyvalent DNA vaccine based on five VARV antigens and an attenuated variant of the vaccinia virus (VACV) with targeted deletion of six genes (VACΔ6). Independent experiments demonstrated that triple immunization with a DNA vaccine and double immunization with VACΔ6 provide protection to mice against a lethal dose (10 LD_{50}) of the ectromelia virus (ECTV), which is highly pathogenic for mice. The present work was aimed at comparing the immunity to smallpox generated by various immunization protocols using the DNA vaccine and VACΔ6. It has been established that immunization of mice with a polyvalent DNA vaccine, followed by boosting with recombinant VACΔ6, as well as double immunization with VACΔ6, induces production of VACV-neutralizing antibodies and provides protection to mice against a 150 LD_{50} dose of ECTV. The proposed immunization protocols can be used to develop safe vaccination strategies against smallpox and other human orthopoxvirus infections.

KEYWORDS DNA vaccine, vaccinia virus, virulence genes, protective potential, smallpox.

ABBREVIATIONS CPXV − cowpox virus, ECTV − ectromelia virus, LD_{50} − 50% lethal dose, LIVP − L-IPV vaccinia virus strain, MPXV − monkeypox virus, PCR − polymerase chain reaction, PFU − plaque forming unit, VACV − vaccinia virus, VARV − variola virus.

INTRODUCTION

The *Orthopoxvirus* genus of the Poxviridae family includes human-pathogenic species, such as the variola virus (VARV), monkeypox virus (MPXV), cowpox virus (CPXV), and vaccinia virus (VACV). Mass vaccination with a conventional VACV-based vaccine protects not only from VARV, but also from the closely related MPXV and CPXV [1]. After 1980, the share of the population sensitive to VARV and other orthopoxviruses pathogenic to humans has constantly increased due to the eradication of smallpox and cessation of widespread immunization against the disease. This is evidenced in the increasingly more frequent multiple cases of orthopoxvirus infections in humans caused by such viruses as MPXV, CPXV, and VACV [2−6]. Moreover, VARV is considered a potential agent of bioterrorist attacks, which could have catastrophic consequences for the entire world population [6]. The lack of effective antiviral drugs and the risk associated with con-

ventional VACV-based live vaccines, because of severe postvaccinal complications, necessitate the development of modern, safe orthopoxvirus vaccines and protocols for their use [7, 8].

Earlier, we developed a recombinant variant VACΔ6 with targeted knockdown of six genes, encoding hemagglutinin *(A56R)*, the gamma-interferon-binding protein *(B8R)*, thymidine kinase *(J2R)*, the complement-fixing protein *(C3L)*, the Bcl2-like apoptosis inhibitor *(N1L)*, and the *A35R* gene, which controls antigen presentation by the class II major histocompatibility complex (MNSII), based on the LIVP VACV strain used in the Russian Federation for the vaccination of humans. It has been shown that inactivation of selected virulence genes does not affect the reproductive properties of VACV in mammalian cell cultures. The VACΔ6 strain is significantly less reactogenic and neurovirulent and more immunogenic compared to the parent LIVP strain. Double subcutaneous injection of recom-

binant variant VACΔ6 induces significantly higher levels of virus-neutralizing antibodies in mice than the parental LIVP strain and provides complete protection to mice against the highly pathogenic ectromelia virus (ECTV), as opposed to the effect of the LIVP strain in this model, which is approved as a smallpox vaccine [9, 10].

Earlier, we implemented another independent approach to vaccinal prevention of smallpox. We developed a polyvalent DNA vaccine based on a mixture of recombinant plasmids containing the genes of five virion proteins of the VARV: A30, F8, M1, which are constituents of the surface membrane of intracellular virions, and A36, B7, which are located on the membrane of the extracellular form of the virus, under the control of the CMV promoter. Triple intradermal immunization with a polyvalent DNA vaccine induced the production of virus-neutralizing antibodies and provided complete protection to mice against ECTV infection at a dose of $10 \, LD_{50}$ [11−13].

Along with the development of fundamentally new vaccines, a combination of various types of vaccines which can complement each other and induce strong and broad immunity is another promising avenue in improving the efficacy of smallpox vaccination [14]. Such a heterologous immunization strategy (prime-boost), where the subunit vaccine (DNA vaccine) is used to prime the immune system and where the attenuated variant of VACV is used for subsequent booster vaccination, is considered promising.

This study compared immunity against smallpox induced by double immunization with various combinations of polyvalent DNA vaccines and a highly attenuated VACΔ6 strain.

EXPERIMENTAL
Bacteria, viruses, cell cultures
In this study, we used *Escherichia coli XL2-blue,* the VACΔ6 strain [10], the LIVP VACV strain (derived from a Lister strain obtained from the Institute of Viral Preparations, Moscow), and a K-1 ECTV strain from the collection of SRC VB "Vector," continuous cell culture 4647 of African green monkey kidney cells [15] from the collection of cell cultures of SRC VB "Vector" cultivated on a DMEM medium supplemented with 10% fetal bovine serum.

Polyvalent DNA vaccine
A set of recombinant plasmids based on the vector plasmid pcDNA3.1, bearing genes of five VARV antigens, including A30, F8, M1 antigens of the surface membrane of intracellular virions and A36, B7 antigens of the membrane of extracellular forms of the virus under the control of a cytomegalovirus promoter, was

obtained previously [11−13]. Preparative quantities of plasmid DNA were accumulated in *E. coli* cells and purified using the EndoFree Plasmid Giga Kit (Qiagen, USA) according to the manufacturer's recommendations. Plasmid DNA concentration was measured spectrophotometrically on a Ultrospec 3000 pro instrument (GE Healthcare Life Sciences, USA).

Accumulation and purification of viruses
A monolayer of 4647 cells grown in culture flasks with a growth surface of 175 cm² (volume of 650 ml) was infected with VACV (VACΔ6 or LIVP strain), and the multiplicity of infection was 1 PFU/cell. The virus was incubated in a DMEM medium with 2% fetal bovine serum for 48 hours at 37°C until complete cytopathic effect, followed by the obtaining of a cryolysate (three freezing-thawing cycles) of the infected cells, and double or triple sonication of the latter in the 22 kHz MSE 500 disintegrator for 10−15 seconds. Cell debris was removed by low-speed centrifugation (10 min at 4,000 *g)*. The supernatant was centrifuged for 1.5 hours at 30,000 *g*. The precipitated virus was re-suspended in 4 ml of saline. Infectious virus titer was determined using the agar-free plaque technique in a 4647 cell monolayer.

Study of the immunogenicity and protectivity
In this study we used Balb/c mice (females, weight 14−16 g, 5−6 weeks old) from the mouse bank of the SRC VB "Vector." Mice were divided into groups of 10 animals. They were immunized with a polyvalent DNA vaccine subcutaneously and with a mixture of pcDNA-A30, pcDNA-A36, pcDNA-M1, pcDNA-F8, and pcDNA-B7 plasmids (50 μg of each plasmid, a total dose of 250 μg/100 μl per mouse) intradermally. The mice were immunized subcutaneously with VACΔ6 or LIVP strain at a dose of 10^7 PFU/100 μl per mouse. Control group mice were injected with a volume equal to that of the saline that was used to prepare virus dilutions. Immunization was performed twice at an interval of 21 days as shown in *Table 1*.

Blood samples were collected from the retrobulbar venous plexus of pre-anesthetized mice 19 days after the second immunization, incubated at 4°C for 24 hours to form a fibrin clot, and centrifuged for 10 min at 5,000 *g*. Serum preparations from one group of animals were then pooled and heated at 56°C for 30 min. Titer of VACV-neutralizing antibodies was determined on a 4647 cell culture according to [16], using serial five-fold dilutions of sera, which were mixed with an LIVP strain of VACV at the working dilution of 50 PFU/well. The effectiveness of the neutralization was calculated with respect to the number of plaques in the sera-free wells as -lg of the highest serum dilution, which provides 50% neutralization of VACV.

The animals under mild ether anesthesia were subjected to intranasal inoculation with ECTV, which is highly pathogenic to mice, at a dose of 150 $LD_{50}/20$ μl per mouse according to [17] 21 days after the second immunization. The mice were followed for 14 days, and the number of survived and dead mice was recorded.

Data analysis

The statistical significance of the experimental data was evaluated based on the Student's t-test using the Origin Professional 8.1.10.86 software. The differences were considered statistically significant at $P < 0.05$ [18].

RESULTS

Preparative quantities of pre-engineered pcDNA-A30, pcDNA-A36, pcDNA-M1, pcDNA-F8, and pcDNA-B7 plasmids were accumulated in *E. coli* cells and purified using the EndoFree Plasmid Giga Kit (Qiagen, USA) according to the manufacturer's instructions, followed by confirmation of the accuracy of insertions by restriction analysis using AsuNHI and HindIII endonucleases *(Fig. 1)* and sequencing.

VACΔ6 and LIVP vaccinia virus strains were produced in a 4647 cell culture recommended for the production of a smallpox vaccine [19] and purified according to the aforementioned method. The strains were identified using a PCR analysis based on the loci of six inactivated genes *(Tab. 2, Fig. 2)*.

The immunogenicity of double immunization with various combinations *(Tab. 1)* of the polyvalent DNA vaccine and a highly attenuated strain VACΔ6 was assessed based on the level of induced virus-neutralizing antibodies in the mice serum sampled 21 days after the second immunization. As can be seen from the data shown in *Fig.* 3, the combination of DNA & VACΔ6 vaccines induced the accumulation of VACV-neutralizing antibodies whose level was comparable to the level of antibodies induced by double vaccination with a parent-strain LIVP. Moreover, double immunization with a VACΔ6 strain induced significantly higher levels of neutralizing antibodies, which is consistent with our previous results [10].

As shown in our previous studies, triple immunization with a polyvalent DNA vaccine or double immunization with the VACΔ6 strain provides 100% protection to mice subsequently infected with ECTV at a dose of 10 LD_{50}/mouse [10, 12]. For this reason, a significantly higher resolving dose of ECTV was used, 150 LD_{50}/mouse, in order to assess the differences in the effectiveness of the used immunization protocols. As a result, a partial protective effect of double immunization (DNA & VACΔ6, LIVP & LIVP and VACΔ6 & VACΔ6) was observed in three test groups *(Fig.* 4.).
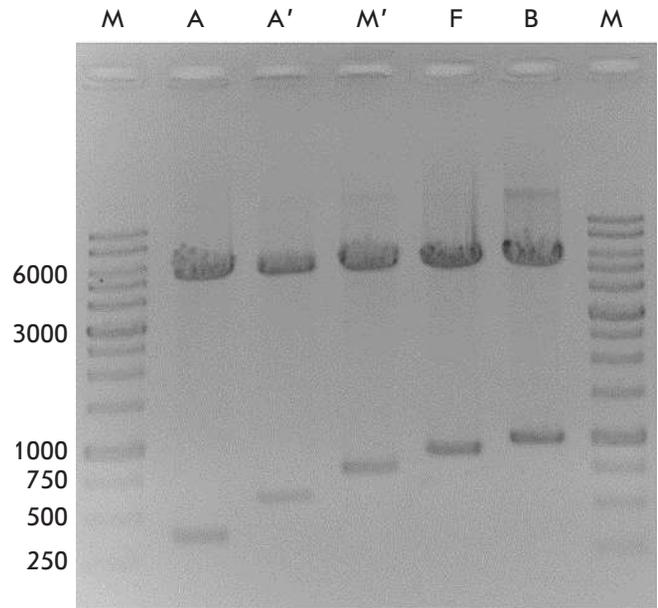
Fig. 1. Result of electrophoretic separation of DNA fragments produced after hydrolysis of the recombinant plasmid with the restriction endonucleases *Asu*NHI and *Hind*III on a 1.2% agarose gel. A, A', M', F, B – DNA fragments obtained for the recombinant plasmids pcDNA-A30, pcDNA-A36, pcDNA-M1, pcDNA-F8, and pcDNA-B7, respectively. M – DNA ladder, fragment length in bp is shown on the left
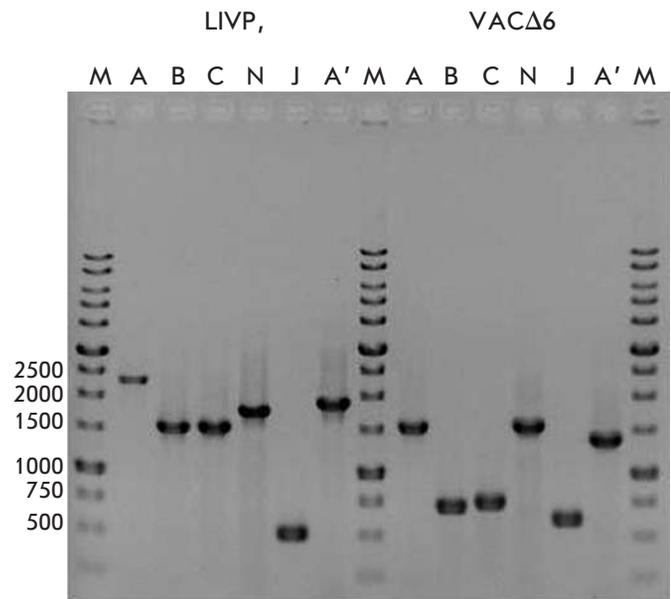
Fig. 2. Verification of deletions/insertions by PCR. PCR products formed from DNA of the parent clone VACV LIVP and VACΔ6 with deletion of six virulence genes. A, B, C, N, J, A' – PCR products obtained with the appropriate primer pairs for the *A56R, B8R, C3L, N1L, J2R, and A35R* genes. M – DNA ladder, fragment length in bp is shown on the left

Table 1. Testing scheme to assess the immunogenicity and protection of the vaccines in animal experiments

Group	Vaccine, dose per animal		Protectivity test, day 42
	1st immunization, day 1	2nd immunization, day 21	
DNA&DNA	DNA vaccine 250 μg	DNA vaccine 250 μg	К-1 strain of ECTV, 150 LD$_{50}$
DNA&VACΔ6	DNA vaccine 250 μg	VACΔ6 strain 10^7 PFU	К-1 strain of ECTV, 150 LD$_{50}$
VACΔ6&VACΔ6	VACΔ6 strain 10^7 PFU	VACΔ6 strain 10^7 PFU	К-1 strain of ECTV, 150 LD$_{50}$
LIVP&LIVP	LIVP VACV strain 10^7 PFU	LIVP VACV strain 10^7 PFU	К-1 strain of ECTV, 150 LD$_{50}$
K-	Saline	Saline	К-1 strain of ECTV, 150 LD$_{50}$

Table 2. PCR analysis aimed at identification of the recombinant VACV

Gene	Primer, nucleotide sequence (5' → 3')	LIVP strain, bp	VACΔ6 strain, bp
A56R	GTGGTATGGGACACCACAAATCCAA ATTAAACATTCCTAGAATTAATCCCGCTC	2366	1425
B8R	TCACAAATATGATGGTGATGAGCGA CGTGATATACCCTAGCCATAGGCAT	1555	737
C3L	TCGCGCTTTACATTCTCGAATCT TGTTCGTGTGTTCTTGCGGTGA	1542	751
N1L	GGGTTGGATCCTTTACACATAGATCTACTACAGGCGGAACA GGGAAAGCTTAATTTGTGAAGATGCCATGTACTACGCT	1784	1431
J2R	ATATGTTCTTCATGCCTAAACGA ATGAAGGAGCAAAAGGTTGTAAC	512	617
A35R	ACGACGGATGCTGAAGCGTGTTATA AAACGATGTTACCAATCGTTTGCTAGGT	1880	1360

Maximum survival was observed in the VACΔ6&VACΔ6 group animals, who received double vaccination with VACΔ6, and in DNA & VACΔ6 group animals, wherein the immune system was primed using the polyvalent DNA vaccine, and attenuated VACΔ6 was used for subsequent booster vaccination. All control-group animals died on the 8th day, and all DNA & DNA-group animals died on the 9th day after infection with the ectromelia virus. The lack of complete protection can be explained by the use of extremely high doses of the ectromelia virus, which is heterologous to VACV.

DISCUSSION

Variolation, i.e. intradermal injection of infectious material from smallpox patients to healthy people, was the first method used to protect people from devastating epidemics of smallpox. The disease induced thus had a short incubation period and was relatively mild compared to conventional human-to-human respiratory transmission of the virus. The mortality caused by the

inoculation was 0.5–2% as opposed to the 20–30% observed during variola virus epidemics [20]. Discovery of the possibility of human vaccination by inoculation with the cowpox virus and later with the vaccinia virus resulted in a significantly lower risk of severe adverse reactions. In the second half of the XXth century, when VACV was used for immunization, mortality was 1–25 per 1 million vaccinated people [21]. In the case of this vaccination, the risk group included primarily people with immunodeficiency, such as transplant patients, HIV-infected patients, individuals taking immunosuppressive drugs, and others. In this regard, modified vaccines with improved safety characteristics were developed based on VACV. For example, late in the XXth century, Russian researchers developed a live vaccine based on the recombinant strain LIVP VACV, which was tested on humans [22].

To date, there has been no mass vaccination against smallpox. However, there are categories of people who are at risk of becoming infected with smallpox or other
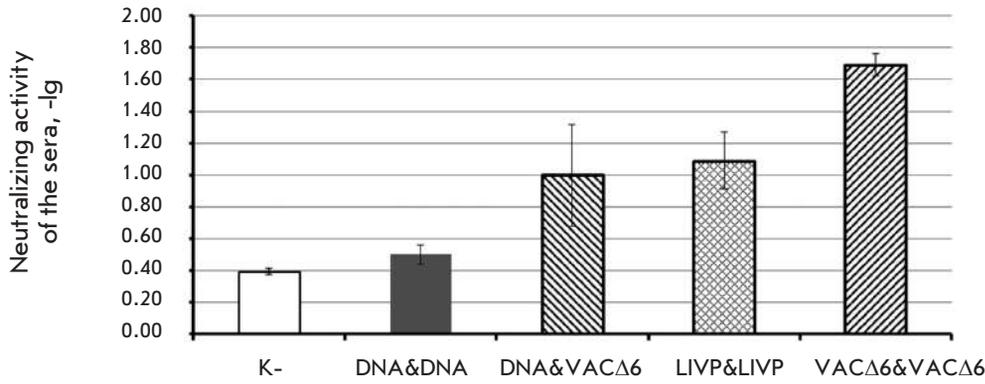
Fig. 3. The level of serum-neutralizing activity against VACV, following double immunization with study preparations (DNA vaccine, VACΔ6 and LIVP VACV strains)

pathogenic orthopoxviruses based on their professional occupation. These categories comprise the risk group, and they should undergo obligatory vaccination against smallpox. First, this concerns personnel involved in epidemiological surveillance, the medical staff of infectious departments at hospitals, and employees of virology laboratories dealing with orthopoxviruses. In the case of smallpox outbreaks (e.g., as a result of a bioterrorist attack), all inhabitants of a region must be vaccinated. The conventional first-generation smallpox vaccine based on the LIVP strain, which is currently used for vaccination, has a lot of contraindications and can cause complications with varying severity. It is worth noting that it is somewhat difficult to demonstrate protective immunity against smallpox induced by vaccination of new preventive medication, since the smallpox has been eliminated, and it is impossible to test the efficacy of these vaccines against the natural disease in the absence of epidemics.

Previously, we implemented two independent approaches to the development of safe vaccines against human orthopoxvirus infections. We developed a highly attenuated variant of the vaccinia virus, VACΔ6, with targeted knockdown of six genes, and a polyvalent DNA vaccine based on five antigens of the variola virus. Independent experiments demonstrated that triple immunization with a DNA vaccine and double immunization with VACΔ6 provide protection to mice against a lethal dose (10 LD_{50}) of the ectromelia virus, which is highly pathogenic to mice [10, 12].

In this study, we compared the immune response developed against orthopoxvirus using various immunization protocols with a DNA vaccine and VACΔ6. The product of the *A35R* gene, one of the six genes deleted in the recombinant variant VACΔ6, reduces the antigen presentation by the class II major histocompatibility complex. Therefore, the VACΔ6 strain induces a higher level of VACV-neutralizing antibodies than the parental clone LIVP, and it is more effective in protecting animals from ECTV infection at a dose of 150 LD_{50}. Combined immunization with a DNA vaccine and a recombinant VACΔ6 variant leads to a lower level of neutralizing antibodies compared to double immunization with VACΔ6. However, it provides the same level of protection. Apparently, this can be attributed to the fact that the DNA vaccine better induces the cell component of the immune response during primary immunization, which is also required for effective orthopoxvirus elimination from the organism [23, 24].
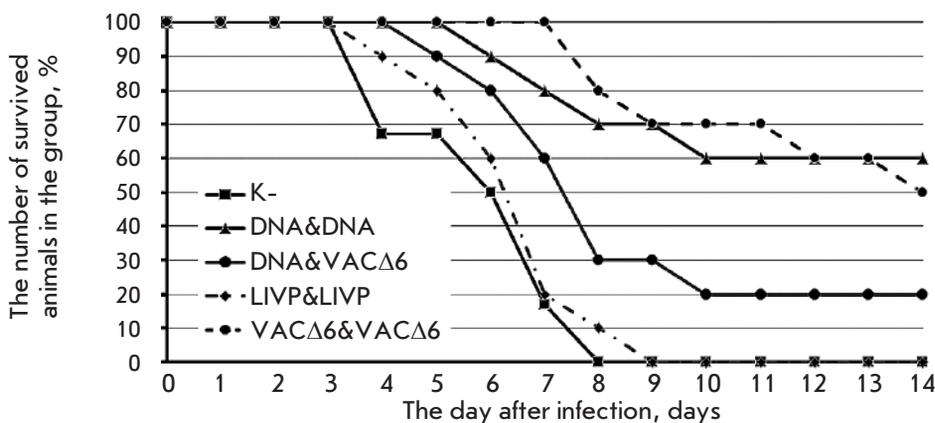
Fig. 4. Time-course of mortality after double immunization of mice with the preparations under study(DNA vaccine, VACΔ6 and LIVP VACV strains), followed by challenge with ECTV at a dose of 150 LD_{50}/mouse

CONCLUSION

In this study, we used a heterologous immunization strategy to enhance the effectiveness of smallpox vaccination, where the immune system was primed using a polyvalent DNA-vaccine based on five VARV genes, and an attenuated version VACΔ6 was used for subsequent booster vaccination. The level of protection induced this way was the same as that in the option with double immunization using a VACΔ6 strain and superior to that induced by double immunization with the LIVP VACV strain used in the Russian Federation for human vaccination. The proposed immunization protocols can be used to develop safe vaccination strategies against smallpox and other human orthopoxvirus infections. DNA vaccination, followed by vaccination with live-attenuated virus VACΔ6 can be considered as advantageous in terms of safety. It should be noted that the double vaccination protocol is not optimal for emergency prevention of smallpox. In this case, single-dose administration of the conventional smallpox vaccine based of the LIVP VACV strain is advisable. ●

REFERENCES
1. Marennikova S.S., Shchelkunov S.N. Pathogenic for humans orthopoxviruses. M: KMK Scientific Press Ltd, 1998. 386 p.
2. Stephenson J. // J. Am. Med. Assoc. 2003. V. 290. P. 23−24.
3. Lewis-Jones S. // Curr. Opin. Infect. Dis. 2004. V. 17. P. 81−89.
4. Campe H., Zimmermann P., Glos K., Bayer M., Bergemann H., Dreweck C., Graf P., Weber B.K., Meyer H., Büttner M., et al. // Emerg. Infect. Dis. 2009. V. 15. P. 777−780.
5. Ninove L., Domart Y., Vervel C., Voinot C., Salez N., Raoult D., Meyer H., Capek I., Zandotti C., Charrel R.N. // Emerg. Infect. Dis. 2009. V. 15. P. 781−784.
6. Shchelkunov S.N. // PLoS Pathogens. 2013. V. 9. e1003756.
7. Shchelkunov S.N. // Vaccine. 2011. V. 29. Suppl. 4. P. D49−D53.
8. Breman J.G., Henderson D.A. // N. Engl. J. Med. 2002. V. 346. P. 1300−1308.
9. Yakubitsky S.N., Kolosova I.V., Maksyutov R.A., Shchelkunov S.N. // Acta Naturae. 2015. V. 7. № 4. P. 113−121.
10. Yakubitsky S.N., Kolosova I.V., Maksyutov R.A., Shchelkunov S.N. // Dokl. Akad. Nauk. 2016. V. 466. P. 241−244.
11. Maksyutov R.A., Shchelkunov S.N. // Russ. J. Immunol. 2010. V. 4. P. 25−32.
12. Maksyutov R.A., Gavrilova E.V., Kochneva G.V., Shchelkunov S.N. // J. of Vaccines. 2013. http://dx.doi.org/10.1155/2013/618324.
13. Maksyutov R.A., Shchelkunov S.N. // Mol. Gen. Mikrobiol. Virusol. 2011. V 2. P. 30−34.
14. Radosevic K., Rodriguez A., Lemckert A., Goudsmit J. // Expert Rev. Vaccines. 2009. V. 8. P. 577−592.
15. Mironova L.L., Hapichev Yu.H. Line heteroploid cell 4647: study, application, prospects. M., 1994. 146 p.
16. Leparc-Goffart I., Poirier B., Garin D., Tissier M.-H., Fuchs F., Crance J.-M. // J. Clin. Virol. 2005. V. 32. P. 47−52.
17. Martinez M.J., Bray M.P., Huggins J.W. // Arch. Pathol. Lab. Med. 2000. V. 124. P. 362−377.
18. Ashmarin I.P., Vorobiev A.A. Statistical methods in microbiological studies. L.: Gos. izd. med. lit., 1962. 186 p.
19. Scarnovich M.O., Radaeva I.F., Vdovichenko G.V., Nechaeva E.A., Sergeev A.A., Petrishenko V.A., Plyasunov I.V., Shishkina L.N., Ternovoi V.A., Smetannikova M.A., et al. // Vopr. virusol. 2007. V. 2. P. 37−40.
20. Fenner F., Henderson D.A., Arita I., Jezek Z., Ladnyi I.D. Smallpox and Its Eradication. Geneva: World Health Organization, 1988. 1460 p.
21. Gurvich E.B. // Vaccine. 1992. V. 10. № 2. P. 96−97.
22. Chernos V.I., Chelyapin N.V., Antonova T.P., et al. // Vopr. virusol. 1990. V. 2. P. 132−135.
23. Kennedy J.S., Frey S.E., Yan L., Rothman A.L., Cruz J., Newman F.K., Orphin L., Belshe R.B., Ennis F.A. // J. Infect. Dis. 2004. V. 190. № 7. P. 1286−1294.
24. Ennis F.A., Cruz J., Demkowicz W.E., Rothman A.L., McClain D.J. // J. Infect. Dis. 2002. V. 185. № 11. P. 1657−1659.

Low-Molecular-Weight NGF Mimetic Corrects the Cognitive Deficit and Depression-like Behavior in Experimental Diabetes

R. U. Ostrovskaya*, S. S. Yagubova, T. A. Gudasheva, S. B. Seredenin
V.V. Zakusov Institute of Pharmacology, Baltijskaya Str., 8, Moscow, 125315, Russia
*E-mail: rita.ostrovskaya@gmail.com
Received September 28, 2016; in final form, February 20, 2017

ABSTRACT Based on the comorbidity of diabetes, depression, and dementia and recognizing that a deficiency of the nerve growth factor (NGF) is involved in all of these kinds of pathologies, we studied the effect of the mimetic of dimeric dipeptide NGF loop 4, GK-2, on a model of streptozotocin-induced type 2 diabetes in C57Bl/6 mice. GK-2 [hexamethylenediamide bis-(N-monosuccinyl-glutamyl-lysine)] was synthesized at the V.V. Zakusov Scientific Research Institute of Pharmacology. The study revealed the ability of GK-2 to ameliorate hyperglycemia induced by streptozotocine (STZ 100 mg/kg i.p.) in C57Bl/6 mice, to restore learning ability in the Morris Water Maze test, and to overcome depression after both intraperitoneal (0.5 mg/kg) and peroral (5 mg/kg) long-term administration. The presence of the listed properties and their preservation in the case of peroral treatment determines the prospects of research. Taking into account the previous findings on the ability of GK-2 to selectively activate PI3K/Akt, these data suggest that Akt-signaling is sufficient for pancreatic beta cell function. GK-2 has been shown to exhibit pronounced neuroprotective activity. The coexistence of neuroprotective and antidiabetic effects is in agreement with the fundamental concept holding that the function of neurons and pancreatic beta cells is controlled by similar mechanisms.

KEYWORDS depression, diabetes, dipeptide NGF mimetic, learning.

ABBREVIATIONS AD − Alzheimer's disease; BDNF − brain-derived neurotrophic factor; i.p. − intraperitoneal; NGF − nerve growth factor; per os − peroral; STZ − streptozotocin; T2D − type 2 diabetes.

INTRODUCTION

In the several decades that have elapsed since it was discovered that neurotrophic factors play a key role in the development and maintenance of the viability of neurons [1], facts showing that they exhibit a similar regulatory activity at the level of non-neuronal systems have been obtained [2]. An understanding of the role of neurotrophins in the development of pancreatic β-cells was one of the important results of these discoveries. The data provide grounds to believe that the similarity between the growth factors and differentiation is responsible for the similarity between pancreatic β-cells and neurons, which form via the same fundamental development program, although they originate from different cell lineages [3]. The regulatory role of neurotrophins in pancreatic β-cells has been confirmed in a number of studies [4, 5]. The effect of the nerve growth factor (NGF) on pancreatic β-cells was found to be mediated by TrkA, the high-affinity neurotrophin receptor [6]. NGF ensures β-cell neogenesis not only during the fetal and neonatal periods, but also in adult organisms [7]. The removal of NGF from a β-cell culture medium [8] and administration of antibodies against this neurotrophic factor [9] enhances β-cell apoptosis. Convincing evidence has been obtained showing that a reduced NGF level in type 2 diabetes mellitus (T2DM) decreases the proliferation of and/or enhances β-cell apoptosis [10−12].

Meanwhile, the comorbidity of T2D and cognitive deficit (reduced information-processing speed, reduced verbal memory and conceptualization), whose risk in T2DM is much higher than in healthy individuals, is well-known. According to epidemiological data, the degree of increase in risk ranges between 50 and 150% [13, 14]. Post-mortem studies have revealed a decreased NGF level in the frontal cortex of patients in the phase that precedes Alzheimer's disease (AD) [15]. A reduced activity of choline acetyltransferase, the enzyme whose activity in cholinergic neurons of the basal brain structures is regulated by NGF, is already in fact in this phase. It has been demonstrated that the level of TrkA receptors in the hippocampus, the brain

structure responsible for the main cognitive functions and memory, in particular, is reduced in patients with mild cognitive impairment [16]. Hippocampal atrophy is an important prognostic sign of an aggravation of the cognitive pathology and a transition from mild cognitive impairment to AD [17]. Deficiency in NGF plays an important role in it, since this neurotrophin prevents the formation of β-amyloid peptide (Aβ1−42) [18]. The decrease in the NGF level accompanying a cognitive deficit is associated with an increased level of its precursor (proNGF) that suppresses the proliferation and differentiation of the basal brain and hippocampal structures [19]. A shift in the proNGF/NGF ratio towards precursor prevalence is regarded as the main reason for cholinergic deficit, leading to cognitive impairment [20].

The risk of depression and depressive-like behavior in T2DM is at least twice as high as that in individuals without resistance to insulin [21]. The bilateral comorbidity of these disorders (depression aggravates the course of diabetes and vice versa) has been studied [22, 23]. In addition to the convincing data on the role of a deficiency in the brain-derived neurotrophic factor (BDNF) in the pathogenesis of depressive states of different etiologies, including in patients with diabetes [24], it has been demonstrated that the activity of NGF drops both in depression and in diabetes, which is considered to be an important factor that determines their comorbidity. A meta-analysis of 21 publications [25] confirmed a statistically significant decrease in the blood level of NGF in depression, which correlated with impairment intensity. It has been suggested that the reduced level of NGF in blood serum should be regarded as a biomarker for major depression [26]. Such a reduction is also observed in patients with bipolar disorder [27] and senile depressions [28]. Post-mortem examinations of brain tissues from suicide victims have revealed an almost two-fold decrease in NGF expression and a more than three-fold decrease in TrkA density [29].

A combination of the reported data demonstrates that NGF could be used in patients with type 2 diabetes mellitus because of its ability to maintain β-cell function, stimulate insulin secretion, and simultaneously impede the development of diabetes mellitus and its comorbidities. However, in their attempts to use native NGF, researchers have faced a problem associated with the unsatisfactory pharmacokinetic properties of this protein molecule (low biological stability and inability to pass through biological barriers when administered systemically) and pleiotropicity of NGF activity, which may result in such side effects as weight loss and hyperalgesia. Meanwhile, the effectiveness of topical administration of NGF in trophic ulcers of diabetic genesis has been reported [30]. As for systemic administra-

tion of NGF, phase I/II clinical trials of recombinant NGF have revealed a tendency towards a favorable effect in patients with diabetic neuropathy; however, side effects and the lack of a therapeutic effect were observed when a broader patient population was used in phase III trials [31].

One of the strategies used to overcome the drawbacks of native neurotrophins involves the design of low-molecular-weight agents that can induce NGF-like therapeutic effects upon systemic administration without the side effects typical of native NGF. Several compounds of this type have been reported; in particular, NGF mimetic of nonpeptide structure, compound MT-2 [32], and peptide NGF-mimetic BB14 [33, 34]. However, the effects of these compounds have been studied only in *in vitro* systems.

A dimeric dipeptide NGF mimetic GK-2 (hexamethylenediamide-*bis*-(N-monosuccinyl-glutamyl-lysine)) has been designed at the V.V. Zakusov Research Institute of Pharmacology on the basis of the structure of the NGF loop 4 β-turn. It exhibited a high neuroprotective activity in *in vitro* experiments, as well as *in vivo* in models of stroke, Alzheimer's and Parkinson's diseases, and had none of the side effects typical of native NGF. GK-2 was shown to activate TrkA receptors [35−37].

Preliminary experiments in rats demonstrated that GK-2 exhibits anti-hyperglycemic activity [38]. On the basis of the comorbidity of diabetes and cognitive impairment and depression, we modeled streptozotocin-induced diabetes in mice and studied the effect of GK-2, the original NGF mimetic, on the cognitive impairment and depressive-like behavior in these animals.

EXPERIMENTAL

Animals
Male C57Bl/6 mice with an initial body weight of 23−28 g purchased from the Stolbovaya breeding farm were used in the experiments. The animals were kept under standard vivarium conditions, with unrestricted access to food (except for 16 h prior to streptozotocin administration) and water. The guidelines for ethical rules in the care and use of animals in research summarized in the European Communities Council Directive 86/609/EEC were followed.

Experiment design
Type 2 diabetes mellitus was induced by intraperitoneal (i.p.) administration of streptozotocin (STZ, Sigma, USA) at a dose of 100 mg/kg, which was effective for C57Bl/6 mice [39].

The mice were randomly divided into four groups: group 1 (passive control, $n = 10$), group 2 (active con-

trol, $n = 11$), and experimental groups 3 ($n = 11$) and 4 ($n = 12$). The mice in the passive control group received saline, either i.p., or perorally (*per os*), for 31 days[1]. The animals from the active control group received saline i.p. for 14 days; a single dose of STZ (100 mg/kg) was administered i.p. on day 15 after 16-hour fasting; then, mice continued to receive saline for 16 days.

The low molecular weight (831 Da) of GK-2 makes it reasonable to study the effects of both i.p. and the peroral route of administration. The effect of GK-2 administered *per os* needs to be studied, since this compound is intended to be used as a drug for long-term clinical application. The freshly prepared GK-2 solution (in 0.9% NaCl) was administered once a day during 14 days: in study group 3, i.p. at a dose of 0.5 mg/kg; in study group 4, *per os* at a dose of 5 mg/kg. On day 15 (30 min after the animals had received the final dose of GK-2), they were i.p. treated with STZ (100 mg/kg) on an empty stomach; then, both groups of mice continued to receive GK-2 for 16 days.

The glucose level in the blood collected from the tail vein was measured using a One Touch Ultra glucometer (USA). The dynamics of the effect of GK-2 was assessed using the indicator of relative antihyperglycemic activity (Ag) according to the formula

$$Ag = gl.STZ - gl.(STZ + GK\text{-}2) \,/\, gl.STZ - gl.saline \times 100\%,$$

where gl.STZ is the blood glucose level in the active control group (group 2); gl.STZ + GK-2 is the blood glucose level in the study group 3 or 4; and gl.saline is the blood glucose level in the passive control group (group 1).

Studying the effect of GK-2 on learning ability in the Morris water maze

Spatial learning and memory were assessed 24 h after the mice had received the final dose of GK-2 (day 17 after administration of STZ) using the Morris water maze [40]. The experimental device consisted of a pool 150 cm in diameter with 60-cm-high walls filled with water (23−25°C). The pool was imaginatively divided into four quadrants. A platform 9 cm in diameter, 1 cm higher than the water level, was placed in the center of one quadrant.

During day 1, the animals were allowed to find the visible platform. If the mouse did not find the platform during the 60 s cut-off, it was placed on the platform and allowed to stay there for 20 s before returning to its home cage. Four trials (one per each quadrant) were used. After 24 h, a platform sub-

merged 1 cm below the water level was placed onto the same spot as in day 1, but water was preliminarily whitened with milk. Identically to day 1, four trials were used, one for each quadrant. The same procedure was repeated on days 3, 4, 5, and 8. The number of animals that found the platform within the 60 s cut-off was recorded.

Studying the effect of GK-2 using the depression model

The depressive-like behavior (the behavioral despair) was assessed using the modified forced swim test on days 45 and 46 after discontinuation of GK-2 [41, 42]. Cylindrically shaped vessels 10 cm in diameter and 30 cm high (OOO Research and Production Company Open Science) were filled with water (23−25°C) to the level of 20 cm from the bottom. On day 1, the animal was placed into the vessel for 10 min and its behavior was video-recorded in the interval between the 2nd and the 6th minute. The test was repeated for 6 min after 24 h. Active swimming and immobilization durations in both sessions were determined using the RealTimer software. According to the definition given by the authors of the test, active swimming implied the periods when the forelimbs moved upward along the cylinder walls, while immobilization implied remaining completely motionless or making the minor movements necessary to maintain the head above water. The total duration of immobilization episodes was the key parameter of the severity of depressive-like behavior in this test.

Exploratory behavior, as well as the general locomotor activity, was assessed using the open field test 2 days prior to performing the Morris water maze. The animals were placed in the center of the open field, and the horizontal motor activity and the numbers of holes and vertical bars were measured during 5 min.

The animals' body weight was measured every 3 days.

Figure 1A shows the order in which the compounds were administered and behavioral tests were performed.

Statistical analysis

The experimental data are shown as mean values, with the mean error and the standard error of the mean (M ± SEM) indicated. The statistical analysis was performed using the Statistica 8.0 software. The statistical significance of intergroup differences was assessed using the nonparametric method, the Mann−Whitney U test. The χ^2 test was used for the parameters measured in %. The results were considered to be statistically significant at $p \leq 0.05$.

---

[1] *No significant differences between i.p. or per os administration of saline during the entire experiment were revealed, so these animals were merged into one group.*
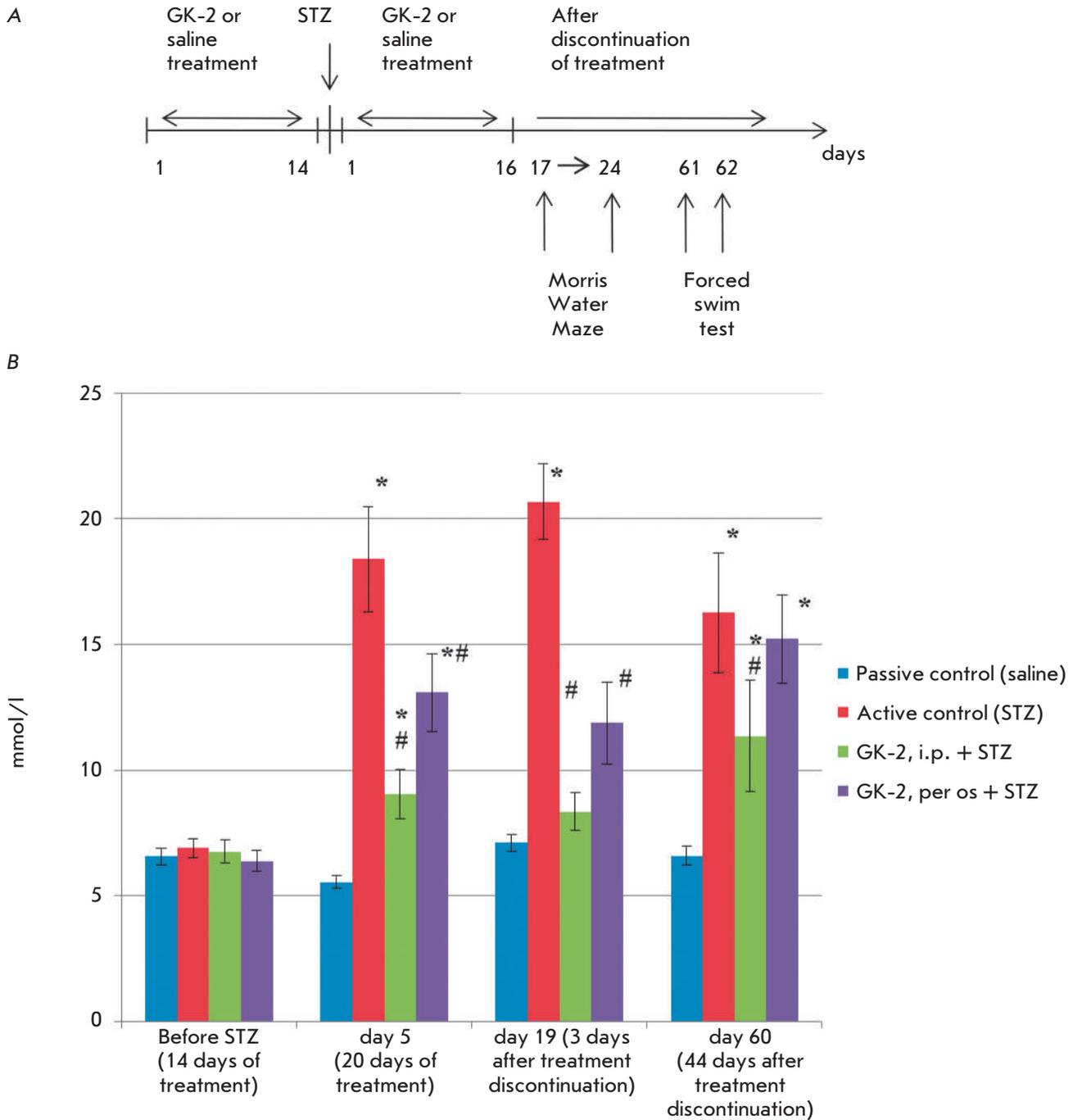
Fig. 1. Design of the experiment (*A*) and the dynamics of the blood glucose level (mmol/l) in C57Bl/6 mice (*B*) in the following groups: passive control (Saline + Saline), active control (Saline + STZ 100 mg/kg, i.p. + Saline), GK-2 treated group 3 (GK-2 0.5 mg/kg, i.p. + STZ 100 mg/kg, i.p. + GK-2 0.5 mg/kg, i.p.), GK-2 treated group 4 (GK-2 5 mg/kg, per os + STZ 100 mg/kg, i.p. + GK-2 5 mg/kg, per os). Data are presented as M ± SEM. The statistical significance of the differences was calculated using the Mann–Whitney U-test: * p<0.05 compared to passive control; # p<0.05 compared to active control (STZ).

RESULTS

Data on the dynamics of the blood glucose level in different groups are presented in *Fig. 1B*. While the glucose level in the peripheral blood of mice in the passive control group was 6−7 mmol/l, administration of STZ at a dose of 100 mg/kg to C57Bl/6 mice increased that blood glucose level to 16−20 mmol/l, which is close to the values obtained earlier in the experiments with rats [38]. In full compliance with the antihyperglycemic effect of GK-2 observed in the experiments with rats, we revealed the antihyperglycemic effect of GK-2 in mice. It is important to emphasize that the antihyperglycemic effects were similar for rats and mice: e.g., the calculated Ag parameter on day 17 after administration of STZ to rats was 80%, being 90% on day 19 in mice.

Assessment of the cognitive function performed 24 h after the final dose of GK-2 had been injected demonstrated (*table*) that, whereas the number of animals that found the platform within 60 s in repeated tests significantly increased in the passive control group, this occurred muchmore slowly in the active control group (the differences between the two groups were statistically significant on days 4 and 8). These results agree with the data on cognitive impairment in STZ-induced diabetes [43]. Intraperitoneal administration of GK-2 caused a statistically significant increase in the number of animals that found the platform on days 2, 4, and 8 of training compared to the animals in the active control group. Upon administration *per os*, the learning ability significantly increased only on test day 2. It should be mentioned that in the beginning of the experiment, the learning ability of mice for both administration routes was even higher than that in the passive control group. The intergroup differences were significant on test days 3 and 5 as well (except for day 5 in the group that received GK-2 *per os*, when the differences between the active control and the study group failed to reach the level of statistical significance).

The effect of GK-2 on the severity of the depression-like behavior was assessed in a long-term period after STZ administration (day 45), since the duration of the depressive-like behavior in the diabetes model was reported to be rather long [25].

Comparison of active swim test parameters and the immobilization duration in different groups revealed the following regularities (*Fig. 2*): In mice in the active control group, immobilization duration increased, while the duration of active swimming decreased compared to the parameters in the passive control group, while i.p. administration of GK-2 reduced the immobilization duration and increased the active swimming duration, making them as high as the control values. The intensity of the effect of GK-2 administered *per os* was the same as upon i.p. administration.

Learning ability of mice in the Morris water maze (the percentage of animals that found the platform within the 60 s cut-off time)

Group	day 2	day 4	day 8
Group 1 Passive control (saline)	14.3%	85.7%	100%
Group 2 Active control (STZ, 100 mg/kg)	9.09%	54.54%*	72.7%*
Group 3 GK-2, 0.5 mg/kg i.p. + STZ	27.3%*#	72.7%*#	90.9%*#
Group 4 GK-2, 5 mg/kg *per os* + STZ	50%*#	50%*	100%#

The statistical significance of differences was assessed using the χ^2 test.
*$p < 0.05$ compared to passive control group (saline).
#$p < 0.05$ compared to active control group (STZ).

Similar regularities were observed on day 2: increased immobilization duration and reduced active swimming duration in the active control group, where GK-2 reduced the severity of depression when administered both i.p. and *per os*.

In order to interpret the results, we needed to understand whether the streptozotocin-induced behavioral disorders were related to the overall wellbeing of the animals (reduced motor activity and body weight loss). In order to answer this question, we performed the open field test 2 days prior to the Morris water maze, where changes in neither the orientational nor exploratory activity and overall mobility were observed in the animals treated with STZ. GK-2 upon both administration routes had no effect on these indicators. It was demonstrated that, unlike the passive control group where animal body weight increased during the entire experiment (10.5% with respect to the initial weight by the time the Morris water maze was performed and 16.7% by the time the forced swim test was formed), a slight decrease in body weight by the time of Morris water maze study (−6.7%) and body weight gain by the time of the forced swim test (1.8%) were observed in the active control group. GK-2 reduced this effect of STZ administered both i.p. (−2 and 4.6%, respectively) and *per os* (1 and 10%, respectively). Therefore, the resulting data allow one to rule out the changes in the overall wellbeing of animals as the reason for the STZ-induced behavioral disorders and their normalization due to the administration of NGF mimetic.
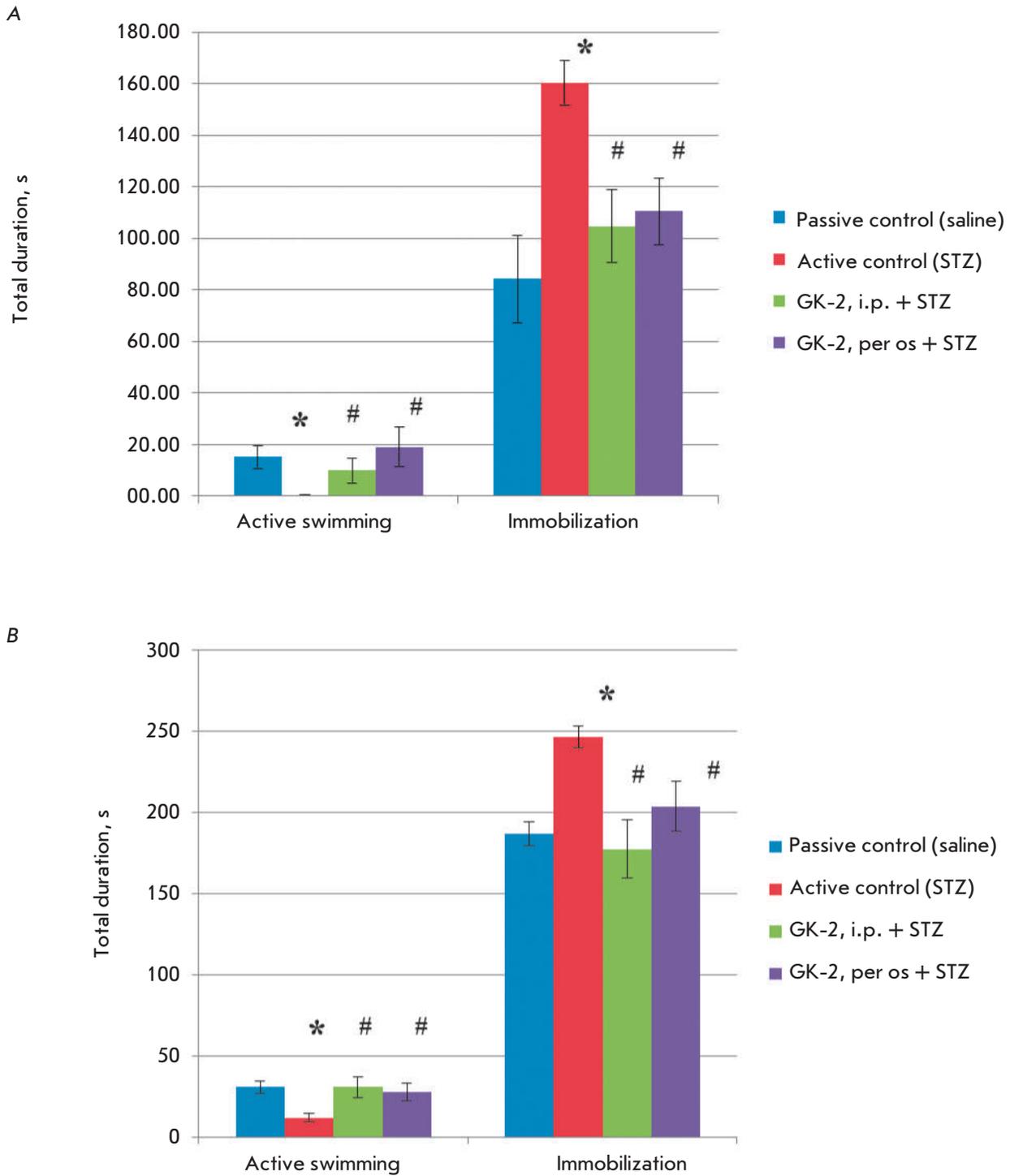
Fig. 2. Indicators of a depressive-like status in C57Bl/6 mice: the total duration of active swimming and immobilization (s) on days 61 (A) and 62 (B) after STZ administration. The groups and statistical data are similar to those in *Fig. 1B*

DISCUSSION

We reproduced the known model of diabetes mellitus with specific behavioral signs [25, 43] and described for the first time the ability of GK-2, the low-molecular-weight mimetic of the nerve growth factor, to eliminate these behavioral disorders. The main role in the development of a deficiency of NGF in diabetes is known to be its reduced formation from the proNGF precursor as a result of hyperglycemia-induced oxidative stress [44, 45], which suppresses protease activity and shifts the proNGF/NGF ratio towards the precursor prevalence that promotes apoptosis of insulin-secreting cells, contrary to mature NGF, exerting an antiapoptotic effect (*Fig. 3*).

Streptozotocin facilitates free radical formation and alkylates DNA [46]. Administration of STZ reproduces not only the reduced NGF level typical of diabetes [47], but also the increased proNGF level [48]. It has been experimentally demonstrated that the degrees to which the proNGF level increases and mature NGF and phosphorylated TrkA receptors decrease correlate with the severity of the cognitive impairment [49]. The shift in the proNGF/NGF ratio towards the precursor is considered to be the main reason behind the cholinergic deficit that causes cognitive impairment [20].

Identically to the native NGF molecule, GK-2 activates TrkA receptors and alleviates the toxic effects of H_2O_2 [35]. In addition, it reduces the blood level of malonic dialdehyde in diabetic mice [50]. An assumption can be derived from these data that the antihyperglycemic effect of GK-2 is caused both by its direct effect on NGF receptors and by its ability to eliminate the toxic effect of free radicals, which can normalize the formation of NGF from its precursor.

We experimentally reproduced the main metabolic effect of STZ − the hyperglycemic effect − and also its behavioral effects imitating the behavioral disorders in diabetic patients: namely, cognitive impairment [14, 16] and development of a depressive-like behavior [51−53]. The ability of GK-2 to attenuate the severity of the cognitive deficit accompanying a diabetes model was revealed. This fact agrees with the positive cognitive effect of GK-2 observed in the Alzheimer's disease model [54]. The antidepressant effect of GK-2 was described for the first time. The combination of the antidiabetic and antidepressant activities of GK-2 is especially important, because conventional antidepressants not only do not attenuate diabetes signs, but can also increase the risk of its development [55].

It is important to emphasize that the activity of GK-2 is maintained in the case of peroral administration, which is a requisite for drugs used to treat chronic conditions. The combination of the antidiabetic activity of GK-2 with its long-term positive effect on the
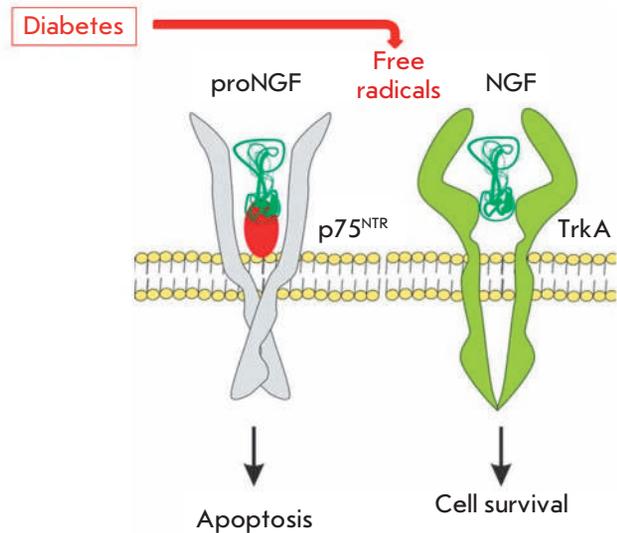
Fig. 3. NGF is synthesized from the precursor, pro-NGF. NGF binds to the TrkA receptor and this interaction induces the activation of the signaling pathway of β-cell survival. Diabetes-induced hyperglycemia is known to cause the oxidative stress that decreases protease activity, thus provoking pro-NGF accumulation resulting in β-cell apoptosis (modified from [19, 48])

cognitive function and antidepressant properties is an important additional characteristic of this compound. GK-2 is intended for use in the therapy of post-stroke sequelae, since it is known that stroke and diabetes are comorbid and that there is a high rate of development of cognitive deficit and depressive disorders during the post-stroke period [56].

It has been demonstrated previously [57] that the NGF mimetic GK-2 selectively activates only one of the two main signaling pathways, the PI3K/Akt pathway involved in the neuroprotective effects of neurotrophins, by activating TrkA [58]. The data on the antidiabetic activity of GK-2 allow one to suggest that Akt signalization is sufficient to maintain the β-cell function. The significance of these data mainly consists in the fact that they can lead to new concepts of diabetes development mechanisms and could serve as a basis for the design of antidiabetic agents that exhibit cytoprotection of β-cells. The combination of the neuroprotective and antidiabetic effects of GK-2 is consistent with the earlier stated fundamental concept that the mechanisms of regulation of the function of neurons and pancreatic β-cells are similar [59] and the subsequent conclusion about the reasonability of studying the potential antidiabetic properties of neuroprotective agents that eliminate the deficiency of neurotrophic factors [60].

CONCLUSIONS

The hyperglycemic, amnestic, and depressive-like effects of STZ were reproduced in this study. The ability of GK-2, the dimeric analog of nerve growth factor loop 4, to have an antihyperglycemic effect and attenuate the severity of the cognitive deficit that develops in a diabetes model has been revealed. The anti-depressant activity of the compound has been established for the first time. Further development of GK-2 is promising due to its combination of antidiabetic activity and positive effect on cognitive functions, as well as antidepressant properties and maintenance of activity when administered *per os*.

In view of the data on the pronounced neuroprotective activity of GK-2 previously obtained at the Research Institute of Pharmacology, the antidiabetic activity of this compound can be regarded as an important argument in support of the fundamental concept that the function of neurons and pancreatic β-cells is controlled by similar mechanisms. ●

REFERENCES

1. Levi-Montalcini R. // Science. 1987. V. 237. P. 1154–1162.
2. Yamamoto M., Sobue G., Yamamoto K., Terao S., Mitsuma T. // Neurochem. Res. 1996. V. 21. № 8. P. 929–938.
3. Polak M., Scharfmann R., Seilheimer B., Eisenbarth G., Dressler D., Verma I.M., Potter H. // Proc. Natl. Acad. Sci. USA. 1993. V. 90. № 12. P. 5781–5785.
4. Rosenbaum T., Vidaltamayo R., Sanchez-Soto M.C., Zentella A., Hiriart M. // Proc. Natl. Acad. Sci. USA. 1998. V. 95. P. 7784–7788.
5. Gezginci-Oktayoglu S., Karatug A., Bolkent S. // Diabetes Metab. Res. Rev. 2012. V. 28. № 8. P. 654–662.
6. Kanaka-Gantenbein C., Dicou E., Czernichow P., Scharfmann R. // Endocrinology. 1995. V. 136. № 7. P. 3154–3162.
7. Paris M., Tourrel-Cuzin C., Plachot C., Ktorza A. // Exp. Diabesity Res. 2004. V. 5. № 2. P. 111–121.
8. Pierucci D., Cicconi S., Bonini P., Ferrelli F., Pastore D., Matteucci C., Marselli L., Marchetti P., Ris F., Halban P., et al. // Diabetologia. 2001. V. 44. № 10. P. 1281–1295.
9. Gezginci-Oktayoglu S., Karatug A., Bolkent S. // Pancreas. 2015. V. 44. № 2. P. 243–249.
10. Faradji V., Sotelo J. // Acta Neurol. Scand. 1990. V. 81. P. 402–406.
11. Vidaltamayo R., Mery C.M., Angeles-Angeles A., Robles-Díaz G., Hiriart M. // Growth Factors. 2003. V. 21. № 3–4. P. 103–107.
12. Chaldakov G.N. // Archives Italiennes de Biologie. 2011. V. 149. № 2. P. 257–263.
13. Biessels G.J., Staekenborg S., Brunner E., Brayne C., Scheltens P. // Lancet Neurol. 2006. V. 5. № 1. P. 64–74.
14. Li X., Song D., Leng S.X. // Clin. Intervent. Aging. 2015. V. 10. P. 549–560.
15. Hellweg R., Gericke C.A., Jendroska K., Hartung H.D., Cervós-Navarro J. // Int. J. Dev. Neurosci. 1998. V. 16. № 7–8. P. 787–794.
16. Mufson E.J., He B., Nadeem M., Perez S.E., Counts S.E., Leurgans S., Fritz J., Lah J., Ginsberg S.D., Wuu J., et al. // J. Neuropathol. Exp. Neurol. 2012. V. 71. № 11. P. 1018–1029.
17. Devanand D.P., Pradhaban G., Liu X., Khandji A., De Santi S., Segal S., Rusinek H., Pelton G.H., Honig L.S., Mayeux R., et al. // Neurology. 2007. V. 68. № 11. P. 828–836.
18. Triaca V., Sposato V., Bolasco G., Ciotti M.T., Pelicci P., Bruni A.C., Cupidi C., Maletta R., Feligioni M., Nisticò R., et al. // Aging Cell. 2016. V. 15. № 4. P. 661–672.
19. Budni J., Bellettini-Santos T., Mina F., Garcez M.L., Zugno A.I. // Aging Dis. 2016. V. 6. № 5. P. 331–341.
20. Allard S., Leon W.C., Pakavathkumar P., Bruno M.A., Ribeiro-da-Silva A., Cuello A.C. // J. Neurosci. 2012. V. 32. № 6. P. 2002–2012.
21. Kahn L.S., McIntyre R.S., Rafalson L., Berdine D.E., Fox C.H. // Depression Research and Treatment. 2011. e862708.
22. Pouwer F., Nefs G., Nouwen A. // Endocrinol. Metab. Clin. North Am. 2013. V. 42. № 3. P. 529–544.
23. Holt R.I., de Groot M., Lucki I., Hunter C.M., Sartorius N., Golden S.H. // Diabetes Care. 2014. V. 37. № 8. P. 2067–2077.
24. Wang J., Zhao X., He M. // Med. Hypotheses. 2012. V. 79. № 2. P. 255–258.
25. Chen Y.W., Lin P.Y., Tu K.Y., Cheng Y.S., Wu C.K., Tseng P.T. // Neuropsychiatric Disease and Treatment. 2015. V. 11. P. 925–933.
26. Wiener C.D., de Mello Ferreira S., Pedrotti Moreira F., Bittencourt G., de Oliveira J.F., Lopez Molina M., Jansen K., de Mattos Souza L.D., Rizzato Lara D., Portela L.V., et al. // J. Affect. Disord. 2015. V. 184. P. 245–248.
27. Barbosa I.G., Huguet R.B., Neves F.S., Reis H.J., Bauer M.E., Janka Z., Palotás A., Teixeira A.L. // World J. Biol. Psychiatry. 2011. V. 12. № 3. P. 228–232.
28. Diniz B.S., Teixeira A.L., Machado-Vieira R., Talib L.L., Gattaz W.F., Forlenza O.V. // Am. J. Geriatr. Psychiatry. 2013. V. 21. № 5. P. 493–496.
29. Banerjee R., Ghosh A.K., Ghosh B., Bhattacharyya S., Mondal A.C. // Clin. Med. Insights Pathol. 2013. V. 6. P. 1–11.
30. Tiaka E.K., Papanas N., Manolakis A.C., Maltezos E. // Int. J. Burns Trauma. 2011. V. 1. № 1. P. 68–76.
31. Pittenger G., Vinik A. // Exp. Diabesity Res. 2003. V. 4. № 4. P. 271–285.
32. Scarpi D., Cirelli D., Matrone C., Castronovo G., Rosini P., Occhiato E.G., Romano F., Bartali L., Clemente A.M., Bottegoni G., et al. // Cell Death Dis. 2012. V. 3. e339.
33. Cirillo G., Colangelo A.M., Bianco M.R., Cavaliere C., Zaccaro L., Sarmientos P., Alberghina L., Papa M. // Biotechnol. Adv. 2012. V. 30. № 1. P. 223–232.
34. Cirillo G., Colangelo A.M., De Luca C., Savarese L., Barillari M.R., Alberghina L., Papa M. // PLoS One. 2016. V. 11. № 3. e0152750.
35. Gudasheva T.A., Antipiva T.A., Seredenin S.B. // Doklady Biochemistry and Biophysics. 2010. V. 434. № 4. P. 262–265.
36. Seredenin S.B., Gudasheva T.A. // Zhurnal nevrologii i psikhiatrii. 2015. V. 6. P. 63–70.
37. Gudasheva T.A., Povarnina P.Y., Antipova T.A., Firsova Y.N., Konstantinopolsky M.A., Seredenin S.B. // J. Biomed. Sci. 2015. V. 22. e106.

38. Seredenin S.B., Gudasheva T.A., Ostrovskaya R.U., Povarnina P.Y., Ozerova I.V. // Patent RU № 2613314. 2017. A61K38/05, A61P3/10.

39. Hayashi K., Kojima R., Ito M. // Biol. Pharm. Bull. 2006. V. 29. № 6. P. 1110−1119.

40. Morris R. // J. Neurosci. Meth. 1984. V. 11. № 1. P. 47−60.

41. Porsolt R.D., Bertin A., Jalfre M. // Eur. J. Pharmacol. 1978. V. 51. P. 291−294.

42. Perona M.T., Waters S., Hall F.S., Sora I., Lesch K.P., Murphy D.L., Caron M., Uhl G.R. // Behav. Pharmacol. 2008. V. 19. № 5−6. P. 566−574.

43. Du G.T., Hu M., Mei Z.L., Wang C., Liu G.J., Hu M., Long Y., Miao M.X., Chang Li J., Hong H. // J. Pharmacol. Sci. 2014. V. 124. № 4. P. 418−426.

44. Vincent A., Brownlee M., Russell J. // Ann. N.Y. Acad. Sci. 2002. V. 959. P. 368−383.

45. Ali T.K., Matragoon S., Pillai B.A., Liou G.I., El-Remessy A.B. // Diabetes. 2008. V. 57. № 4. P. 889−898.

46. Szkudelski T. // Physiol. Res. 2001. V. 50. № 6. P. 536−546.

47. Sposato V., Manni L., Chaldakov G.N., Aloe L. // Arch. Italiennes Biol. 2007. V. 145. P. 87−97.

48. Al-Gayyar M.M., Mysona B.A., Matragoon S., Abdelsaid M.A., El-Azab M.F., Shanab A.Y., Ha Y., Smith S.B., Bollinger K.E., El-Remessy A.B. // PLoS One. 2013. V. 8. № 1. e54692.

49. Terry A.V.Jr., Kutiyanawalla A., Pillai A. // Physiol. Behav. 2011. V. 102. № 2. P. 149−157.

50. Yagubova S., Zolotov N., Ostrovskaya R. // J. Diabetes Metab. 2016. V. 7. Suppl 7. P. 76.

51. Kamei J., Miyata S., Morita K., Saitoh A., Takeda H. // Pharmacol. Biochem. Behav. 2003. V. 75. № 2. P. 247−254.

52. Ates M., Dayi A., Kiray M., Sisman A.R., Agilkaya S., Aksu I., Baykara B., Buyuk E., Cetinkaya C., Cingoz S., et al. // Biotech. Histochem. 2014. V. 89. № 3. P. 161−171.

53. Luchsinger J.A. // J. Alzheimers Dis. 2012. V. 30. P. 185−198.

54. Povarnina P.Y., Vorontsova O.N., Gudasheva T.A., Ostrovskaya R.U., Seredenin S.B. // Acta Naturae. 2013. V. 5. № 3. C. 84−91.

55. Barnard K., Peveler R.C., Holt R.I.G. // Diabetes Care. 2013. V. 36. № 10. P. 3337−3345.

56. Li W.A., Moore-Langston S., Chakraborty T., Rafols J.A., Conti A.C., Ding Y. // Neurol. Res. 2013. V. 35. № 5. P. 479−491.

57. Gudasheva T.A., Antipova T.A., Konstantinopolsky M.A., Povarnina P.Y., Seredenin S.B. // Dokl Biochem Biophys. 2014. V. 456. № 1. P. 88−91.

58. Kaplan D.R., Miller F.D. // Curr. Opin. Neurobiol. 2000. V. 10. № 3. P. 381−391.

59. de la Monte S.M. // Front. Biosci (Elite Ed.). 2012. V. 4. P. 1582−1605.

60. Ostrovskaya R.U., Yagubova S.S. // Psychiatry. 2014. №1 (61). P. 35−43.

GENERAL RULES

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

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

The journal *Acta Naturae* is on the list of the leading periodicals of the Higher Attestation Commission of the Russian Ministry of Education and Science. The 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 60,000 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 (8 pages in A4 format including no more than 12 references). It should contain no more than four figures.

The manuscript and the accompanying documents should be sent to the Editorial Board in electronic form:
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. Ab-

stract. The structure of the abstract should be very clear and must reflect the following: it should introduce the reader to the main issue and describe the experimental approach, the possibility of practical use, and the possibility of further research in the field. The average length of an abstract is 20 lines (1,500 characters).
- Keywords (3 − 6). These should include the field of research, methods, experimental subject, and the specifics of the work. List of abbreviations.
- INTRODUCTION
- EXPERIMENTAL PROCEDURES
- RESULTS AND DISCUSSION
- CONCLUSION
  The organizations that funded the work should be listed at the end of this section with grant numbers in parenthesis.
- REFERENCES
  The in-text references should be in brackets, such as [1].

RECOMMENDATIONS ON THE TYPING AND FORMATTING OF THE TEXT

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

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

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

REVIEW AND PREPARATION OF THE MANUSCRIPT FOR PRINT AND PUBLICATION
Articles are published on a first-come, first-served basis. The 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.