Multifunctional Proteins

A. B. Uzdensky*

Laboratory of Molecular Neurobiology, Academy of Biology and Biotechnology, Southern Federal University, Rostov-on-Don, 344090 Russia

*e-mail: auzd@yandex.ru

Abstract—A surprisingly small number of human genes (19–20 thousand) is not consistent with a much larger number of proteins and the number of their functions. One of the factors of functional diversity is the multifunctionality of proteins. An important subclass of such proteins is moonlighting proteins, in which one polypeptide chain performs two or more functions under different conditions. Often, these are various housekeeping proteins – glycolytic, ribosomal, and others, which are abundant in the cell. In this review we consider the structures and functions of several such proteins – the rpS3 ribosomal protein, cytochrome c, glyceraldehyde-6-phosphate dehydrogenase (GAPDH), transcription factors STAT3, β-catenin and p53. A switching of their functions occurs due to violation of the balance between their synthesis, use, and degradation, intracellular relocalization, and post-translational modifications. A significant role of the internal disordered regions in the formation of intermolecular complexes with other proteins and nucleic acids is noted. The emergence of multifunctional proteins during evolution is discussed.

INTRODUCTION

Molecular biology of the twentieth century was based on the classical postulates, such as *"DNA contains the information on the structure of all proteins and RNA in the organism." Data on all proteins and RNA of the body is found in DNA*. One more postulate was formulated by Beadle and Tatum (1941): "*One gene – one enzyme*" [1], which was later transformed into: "*One gene—one protein*," and after refinement— "*One cistron—one polypeptide chain*." Another important postulate was formulated by Anfinsen [2]: "*The primary protein structure, i.e. the amino acid sequence, which is unambiguously determined by the nucleotide sequence in DNA, unambiguously determines its spatial structure and functional activity.*"

However, the last data showed that these statements are not accurately consistent with the current state of molecular biology. It was early assumed the existence of about a million genes in a human genome. Then their number was reduced to 40–100 thousand.

But in the early 21st century the genome sequencing of humans and other organisms have led to the striking discovery: a human genome has only about 20 thousand genes encoding proteins [3]. This value is of the same order as that of the primitive nematode *Caenorhabditis elegans*. Where does the complexity of the organism come from? What, if not the genes, is the reason why a human differs from a worm? Probably, the matter is not so much the genes as the proteins. Indeed, the number of proteins is significantly greater than the number of genes. There is no simple correspondence. The exact amount of protein is unknown. In the UniProtKB/Swiss-Prot database, about 500 thousand protein amino acid sequences were recorded in 2014 [4]. The Protein Data Bank presents more than 150 thousand three-dimensional protein structures [5]. The number of structural proteoforms, i.e. local variations in the primary amino acid sequence, which are determined in tissues by mass spectrometry, liquid chromatography, two-dimensional or capillary electrophoresis, and other methods, amounts to several million [6, 7]. And the number of identified polymorphisms in the human genome exceeds 150 million [7]. Nevertheless, there is no doubt that information on all cellular proteins is contained in the genome. But why is the number of proteins many times greater than the number of genes?

The main sources of the diversity of proteoforms are: alternative splicing of mRNA, single nucleotide

Abbreviations: P—platform; PTM—posttranslational modification; E—enzyme; Apaf-1—Apoptotic protease-activating factor 1; CTD—C-terminal domain; DBD—DNA-binding domain; GAPDH—glyceraldehyde-6-phosphate dehydrogenase; IDR—intrinsically disordered region; MLP—moonlighting proteins; NLS—nuclear localization sequence; p53RE—p53-response element; SLiM—short linear motif; SNP—single nucleotide polymorphism; STAT3—signal transducer and activator of transcription 3; TAD—transactivation domain; TET—tetramerization domain.

Table 1. Moonlighting protein databases

MLP databases	Number of MLP proteins (December 2018)	URL
MultitaskProtDB-II (Multifunctional proteins including MLP)	694	http://wallace.uab.es/multi taskII
MoonProt	361	http://moonlightingproteins.org/
MoonDB v2.0	351	http://moondb.hb.univ-amu.fr/

polymorphism (SNP), and covalent post-translational modifications (PTM). In the case of PTM, small molecular groups − methyl, acetyl, phosphate, etc., or long polymer chains − lipid, carbohydrate (glycosylation), protein (ubiquitination), nucleotide (poly-ADP-ribosylation) add to proteins. Small inhibitory peptides can also be cleaved from the protein molecules. Nevertheless, despite the huge variety of proteoforms, the number of canonical proteins that are more common than other alternative forms, and that are similar to orthologic forms in other organisms [7], significantly exceeds the number of genes. Each cell has its own set of proteins, the composition and quantity of which (protein profile), is dynamic and is determined by the complex network of signaling cascades that respond to external impacts, which change metabolism and homeostasis.

Protein diversity can be conditionally divided into a variety of structures and a variety of functions. Different protein structures, for example, different splice forms, can, although with different efficiencies, perform the same function. On the other hand, one protein can perform different functions under different conditions. However, the presence of a large number of structural proteoforms does not mean the equal number of performed functions. Many proteoforms are either inactive, or are eliminated in the cell. Amino acid substitutions may be indifferent and not influence protein activity, although in many cases they can impair functions and lead to cell death [7].

The important factor of protein diversity and origin of new proteins is alternative splicing, in which RNA fragments, introns, are removed from pre-mRNA, whereas remaining fragments, exons, cross-link. Mature mRNA is then formed after an additional processing. Each gene contains 4−5 exons in average; sometimes more than 10. Their combinations can, in principle, produce new proteins [8]. The numbers of spliceforms (RNA transcripts) in human cells were estimated to be around 80−200 thousand, i.e. from 4 to 10 for each gene [9]. However, the studies of expression of diverse genes in cells showed that only one protein isoform is actually produced. It corresponds optimally to the biological purpose, is more conservative and does not have disturbances in functional domains. Alternative splicing is tissue and species-specific. Alternative spliceforms may predominate in various tissues or organisms, but their number is usually lim-

ited. Although many pre-mRNAs can undergo alternative splicing, most spliceformss in cells are not really translated and not produced [9]. For example, the amyloid precursor protein gene (APP) contains 18 exons. However, only three alternative isoforms are produced in cells: APP695, APP751 and APP770. APP695 predominates in the brain [10]. Apparently, alternative splicing plays a limited role in complicating of the cellular proteome.

Multifunctional Proteins

The multifunctionality of proteins is provided by their structure, domain organization. Switching of protein functions can occur upon binding of allosteric regulators, post-translational modifications, changes in intracellular localization and microenvironment, under external influences, etc. This significantly increases the functional predetermination given by genes. According to December 2018 data, the MultitaskProtDBII database collected the infirmation on about 700 multifunctional proteins (Table 1) [11, 12].

Schematically, proteins can be divided into two categories:

(1) "Enzymes" (E) that carry out biochemical or biophysical processes with transport of charged particles (electrons, protons, ions) or molecules, rupture or formation of chemical bonds as a result of rearrangement of electron clouds in the active center.

(2) "Platforms" (P), in which the surface regions recognize the complementary surfaces of other molecules during the assembly of supramolecular complexes. The assembly includes the search and approach of interacting molecules due to long-range forces (electrostatic, van der Waals); docking; binding of complementary surfaces mediated by electrostatic, van der Waals, hydrogen and hydrophobic interactions; conformational adjustment, and final activation of the complex. The self-assembly of protein-protein complexes of nucleic acid-protein complexes (ribosomes, nucleosomes, transcriptional and repair complexes, etc.) is based on the recognition and interaction of "platforms." Depending on the strength of the interaction, such complexes can be stable or short-lived.

From a physiological point of view, many signaling proteins such as protein kinases, phosphatases, acetyltransferases or deacetylases are multifunctional. Carrying out post-translational modifications (phosphor-
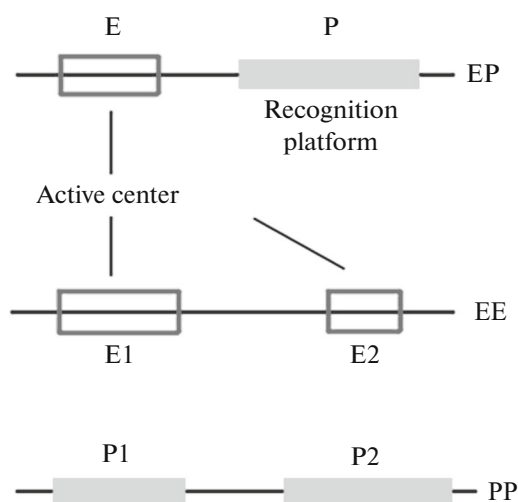
Fig. 1. Combinations of the domains of the enzyme type (E) and/or the recognition platform type (P) in MLP proteins: (1) Region A—enzyme; Region B—platform (EP); (2) Region A—enzyme; Region B—enzyme B (EE); (3) Two platforms A and B (PP).

ylation, methylation, acetylation, etc.), they activate or inhibit different proteins, turn on, turn off or switch their functions and thereby regulate a variety of cellular processes. A higher level of regulation is performed by transcription factors that regulate the expression of many functionally related proteins. They carry out complex cell reactions: changes in functional activity, metabolic regulation, division, apoptosis, etc. The physiological functions of such regulatory proteins are mediated by the set of proteins, which are modified by them.

We consider the proteins whose multiple functions are not mediated by other proteins. They transform substrates directly or recognize other molecules in the course of assembly of supramolecular complexes. Their functions switch upon changes of conditions.

Moonlighting Proteins (MLP)

In recent years, moonlighting proteins (MLP) attract the increasing attention. This term means a moonlighting, or second work. MLP is a subgroup of multifunctional proteins in which one polypeptide chain encoded by one gene can form different spatial structures and, under different conditions, perform two or more biochemical and biophysical functions. Each function is regulated independently of the others [13–18]. From the functional point of view, MLP is a combination of the enzyme-type domains (E) and/or the recognition platforms (P) (Fig. 1).

Several hundred MLPs have already been identified at the all levels of the evolutionary tree from bacteria to higher organisms. Many of them are very conservative [11, 12]. MLP perform diverse functions. They are often the abundant housekeeping proteins, which are expressed in large quantities in cells. Some components of glycolysis, Krebs cycle, chaperones, ribosome proteins, transcription factors, cytoskeleton components, cell surface receptors, cell adhesion proteins, etc. are MLP [13–17]. For example, 7 out of 10 proteins of glycolysis, 7 out of 8 the Krebs cycle proteins, or various ribosomal proteins (rpS3, rpL10A, rpL13a, etc.) are moonlighting proteins. The information on more than 350 MLPs is presented in the databases MultitaskProtDB-II, MoonProt and MoonDB v2.0 (Table 1) [11, 12]. Examples of different MLPs are shown in Table 2. We consider here some of them.

Ribosomal Protein rpS3

More than a dozen ribosomal proteins participate in various non-ribosomal protein complexes in the cell nucleus. They are involved in DNA repair, regulation of transcription and other functions. For example, the rpS3 protein, a component of the 40S ribosome subunit, performs a variety of additional functions [19–21]. After production in the cytoplasm, it is transferred to the nucleus, where it used for the ribosome synthesis. However, some its amount activates the transcription factor NF-κB. It enhances the affinity of p65, the NF-κB subunit, to gene promoters. After p65 binding to the rpS3 domain KH (Fig. 2), NF-κB binds to promoters of genes, which it regulates. The C-terminal domain of rpS3 is also involved in DNA repair (Fig. 2) [21]. The different functions of rpS3 are mediated by phosphorylation of different amino acids and binding of various proteins to its domains. So, after phosphorylation of serine 209 and threonines 42 and 221 by protein kinases IKKβ, ERK and PKCδ, respectively (Fig. 2), rpS3 is involved in the recognition of DNA damage and, due to its endonuclease activity, contributes to excision repair. In the case of rpS3 excess, its KH domain can bind p53 and stimulate apoptosis. Overexpression of rpS3 leads to DNA condensation, degradation of PARP protein and A/C lamin, as well as the expression of proapoptotic caspases 3, 8, and 9 [21]. rpS3 can also bind to bacterial proteins NleH1 and NLeH2, thereby regulating the cell response to microbial pathogenesis [29]. rpS3 exhibits MLP functions as PP or PPP protein rather than EP protein.

Other ribosomal proteins also exhibit MLP properties. So, during ribosomal stress, when the ribosome synthesis is disrupted, the ribosomal proteins rpL5, rpL11 and others suppress the ubiquitin ligase MDM2, which initiates the proteasomal degradation of the proapoptotic protein p53. This leads to stabilization and activation of p53 and stimulation of apoptosis. Fourteen ribosomal proteins demonstrated this property. Probably this response of the organism to genome instability was developed during evolution. In eukaryotic cells, genes encoding various ribosomal proteins are scattered across different chromosomes. However, for the assembly of a ribosome the amount of all ribosomal proteins must be equal. The genomic

Table 2. Selected examples of MLP proteins

Protein	Function 1	Functions 2 +	Type
ε-Cristallin	Element of lens	Lactate dehydrogenase (glycolysis)	PE
τ-Cristallin	Element of lens	Enolase (glycolysis)	PE
Lactate dehydrogenase	Glycolysis	ssDNA-binding protein	EP
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	Glycolysis	15–20 different functions including transferring receptor that mediates iron uptake; timidin hydrohylase; control of vesicular traffic; tRNA export from the nucleus; ssDNA-binding protein, etc. Adhesin, toxin, and invasion in bacteria	EE, EP
Aconitase	Krebs cycle (cytrate → isocytrate)	Fe homeostasis: Fe deficit activates expression of genes that regulate Fe uptake. Maintaining the stability of mitochondrial DNA	EE EP
Cytochrome *c*	Electron transport in mitochondria	Forms the complex with Apaf-1, dATP and procaspase 9 in the cytosol that initiates apoptosis	EP
rpS3	The protein of a small ribosome subunit	After phosphorylation by protein kinase PKCδ acquires endonuclease activity and participates in DNA repair; in a complex with NF-κB regulates gene expression	EP PP
STAT3	Transcription factor	Activates electron transport in mitochondria	PP
β-Catenin	Cell adhesion	Transcription factor	PP
p53	Transcription factor	Regulates cell cycle and DNA reparation; Stimulates apoptosis	PP

instability can lead to an imbalance that disrupts the assembly of functioning ribosomes. Thus, influencing the ability of MDM2 and p53 proteins to regulate apoptosis, the organism gets rid of cells with genome instability [22].

Cytochrome c

Cytochrome *c* is an ancient protein that is present at all levels of the evolutionary tree from microorganisms to eukaryotes. It plays a significant role in cellular bioenergetics. Cytochrome *c* transfers electrons in mitochondria, after translocation into the cytosol, it stimulates apoptosis [23, 24]. In eukaryotic cells,

apocytochrome *c*, produced in the cytosol, is transferred to the intermembrane space of mitochondria. There, two of its cysteine residues bind heme, thus creating cytochrome *c*, which diffuses along the surface of the inner mitochondrial membrane and transfers electrons from cytochrome c_1 of the complex III to the cytochrome oxidase complex IV. This reaction is redox transformation of the heme iron: $Fe^{3+} + e^- \leftrightarrow Fe^{2+}$. The three-dimensional structure of mitochondrial cytochrome *c* is conservative. Due to its simplicity, stability and accessibility the cytochrome *c*-mediated electronic transport was thoroughly studied at the atomic level. However, its moonlighting properties and structural changes in different situations have not
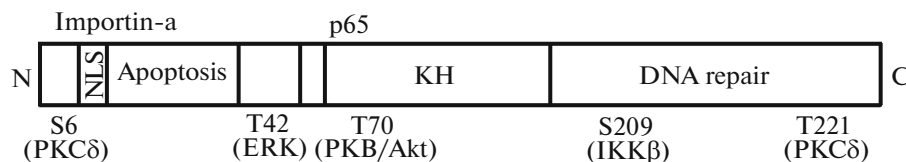

Fig. 2. Structural scheme of the ribosomal protein rpS3 and sites of post-translational modifications. There are sites of serine S6 and threonine T221 phosphorylation by protein kinase PKCδ; serine S209 by kinase IKKβ; threonines T42 and T70 by kinases ERK and PKB/Akt, respectively, indicated.
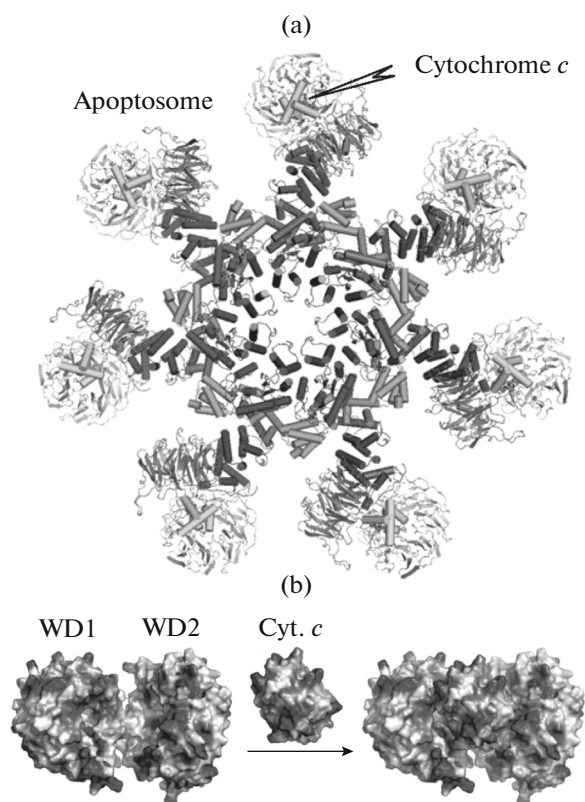
Fig. 3. The structure of the apoptosome formed by seven proteins Apaf-1 with integrated cytochromes *c* (a). Complementary binding of cytochrome to the WD1 and WD2 domains in the Apaf-1 protein (according to: Zhou et al., 2015 [24]).

yet been studied in detail. Cytochrome *c* has a relatively flexible structure. It rearranges easily during various molecular interactions and changes of intracellular localization. A number of structural modifications of cytochrome *c* were revealed under different conditions. It can also undergo various post-translational modifications: phosphorylation, nitrosylation, etc. Cytochrome *c* can be transferred into an alternative, partially expanded form, in which it does not transfer electrons, but is involved in moonlighting functions [23].

Cytochrome *c* may be released from mitochondria to the cytosol through the disruptions or Bax and Bak-mediated megachannels in the outer mitochondrial membrane. Cytoplasmic cytochrome *c* forms apoptosomes, a multi-protein complex that stimulates apoptosis [23–25]. Due to electrostatic interactions, it binds to two "propeller" domains (WD) of the cytosolic Apaf-1 (Apoptotic protease-activating factor 1) monomers. This changes the conformation of Apaf-1, in which bound ADP is replaced by dATP. As a result, an apoptosomes, a stable heptameric complex of 7 subunits "Apaf-1−dATP−cytochrome *c*," is formed (Fig. 3) [23−26]. Apoptosome cleaves procaspase 9, and active caspase 9 stimulates caspase 3, which fur-

ther triggers the apoptotic processes. Interestingly, such apoptosome structure is a late evolutionary acquirement. In nematode *C. elegans*, the apoptosomes consist of 4 CED-4 proteins, Apaf-1 homologs, whereas in Drosophila they consist of 8 Apaf-1-like proteins Dronc. These apoptosomes do not contain cytochrome c, which is present only in vertebrates [25, 27]. As shown recently, when DNA is damaged, cytochrome *c* can translocate to the nucleus, where it impedes the assembly of nucleosomes and thus reduces cell survival. However, the molecular mechanisms of this translocation and structural transformations of cytochrome *c* have not been elucidated yet [23].

Thus, cytochrome *c* is the EP-type MLP.

Bifunctional Proteins

Bifunctional proteins that are encoded by one gene have usually two active centers in the domains separated by a polypeptide chain, which perform different functions. This are EE-type MLP. Often they catalyze the sequential reactions of the same metabolic pathway, in which the substrate or intermediate is transferred between domains (substrate channeling). This can shorten their path and prevent their leakage into the environment and the interactions with the external metabolites. Moreover, the catalytic act in the center A stimulates the intermediate transfer and induces the opening of the "gate" for the intermediate in the center B. The linker connecting these domains plays an important role in stabilizing the conformation of one of the domains and transmitting a signal of the activation of the first domain to another one [28−32]. Aromatic amino acids are often involved in the "gate" mechanism due to their rotational ability. Their small rotation can cause a significant displacement of the side chains and a conformational transition that increases the width of the "gate" [29, 30].

Tryptophan synthase is an example of such a bifunctional protein. In it, the α domain cleaves indole-3-glycerol phosphate into indole and D-glyceraldehyde-3-phosphate. After the conformational transition and movement of indole along the 25-angstrom tunnel, the β domain catalyzes the binding of indole and serine and formation of tryptophan. In the middle of the tunnel in domain β, the walls of the tunnel contain Tyr279 and Phe280. The twists and turns of their aromatic rings provide a dynamic balance between alternative conformations that open or close the tunnel. The binding of the substrate to the α-subunit stimulates the tunnel closing. After substrate cleavage, the equilibrium between the open and closed forms is restored, and indole is transferred to the β-subunit [29].

Another bifunctional enzyme is glucosamine-6-phosphate synthase. In its N-domain, NH_3 is released from glutamine. This causes a conformational transi-
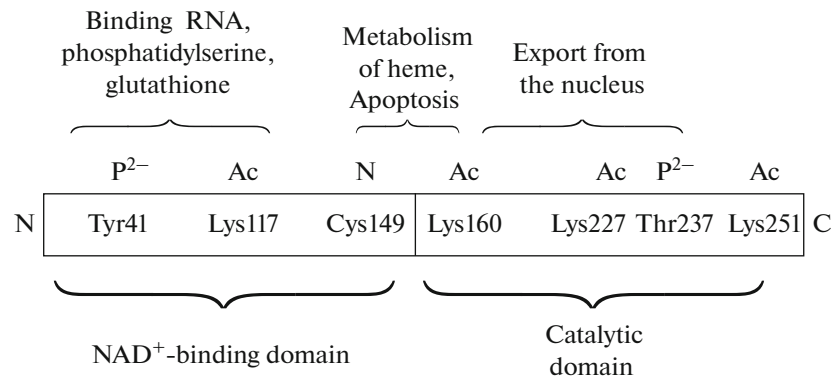
Fig. 4. The structure, functions and post-translational modifications of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Amino acids that are phosphorylated (P^{2-}), acetylated (Ac) and nitrosylated (N) indicated.

tion with the opening of an 18-angstrom tunnel, through which NH_3 is transferred on fructose-6-phosphate in the C-domain. This results in the synthesis of glucosamine-6-phosphate [29].

Metamorphic Proteins

The so-called metamorphic proteins can switch reversibly between fixed conformations under physiological conditions. This is possible if the potential barrier is not very large and the depth of the potential wells does not differ much. This case contrasts to the irreversible transition into a deeper potential well, a trap, that is characteristic for the stable conformation of a misfolded protein (misfolding trap). In metamorphic proteins a dynamic equilibrium is established between these forms. The structural transition is facilitated by the presence of a flexible region in the middle of the polypeptide chain.

A well-studied example of a metamorphic protein is the checkpoint protein Mad2, which controls the correct assembly of the mitotic spindle and initiates mitosis arrest if the assembly is incorrect [33]. Inactive open (O-Mad2) and active closed (C-Mad2) conformations of this protein are in dynamic equilibrium. The central core of Mad2 retains its structure, whereas the C- and N-terminal undergo to significant structural rearrangements. The transition between open and closed conformations ensures the correct attachment of microtubules to kinetochores, which is necessary for accurate chromosome separation during mitosis [33].

GAPDH

A striking example of MLP is the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH). This multifunctional protein is called the "quintessence" of MLP. Except glycolysis, it is involved in about 20 different functions. Such variety of its functions is determined by oligomerization,

intermolecular interactions, different microenvironments in various cellular compartments, and post-translational modifications [34–40]. On the cell surface, GAPDH forms a complex with the transferrin receptor, which mediates the iron uptake [38]. Binding of GAPDH to the cell membrane promotes membrane fusion and endocytosis. Cytoplasmic GAPDH is involved in vesicle transport from the endoplasmic reticulum to the Golgi complex. It also regulates the stability of mRNA. In the nucleus, GAPDH is involved in maintaining DNA integrity, regulation of gene expression, tRNA export from the nucleus. In addition, it regulates apoptosis [34–37]. Therefore, GAPDH may be considered as a central regulator of cell metabolism and as an information center [34–36, 39, 40].

GAPDH is a ubiquitous and abundant protein, highly conserved throughout the evolutionary tree. In somatic cells, it is encoded by a single gene. No alternative transcripts were found. GAPDH consists of 335 amino acids and has a mass of 37 kDa. Its coenzyme is NAD^+. GAPDH contains the NAD^+-binding domain (amino acids 1–150) and the catalytic domain (amino acids 151–335) (Fig. 4). The NAD^+-binding domain participates in mRNA stabilization and regulation of translation. The membrane functions of GAPDH are based on binding of phosphatidylserine by amino acids 70–94 of the NAD^+-binding domain at the inner side of the cellular membrane. Membrane-bound GAPDH catalyzes membrane fusion and regulates Fe^{2+} uptake, transport, and metabolism. GAPDH also binds glutathione by amino acids 67–77. Amino acids 258–270 in the catalytic domain are involved in the export of tRNA from the nucleus [37, 38, 41].

GAPDH expression is dynamic and sensitive to conditions in the cell: Ca^{2+} level, hypoxia, iron concentration, cell cycle stage, etc. [37]. The functional activity of GAPDH is significantly regulated. Post-translational modifications, such as acetylation of lysines 117, 160, 227 and 251, phosphorylation of thre-

onine 237 and tyrosine 41, nitrosylation of cysteine 149, etc. (Fig. 4) regulate multiple GAPDH functions. For example, cysteine 149 is very sensitive to oxidation, which makes GAPDH to be a redox sensor and homeostasis regulator in the cell [39, 40]. Its oxidation increases the GAPDH ability to bind to tRNA and DNA [41]. NO-mediated nitrosylation regulates heme metabolism, cell response to oxidative stress and apoptosis [38]. Acetylation of lysine 160 influences the expression of a number of genes encoding proapoptotic proteins p53, PUMA, and Bax. GAPDH phosphorylation affects vesicular transport, synaptic transmission [34, 35]. However, the exact biophysical and biochemical mechanisms of such diverse functions have not been fully elucidated yet.

Currently, more than 100 studies of the three-dimensional structure of GAPDH from microbes to humans, both separately and together with different substrates, intermediates, products, inhibitors, etc., have been carried out. GAPDH is usually a homotetramer, or to say more precisely, a dimer of two dimers, consisting of 4 identical subunits [37]. In the 3D structure of the GAPDH tetramer, one can see two transverse grooves with a length of 70 angstrom, in which the sites of NAD^+ and substrate binding are located. These regions can bind nucleic acids, since they contain many positively charged lysine and arginine residues. A narrow central channel (4×10 angstroms) well binds small molecules. Despite the abundance of structural data, much still needs to be clarified to understand the biophysical and biochemical mechanisms of reactions mediated by GAPDH. In glycolysis, GAPDH phosphorylates D-glyceraldehyde-3-phosphate (GAP) and converts it to 1,3-bisphosphoglycerate. This possibly occurs in two stages. First, cysteine C152 of GAPDH carries out a nucleophilic attack on GAP with transfer of hydride to NAD^+. H179 is the main catalyst in this process. Then, inorganic phosphate performs a nucleophilic attack on the carbonyl group of the thioether and phosphorylates it [37]. The mechanisms of other moonlighting functions of GAPDH are not elucidated so far. It is still unknown how the switching of functions and relocalization of GAPDH in cells are carried out. For example, it is unknown, how GAPDH that does not contain the nuclear localization signal is transported to the nucleus. Phosphorylation of serine 122 in GAPDH by protein kinase AMPK was shown to induce it nucleus localization. On the other hand, phosphorylation of threonine 237 by protein kinase Akt prevents GAPDH relocalization. The detail mechanisms of these processes are still unclear [37].

TRANSCRIPTION FACTORS

Transcription factors initiate the expression of many proteins that perform or regulate the complex cell function. Some transcription factors have relatively narrow action spectrum. They target genes involved in a certain cell function. For example, Nrf-2 regulates the expression of a group of antioxidant genes. However, other transcription factors are master regulators of a wide variety of cellular processes. The exceptional examples are c-Myc and p53, which regulate the expression of up to $10-15\%$ of all cellular genes [42, 43]. In addition to regulation of gene expression, different transcription factors are moonlighting proteins and can also perform other functions.

STAT3

The protein STAT3 (signal transducer and transcription activator 3) plays a key role in cell growth and survival. In the nucleus it functions as a transcription factor, but in mitochondria it regulates oxidative phosphorylation [44, 45]. The binding of interleukins 6 or 10 to the cytokine receptor on the cell surface activates the cytoplasmic kinase JAK, which phosphorylates tyrosine in STAT3. This initiates STAT3 dimerization and transfer to the cell nucleus, where it controls the expression of a number of genes [45]. However, by means of STAT3 various signaling pathways regulate bioenergetic processes in mitochondria. Indeed, phosphorylation of serine residues by cytoplasmic protein kinases initiates STAT3 translocation into mitochondria, where it stimulates electron transfer from complexes I and II to complex III and thereby activates oxidative phosphorylation (Fig. 5b) [44].

Two isoforms STAT3α and STAT3β with a molecular mass of 770 and 705 kDa, respectively, are known. The main domains in STAT3α are shown in Fig. 5a. The N-terminal domain (NTD; amino acids $1-137$) is involved in dimerization and tetramerization of STAT3 that enhances its transcriptional activity. The supercoiled domain CDD (amino acids $138-320$) binds various regulatory proteins. DBD domain binds to specific DNA regions (amino acids $321-493$). Linker domain LD (amino acids $494-582$) links the domain DBD to the SH2 domain (amino acids $583-687$) that binds STAT3 to phosphorylated receptors. Phosphorylation of tyrosine in STAT3 by protein kinases JAK and Src induces its dimerization and transfer to the nucleus. The C-terminal transactivation domain TAD (amino acids $688-770$ in STAT3α) is intrinsically disordered. Such intrinsically disordered region (IDR) folds differently when interacting with different molecular partners under formation of the transcriptional complexes [45, 46]. When interacting with DNA, STAT3 dimerizes so that DBD domains enwrap DNA. These domains can also bind with importin or exportin that import STAT3 into the nucleus or export it back to the cytoplasm [45, 46]. So, STAT3 is a moonlighting protein of the PP type.

β-Catenin

β-Catenin is a highly conserved protein of 781 amino acids. Its structure is rather particular. The
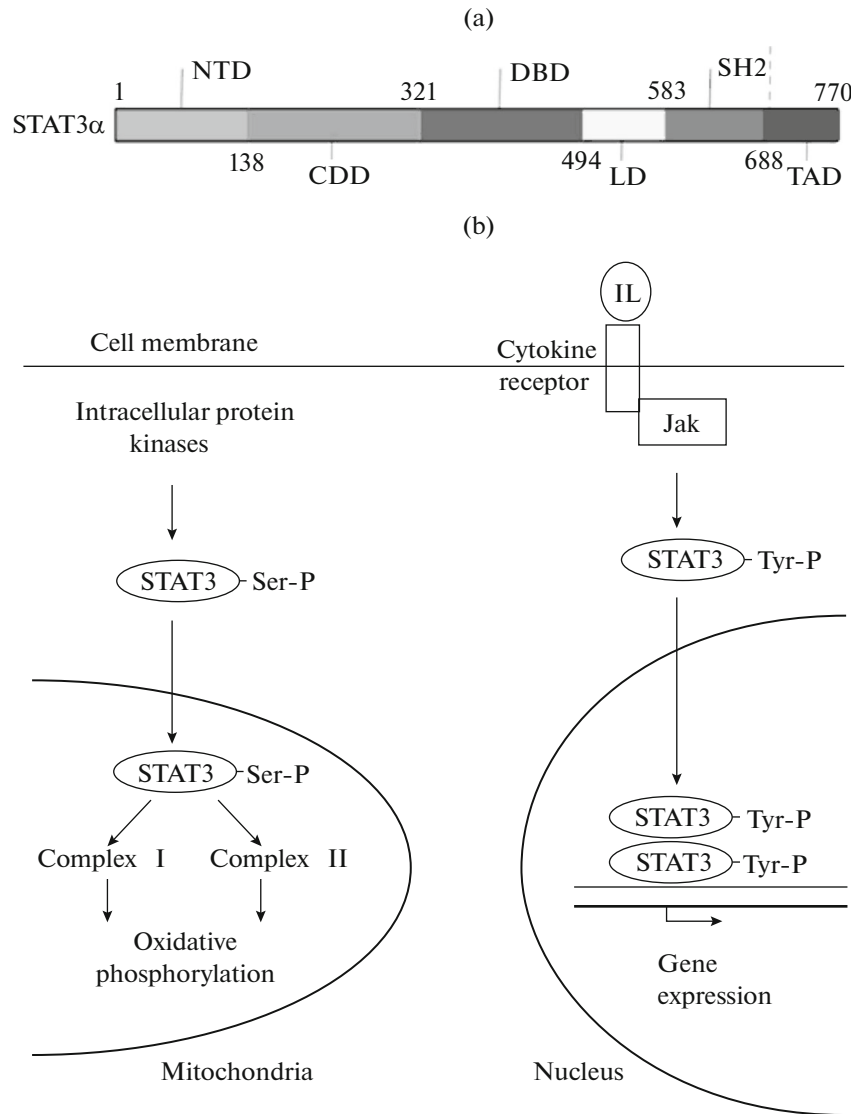
(a)

(b)

Fig. 5. Structure and functions of STAT3. (a) Scheme of the main structural domains of STAT3. NTD − N-terminal domain; CDD − supercoiled domain; DBD − DNA binding domain; LD − linker domain; SH2 − SH2 domain; TAD − transactivation domain. (b) Alternative functions of STAT3. When interleukines activate the cytokine receptor on the cell membrane, the JAK protein phosphorylates tyrosine in STAT3. This causes STAT3 dimerization, translocation to the nucleus, and stimulation of gene expression. If cytoplasmic protein kinases phosphorylate serine in STAT3, it translocates to mitochondria and activates electron transport from complexes I and II to complex III that increases the rate of oxidative phosphorylation.

central region contains several repetitions (12 in humans) of 40 amino acids that form a rigid elongated domain ARM (from the word armadillo) [47−49]. ARM has a slight curvature (Fig. 6). Its inner surface can bind various proteins or nucleic acids. The N-terminal and C-terminal fragments of β-catenin are intrinsically disordered. In solution, they don't have a specific structure. Nevertheless, they play a crucial role in the binding of β-catenin to other proteins or DNA. The N-terminal fragment contains a conserved short linear motif (SLiM, short linear motif), which after phosphorylation binds to ubiquitin ligase β-Trcp, that stimulates its proteasome degradation. The C-ter-

minal region interacts with DNA. It is a potent transcriptional activator. This fragment is not completely disordered: its segment at the C-end forms a stable helix (HelixC) near the ARM domain. This helix is not necessary for the participation of β-catenin in cell adhesion However, it is involved in the regulation of transcription [48, 49].

After synthesis, β-catenin moves to the plasma membrane, where, together with vinculin, α-catenin and catenin p120 it forms a platform for binding of the cytoplasmic fragment of cadherin (a protein that is involved in intercellular adhesion), to the intracellular actin cytoskeleton. Free β-catenin is practically absent
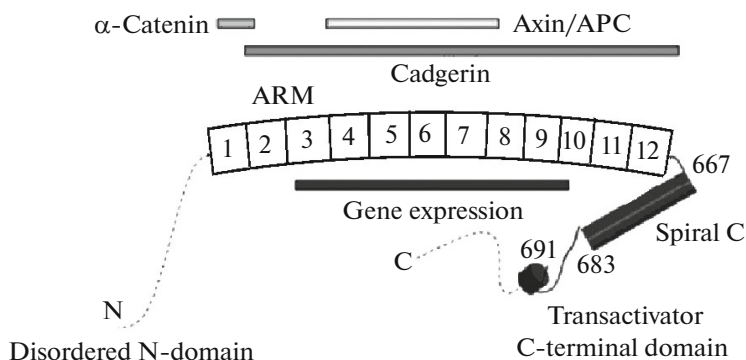
Fig. 6. The structure of human β-catenin with 12 repeats (ARM domain), intrinsically disordered N-domain and transactivation C-domain.
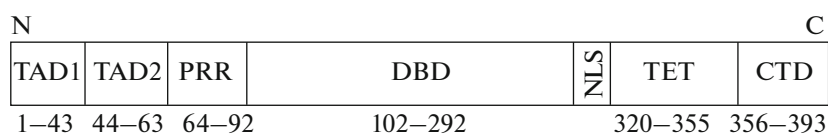

Fig. 7. Scheme of the main structural domains in p53: TAD1 and TAD2 − intrinsically disordered transactivator domains; PRR is a proline-rich region; DBD − DNA binding domain; NLS − nucleus localization signal; TET − the domain responsible for tetramerization; CTD − intrinsically disordered C-terminal domain.

in the cytoplasm. It binds to the multienzyme complex that consists of proteins APC, axin1, GSK3 and CK1. In this complex protein kinases GSK3 and CK1 phosphorylate it that causes it ubiquitinization by β-TrCP ligase and proteasomal degradation. However, the binding of the extracellular molecule Wnt to its receptor causes disintegration of the APC complex, and free β-catenin appears in the cytoplasm. It is translocated into the nucleus. As a transcription factor it regulates the expression of various genes that encode the components of the Wnt signaling pathway, proteins c-Myc, COX, etc. (Fig. 8d) [47]. So, β-catenin is MLP of the PP type.

p53

p53 protein is especially important and interesting. It is expressed in all cells of the organism, where it regulates basic cellular functions, including metabolism, cell cycle, DNA repair, survival, and apoptosis [50−52]. As a transcription factor, it regulates the expression of hundreds genes [43, 53]. p53 stimulates apoptosis of cells with unrepairable DNA damage [54−57]. This provides elimination of malignant cells and protection of the body from malignant tumors. p53 is a tumor suppressor and guardian of the genome. Mutations in the *TP53* gene that inactivate p53 were found in half cases of human cancer. Except regulation of gene expression, p53 controls mitochondrial functions
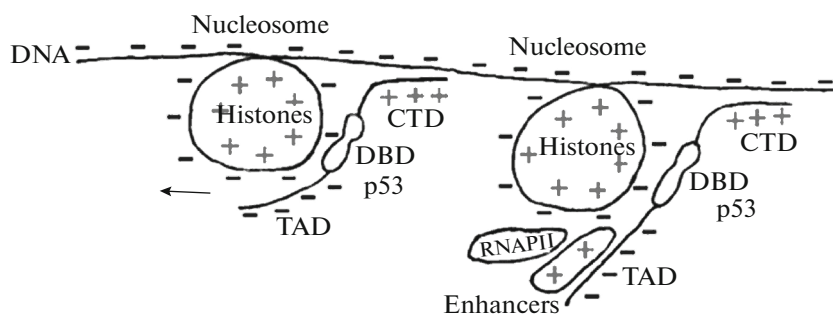

Fig. 8. Hypothetic interaction of p53 with the nucleosome: "Rolling" of the nucleosome and interaction with positively charged domains of proteins bound to DNA.

independently on transcription and induces apoptosis in cells with mitochondrial failure [58−60]. This ancient conservative protein was found at the early evolutionary stages in protozoa and sponges, much earlier than cancerous tumors have emerged. It is assumed that the true role of the p53 family proteins is to maintain the integrity of the genome under adverse effects on the organism [61, 62].

The p53 level in the cytoplasm is typically low. After synthesis in the cytoplasm, it is transported to the nucleus, where it binds to DNA. The unbound p53 forms a complex with MDM2, which ubiquitinates it and transfers back to the cytoplasm, where it additionally ubiquitinized and rapidly degrades in proteasomes [63]. Upon radiation damage to DNA, oxidative stress, or activation of oncogenes, p53 is phosphorylated by protein kinases JNK, p38, ERK, etc. [57]. This prevents its interaction with MDM2 and degradation. The level of p53 in the nucleus significantly increases [51, 62, 64, 65]. It tetramerizes, binds to DNA and stimulates the transcription of large groups of genes that contain a specific nucleotide sequence p53RE (p53-response element) in the regulatory region. More than 600 p53REs are already known [53]. The result of p53 activation is arrest of the cell cycle and DNA replication. This is followed by DNA repair. In the case of a strong stress impact, apoptosis is triggered [57, 58, 64]. The p53 protein is also found in the cell nucleoli, which are known as "ribosome factories". Violation of the synthesis of ribosomes in the nucleus increases the level of p53, which transmits a signal about cell damage to the systems that control cellular metabolism, homeostasis, and survival [66, 67].

p53 is a polypeptide chain from 393 amino acids. In cells, it forms a tetramer of two identical dimers. One can distinguish different functional modules in p53 (Fig. 9a). The transactivation domain TAD at the N-end is subdivided into two subdomains TAD1 and TAD2 (amino acids 1−43 and 44−63). They are followed by a proline-rich region PRR (amino acids 64−92). A DNA-binding domain DBD (amino acids 102−292) recognizes p53RE and binds to it. The nuclear localization sequence NLS, the tetramerization domain TET (amino acids 320−355), and the C-terminal domain CTD (amino acids 356−393) follow further. The DBD and TET domains are highly conserved. The intrinsically disordered domains TAD and CTD are more variable. This allows p53 to interact with many partners flexibly [53, 68−71].

The DBD domains in the p53 tetramer (Fig. 9b) that interact with p53RE differently bend DNA in regions with different nucleotide sequences (Fig. 9c) [72]. In this case, DNA is partially untwisted that facilitates protein binding.

The TAD domain (TAD1 + TAD2) consisting of 63 amino acids is negative due to 17 negatively charged residues Glu− and Asp−. It is intrinsically disordered. Its conformation is determined by the interaction with other proteins or DNA. It is known that human p53 can interact with more than 1000 different proteins, including coactivators, enhancers and other transcriptional regulators, as well as proteins that carry out post-translational modifications: protein kinases, acetyltransferases, etc. [73].

The C-terminal domain CTD from 29 amino acids is also intrinsically disordered. Unlike negative TAD, it contains 7 positively charged amino acid residues (6 Lys+ and Arg+). It can interact with negatively charged DNA phosphates, regardless of the complex shape of DNA fragments: curved, supercoiled, or wound on nucleosomes [69]. CTD plays an important role in the activation, intracellular localization, and degradation of p53 [62]. Due to the relatively weak and non-specific electrostatic binding of a positively charged CTD to negatively charged DNA, p53 can move along DNA in search of the site, where the positive DBD domains can bind to p53RE more robustly and specifically [74]. This facilitates the navigation of transcription factors: the three-dimensional diffusion is replaced by the one-dimensional process [68, 75−77]. (The energy source and the mechanism for determining of the motion direction are unknown. The motion direction in the diffusion process is commonly determined by the concentration gradient, which is absent in the case of movement of a single molecule). Another problem: how proteins moving along the chromosome (RNA polymerases, transcription factors, DNA repair complexes) pass through nucleosomes with wound DNA. One can assume that the negatively charged TAD due to electrostatic repulsion can shift the nucleosome, which carries a negatively charged DNA on the surface. This should open the DNA sites for transcription or reparation. p53 is known to bind often to p53RE located on nucleosomes This probably occurs because the DNA affinity for p53 is increased at the places of its bending on the nucleosome [53, 78, 79]. In this case, the positively charged regions of proteins interacting with DNA (RNA polymerase, DNA repair proteins, coactivators or transcription enhancers) can be inserted between the negatively charged DNA and TAD and neutralize the negative charges of DNA (Fig. 10).

The ability of some transcription factors such as p53 or c-Myc to regulate non-specifically the expression of many hundreds of genes seems mysterious. It is assumed that the regulation of numerous genes is associated not only with specific binding of p53 to p53RE in the promoters of certain genes, but also by its binding to strong enhancers. This can increase its overall transcriptional activity [53]. Nevertheless, the activation of expression of different gene groups by these transcription factors is to some extent specific [80]. What is the structural basis for specific binding of p53 to a certain p53RE? Another one question: what is the mechanism of differential p53 reactions to different stimuli? One hypothesis associates this effect with

intrinsically disordered regions (IDR), which can interact with a variety of proteins. The presence of IDR is characteristic for various transcription factors [81]. In p53, TAD can interact with many proteins that ensure the p53 sensitivity to different stimuli and its ability to stimulate the expression of various genes [53]. Due to the rigid proline linker PRR, TAD protrudes outward from the central DBD/DNA complex and can interact with a variety of coactivator and enhancer proteins [70]. The affinity of p53 to different partners is regulated by different post-translational modifications, especially phosphorylation and acetylation of TAD [70]. It is also assumed that the p53-regulated differential expression of various gene groups, which control different cellular processes, is associated with the local differences in chromatin structure: DNA topology, p53RE location, DNA methylation of promoters, approaching of distant enhancers, and accessibility of gene promoters for RNAPII, etc. [53]. In particular, the selectivity of p53 action may be related to its ability to differently bend DNA in sites with different nucleotide sequences [72, 74].

Various post-translational modifications regulate p53 activity In p53 36 amino acids can be phosphorylated, methylated, acetylated, glycosylated, etc. [82]. As noted above, phosphorylation of serines 18 and 20 in TAD prevents MDM2 binding that leads to p53 degradation [57]. Acetylation of various lysines differentially enhances the transcriptional activity of p53 and leads to expression of various groups of genes that causes different cellular responses. The non-acetylated protein p53 stimulates the expression of MDM2 and other proteins. This prevents its excessive activation. In the case of DNA damage, p53 is acetylated by acetyltransferase CBP/p300 and, together with proteins Tip60 and MOF, stimulates the expression of proteins p21 and GADD45, which induce the cell cycle arrest and DNA repair [82]. If the DNA damage caused by harmful factors (hypoxia, excitotoxicity, oxidative stress, ionizing radiation or impaired nucleolar function) is extremely strong and unreparable, the expression of p53 increases. In this case, almost all lysine residues are acetylated. p53 stimulates the expression of genes encoding proapoptotic proteins: caspases 6, Apaf-1, HtrA2, Bax, Bid, NOXA, PUMA, Fas, DR4, DR5, etc. This causes apoptotic cell death [64−67]. However, p53 can induce apoptosis not only through transcriptional activity, but also independently of transcription. The cytoplasmic p53 binds directly to the outer mitochondrial membrane, inhibits anti-apoptotic proteins Bcl-2 and Bcl-XL, activates proapoptotic proteins Bax and Bid, and stimulates the Bax/Bak-mediated formation of the megapores in the outer mitochondrial membrane. Through these pores cytochrome c, SMAC/Diablo and other proapoptotic can be released into the cytosol and cause apoptosis [54, 58−60].

Thus, p53 is a multifunctional moonlighting protein that acts as a transcription factor. Its functions are mediated by the expression of many proteins. As a proapoptotic agent, it can directly affect mitochondria.

CONCLUSION

The multifunctionality of proteins is not a rare phenomenon. It is characteristic of hundreds proteins. This expands significantly the functional diversity, which is limited, for the first glance, by a small number of genes. Why and how moonlighting and more complex multifunctional proteins appeared are an exciting question, which, like other evolutionary problems, has not been answered yet. According to one idea, the indifferent mutations that affect the surface regions of proteins, which are not critical for their canonical function, can lead to the emergence of new functions. There are many examples where point mutations affecting one or two amino acids can significantly change the surface properties, thereby changing the participation of a given protein in the supramolecular complex or its localization in the cell. Such mutations may be one of the mechanisms of MLP emergence. The following gene duplication, amplification and natural selection occur if these mutations lead to useful properties [14]. The probability of the emergence of new functions is higher in abundant proteins, which are numerous in the cell, and their collision with partner proteins occur more often. This is why many housekeeping proteins, such as proteins of glycolysis, ribosomes, molecular motors, etc., have moonlighting functions [83, 84]. They are found in bacteria, yeast, and in higher organisms [83−87]. The widespread occurrence of moonlighting proteins in nature has been suggested to be a reservoir of functions for the adaptation to a changing environment [87].

An important factor of this process is the imbalance in the synthesis, work, and degradation of such proteins. For example, the overproduction of some ribosomal proteins causes their excess in the cytoplasm. The excessive proteins such as rpS3 that are not integrated into ribosomes can be used in other processes, such as gene repair and regulation of gene expression. A striking and complex example of the imbalance between protein synthesis and degradation is β-catenin. If this protein is not integrated into the adhesive complex in the cell membrane, it degrades. But the binding of the extracellular Wnt signal to the surface receptors prevents degradation of β-catenin. It is transferred to the nucleus, where regulates gene expression. Intracellular localization of proteins and switching of their functions is regulated by post-translational modifications. This is demonstrated by STAT3 protein, which, after tyrosine phosphorylation, moves into the nucleus, where it regulates transcription. However, after phosphorylation of serines by cytoplasmic protein kinases STAT3 moves to mitochondria, where it activates electron transport. Thus, the source of information in the cell is not only the

information on the primary structure of proteins that is recorded in the nucleotide sequence of DNA and determines, which proteins will be synthesized, but also the information coming from outside about the physico-chemical state of the environment (temperature, ionic and gas composition, pH, free radicals, contacts with neighboring cells, etc.). Signaling molecules coming from other cells trigger the intracellular signaling cascades, which regulate gene expression and determine the dynamics and amount of synthesized proteins.

Classical ideas on the role of peripheral amino acid sequences that are not included in the active center and in its immediate environment are currently expanded. The periphery of the protein globule is not only a complementary surface for the assembly of supramolecular complexes or the incorporation of a protein into the membrane. It is involved in a wider range of intermolecular interactions. The computer methods developed in recent years for searching, identifying and predicting of multifunctional proteins [88, 89] have shown that intrinsically disordered regions play an important role in protein interactions that mediate formation of supramolecular complexes, signaling and metabolic chains and networks. IDRs are widespread in many signaling proteins and transcription factors [81, 90−95]. They are often located at the N- or C-ends of the polypeptide chain. To date, more than 650 thousand types of protein-protein interactions have been discovered [89, 95]. The mentioned above transcription factors STAT3, β-catenin and p53 also contain IDR. Some whole proteins, such as calmodulin or chaperones, are also disordered. IDRs are more labile and mutate more often. They are more variable and can play an important role in the evolutionary origin of MLP proteins [96−98]. Conservative short linear sequences (SLiM, short linear motif) of 3−10 amino acids was found to play an essential role in their interactions [96−98]. Protein interactions in SLiM are relatively weak, with affinity in the micromolar range. Therefore, they are short-term and reversible. They are easily modulated by post-translational modifications that switch the structure and functions of the protein. The decisive role in the interactions and conformational changes of MLPs and in switching of their functions is performed by their protein or nucleic partners located within the interaction zone.

Thus, although the number of genes is very limited, many proteins encoded by one gene have developed the ability to perform two or more functions, which are realized depending on the context: cell localization, microenvironment, the formation of complexes with other proteins, post-translational modifications, etc. Multifunctional proteins raise a number of questions on their primary and tertiary structure, on the relationship of structure with functions, on the mechanisms of structural and functional switching, on the folding of proteins consisting of several domains and

their refolding during relocalization, on the self-assembly of multiprotein and nucleoprotein complexes, on the role of MLP in the cell signaling system, on the emergence and transformations of MLP in the course of evolution, etc.

REFERENCES

1. G. W. Beadle and E. L.Tatum, Proc. Natl. Acad. Sci. U. S. A. **27**, 499 (1941).
2. C. B. Anfinsen, Fed. Proc. **16**, 783 (1957).
3. F. Moraes and A. Goes, Biochem. Mol. Biol. Educ. **44**, 215 (2016).
4. UniProt Consortium. Nucleic Acids Res. **43** (Database issue), D204 (2015).
5. S. K. Burley, H. M. Berman, G. J. Kleywegt, et al., Methods Mol. Biol. **1607**, 627 (2017).
6. F. E. Regnier and J. H. Kim, Anal. Chem. **90**, 361 (2018).
7. O. I. Kiseleva, A. V. Lisitsa, and E. V. Poverennaya, Mol. Biol. (Moscow) **52**, 394 (2018).
8. S. Y. Chen, C. Li, X. Jia, and S. J. Lai, Int. J. Mol. Sci. **20**, E3834 (2019).
9. M. L. Tress, F. Abascal, and A. Valencia, Trends Biochem. Sci. **42**, 98 (2017).
10. K. V. Nguyen, Biomol. Concepts **6**, 11 (2015).
11. Multitasking Proteins DataBase (Moonlighting Proteins Database). http://wallace.uab.es/multitaskII.
12. M. Mani, C. Chen, V. Amblee, et al., Nucleic Acids Res. **43**, D277 (2015).
13. C. J. Jeffery, Philos. Trans. R. Soc. Lond. B **373**, 1738 (2018).
14. C. J. Jeffery, J. Proteomics **134**, 19 (2016).
15. K. W. Min, S. H. Lee, and S. J. Baek, Cancer Lett. **370**, 108 (2016).
16. S. D. Copley, Bioessays **34**, 578, (2012).
17. B. Henderson and A. C. Martin, Biochem. Soc. Trans. **42**, 1671 (2014).
18. C. E. Chapple and C. Brun, Oncotarget **6**, 16812 (2015).
19. F. Wan, D. E. Anderson, R. A. Barnitz, et al., Cell **131**, 9279 (2007).
20. 20. B. Pertschy, Microb. Cell **4**, 140 (2017).
21. X. Gao and P. R. Hardwidge, Front. Microbiol. **2**, 137 (2011).
22. T. H. Kim, P. Leslie, and Y. Zhang, Oncotarget **5**, 860 (2014).
23. D. Alvarez-Paggi, L. Hannibal, M. Castro, et al., Chem. Rev. **117**, 13382 (2017).
24. M. Zhou, Y. Li, Q. Hu, et al., Genes Dev. **29**, 2349 (2015).
25. L. Dorstyn, C. W. Akey, and S. Kumar, Cell Death Differ. **25**, 1194 (2018).

26. Y. P. Ow, D. R. Green, Z. Hao, and T. W. Mak, Nat. Rev. Mol. Cell. Biol. **9**, 532 (2008).

27. A. Oberst, C. Bender, and D. R. Green, Cell Death Differ. **15**, 1139 (2008).

28. H-X. Zhou, S. T. Wlodek, and J. A. McCammon, Proc. Natl. Acad. Sci. U. S. A. **95**, 9280 (1998).

29. H.-X. Zhou and J. A. McCammon, Trends Biochem. Sci. **35**, 179.10.007 (2010).

30. F. M. Raushel, J. B. Thoden, and H. M. Holden, Acc. Chem. Res. **36**, 539 (2003).

31. N. K. Nagradova, FEBS Lett. **487**, 327 (2001).

32. N. Nagradova, IUBMB Life **55**, 459 (2003).

33. S. C. Goodchild, P. M. G. Curmi, and L. J. Brown, Biophys Rev. **3**, 143 (2011).

34. M. A. Sirover, Biochim. Biophys. Acta − Gen. Subj. **1810**, 741 (2011).

35. M. A. Sirover, Cancer Metastasis Rev. **37**, 665 (2018).

36. H. He, M. Lee, L. L. Zheng, et al., Biosci. Rep. **33**, e00018 (2013).

37. M. R. White and E. D. Garcin, Subcell. Biochem. **83**, 413 (2017).

38. V. M. Boradia, M. Raje, and C. I. Raje, Biochem. Soc. Trans. **42**, 1796 (2014).

39. M. Schneider, J. Knuesting, O. Birkholz, et al., BMC Plant Biol. **18**, 184 (2018).

40. T. Hildebrandt, J. Knuesting, C. Berndt, et al., Biol. Chem. **396**, 523 (2015).

41. M. A. Sirover, J. Biochem. Cell Biol. **57**, 20 (2014).

42. P. S. Knoepfler, Cancer Res. **67**, 5061 (2007).

43. M. Fisher, Oncogen **36**, 3943 (2017).

44. M. G. Myers, Science **323**, 723 (2009).

45. A. Subramaniam, M. K. Shanmugam, E. Perumal, et al., Biochim. Biophys. Acta **1835**, 46 (2013).

46. J. Sgrignani, M. Garofalo, M. Matkovic, et al., Int. J. Mol. Sci. **19**, E1591 (2018). https://doi.org/10.3390/ijms19061591

47. B. T. MacDonald, K. Tamai, and X. He, Dev. Cell **17**, 9 (2009).

48. Y. Xing, K. Takemaru, J. Liu, et al., Structure **16**, 478 (2008).

49. C. J. Gottardi and M. Peifer, Structure **16**, 336 (2008).

50. M. Napoli and E. R. Flores. Br. J. Cancer **116**, 149 (2017).

51. V. Marcel, F. Nguyen Van Long, and J. J. Diaz, Cancers (Basel) **10**, E152 (2018).

52. F. M. Simabuco, M. G. Morale, I. C. B. Pavan, et al., Oncotarget **9**, 23780 (2018).

53. K. D. Sullivan, M. D. Galbraith, Z. Andrysik, and J. M. Espinosa, Cell Death Differ. **25**, 133 (2018).

54. B. J. Aubrey, G. L. Kelly, A. Janic, et al., Cell Death Differ. **25**, 104 (2018).

55. S. Nicolai, A. Rossi, N. Di Daniele, et al., Aging (NY) **7**, 1050 (2015).

56. L. J. Hofseth, S. P. Hussain, and C. C. Harris, Trends Pharmacol. Sci. **25**, 177 (2004).

57. C. Culmsee and M. P. Mattson, Biochem. Biophys. Res. Commun. **331**, 761 (2005).

58. C. Wan, X. Ma, S. Shi, et al., Toxicol. Appl. Pharmacol. **281**, 294 (2014).

59. D. B. Wang, C. Kinoshita, Y. Kinoshita, and R. S.Morrison, Biochim. Biophys. Acta **1842**, 1186 (2014).

60. Q. Dai, T. T. Luo, S. C. Luo, et al., J. Bioenerg. Biomembr. **48**, 337 (2016)

61. E. Aberg, F. Saccoccia, M. Grabherr, et al., BMC Evol. Biol. **17**, 177 (2017).

62. A. C. Joerger and A. R. Fersht, Annu. Rev. Biochem. **85**, 375 (2016).

63. V. Gottifredi and C. Prives, Science **292**, 1851 (2001).

64. P. Bonini, S. Cicconi, A. Cardinale, et al., J. Neurosci. Res. **75**, 83 (2004).

65. S. Rashi-Elkeles, R. Elkon, S. Shavit, et al., Mol. Oncol. **5**, 336 (2011).

66. R. Parlato and G. Kreiner, J. Mol. Med. (Berl.) **91**, 541 (2013).

67. S. J. Woods, K. M. Hannan, R. B. Pearson, and R. D. Hannan, Biochim. Biophys. Acta − Gene Regul. Mech. **1849**, 821 (2015).

68. M. Kitayner, H. Rozenberg, N. Kessler, et al., Mol. Cell. **22**, 741 (2006).

69. E. Fadda and M. G. Nixon, Phys. Chem. Chem. Phys. **19**, 21287 (2017).

70. A. C. Joerger and A. R. Fersht, Cold Spring Harb. Perspect. Biol. **2**, a000919 (2010). https://doi.org/10.1101/cshperspect.a000919

71. A. C. Joerger and A. R. Fersht, Annu. Rev. Biochem. **77**, 557 (2008).

72. Y. Pan and R. Nussinov, J. Biol. Chem. **282**, 691(2007).

73. Y. S. Tan, Y. Mhoumad, and C. S. Verma. J. Mol. Cell Biol. **11**, 306 (2019).

74. Y. Pan and R. Nussino, J. Mol. Recognit. **23**, 232 (2010).

75. K. McKinney, M. Mattia, V. Gottifredi, and C. Prives, Mol. Cell. **16**, 413 (2004).

76. A. Tafvizi, F. Huang, A. R. Fersht, et al., Proc. Natl. Acad. Sci. U. S. A. **108**, 563 (2011).

77. A. Murata, Y. Itoh, E. Mano, et al., Biophys. J. **112**, 2301 (2017).

78. N. E. Lidor, Y. Field, Y. Lubling, et al., Genome Res. **20**, 1361 (2010).

79. F. Cui and V. B. Zhurkin, Nucleic Acids Res. **42**, 836 (2014).

80. Y. Pan, C. J. Tsai, B. Ma, and R. Nussinov, Trends Genet. **26**, 75 (2010).

81. J. Liu, N. B. Perumal, C. J. Oldfield, et al., Biochemistry **45**, 6873 (2006).

82. J. P. Kruse and W. Gu, Cell **137**, 609 (2009).

83. S. D. Copley, Biochem. Soc. Trans. **42**, 1684 (2014).

84. W. Lu, V.I. Gelfand. Trends Cell Biol. **27**, 505 (2017).

85. B. Henderson, Biochem. Soc. Trans. **42**, 1720 (2014).

86. C. Gancedo, C. L. Flores, and J. M. Gancedo, Biochem. Soc. Trans. **42**, 1715 (2014).

87. M. A. Fares, Biochem. Soc. Trans. **42**, 1709. (2014).

88. C. E. Chapple, B. Robisson, L. Spinelli, et al., Nat. Commun. **6**, 7412 (2015).

89. G. Hu, Z. Wu, V. N. Uversky, and L. Kurgan, Int. J. Mol. Sci. **18**, 2761 (2017).

90. L. M. Iakoucheva, C. J. Brown, J. D. Lawson, et al., J. Mol. Biol. **323**, 573 (2002)

91. R. van der Lee, M. Buljan, B. Lang, et al., Chem. Rev. **114**, 6589 (2014).

92. N. S. Latysheva, T. Flock, R. J. Weatheritt, et al., Prot. Sci. **24**, 909 (2015).

93. A. Balcerak, A. Trebinska-Stryjewska, R. Konopinski, et al., Open Biol. **9**,190096 (2019). https://doi.org/10.1098/rsob.190096

94. E. A. Lemke, J. Mol. Biol. **428** (10, Pt. A), 2011 (2016).

95. N. N. Sluchanko and D. M. Bustos. Prog. Mol. Biol. Transl. Sci. **166**, 19 (2019).

96. A. Zanzoni, D. M. Ribeiro, and C. Brun, Cell Mol. Life Sci. **76**, 4407 (2019).

97. R. J. Weatheritt, N. E. Davey, and T. J. Gibson, Nucleic Acids Res. **40**, 7123 (2012).

98. K. Van Roey, B. Uyar, R. J. Weatheritt, et al., Chem. Rev. **114**, 6733 (2014).