ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 12, pp. 1321-1336. © Pleiades Publishing, Ltd., 2011.
Published in Russian in Biokhimiya, 2011, Vol. 76, No. 12, pp. 1640-1658.

Influence of Intramolecular Interactions on Conformational and Dynamic Properties of Analogs of Heptapeptide AFP_{14-20}

N. T. Moldogazieva[1]*, K. V. Shaitan[2], M. Yu. Antonov[2], I. K. Vinogradova[1], and A. A. Terentiev[1]

[1]*Russian State Medical University, ul. Ostrovityanova 1, 117997 Moscow, Russia; fax: (495) 434-0588; E-mail: nmoldogazieva@mail.ru; aaterent@inbox.ru*
[2]*Biological Faculty, Lomonosov Moscow State University, 119991 Moscow, Russia; fax: (495) 939-2374; E-mail: shaitan@moldyn.org*

Received March 5, 2011
Revision received April 29, 2011

Abstract—Conformational and dynamic properties of proteins and peptides play an important role in their functioning. However, mechanisms that underlie this influence have not been fully elucidated. In the present work we computationally constructed analogs of heptapeptide AFP_{14-20} (LDSYQCT) – one of the biologically active sites of human α-fetoprotein (AFP) – to study their conformational and dynamic properties using molecular dynamics simulation. Analogs were obtained by point substitutions of amino acid residues taking into account differences in their physicochemical properties and also on the basis of analysis of amino acid substitutions in the AFP_{14-20}-like motifs revealed in different physiologically active proteins. It is shown that changes in conformational mobility of amino acid residues of analogs are due to disruption or arising of intramolecular interactions that, in turn, determine existence of steric restrictions during rotation around covalent bonds of the peptide backbone. Substitution of an amino acid by another one with significant difference in physicochemical properties may not lead to remarkable changes in conformational and dynamic properties of the peptide if intramolecular interactions remain unchanged.

DOI: 10.1134/S0006297911120054

Key words: α-fetoprotein, AFP, analogs of heptapeptide AFP_{14-20}, molecular dynamics, intramolecular interactions

Heptapeptide LDSYQCT is the biologically active site of human α-fetoprotein (AFP) that includes amino acid (a.a.) residues 14-20 in the mature protein molecule [1]. The heptapeptide was first discovered by comparison of primary structures of AFP, albumin, and epidermal growth factor (EGF) as the site similar to the part of the receptor-binding site of human EGF [2]. Afterwards the peptide LDSYQCT (designated as AFP_{14-20}) was synthesized by solid-phase chemistry with subsequent testing of biological activity. It was shown that it is able to regulate proliferation and apoptosis of lymphocytes and possesses immunosuppressive activity [3-5].

Presently, it is recognized that the majority of eucaryotic proteins are multimodular and polyfunctional ones [6-8]. The modular character of organization means that the protein molecule may be composed of a few (and even multiple) structurally and functionally independent elements (domains and/or motifs). Each module may be responsible for a certain function of the protein and may act independently. Due to this feature the protein molecule may possess a set of different functions.

It is supposed that proteins that demonstrate similar functions contain similar peptide motifs responsible for these functions [9]. Such linear motifs represent short segments of polypeptide chain of a few (usually up to 8-10) amino acid residues in length. Analysis of structure–function relationships between homologous and non-homologous proteins allows predicting functions of a certain protein on the basis of revealing of short amino acid sequences in its primary structure if they are similar to those in physiologically active proteins.

Motifs similar to the heptapeptide LDSYQCT (or AFP_{14-20}-like motifs) have been revealed in growth factors of the EGF superfamily as well as in EGF-like domains of some cell adhesion proteins, coagulation factors, and enzymes [10]. This may indicate existence of functions common for all these proteins. Thus, the AFP_{14-20}-like motifs may be considered as structural markers of pro-

Abbreviations: AFP, α-fetoprotein; EGF, epidermal growth factor; MD, molecular dynamics.

* To whom correspondence should be addressed.

teins that regulate proliferation, differentiation, migration, and apoptosis of embryonic and tumor cells.

In this connection special attention was paid to studying of the role of individual amino acid residues in functioning of the heptapeptide LDSYQCT and similar linear motifs. This allows revealing mechanisms of their action, including a role of different amino acid residues in protein–protein interactions and receptor binding and also investigating ways of developing new therapeutic agents on the basis of this heptapeptide.

Earlier, using a molecular dynamics (MD) simulation method we studied a set of the heptapeptide LDSYQCT analogs constructed by point substitutions of amino acid residues [11, 12]. Presently, MD methods are widely used to study conformational and dynamic properties of biomacromolecules and mechanisms of their functioning [13, 14]. These methods allow modeling details of conformational changes that take place in molecules, including proteins and peptides.

In the present work we constructed new analogs of the heptapeptide LDSYQCT on the basis of analysis of amino acid residue substitutions in AFP_{14-20}-like motifs of physiologically active proteins. Amino acid sequences of these proteins were extracted from databases UniProtKB/Swiss-Prot, GenBank, etc. [15, 16] by local alignment method using the FASTA and BLAST programs [17-20].

Conformational and dynamic properties of analogs obtained were investigated using the method of equilibrium MD using an implicit solvent model with characteristics of water [14]. Evaluation of changes in dynamic and conformational properties of amino acid residues of the peptides were performed using two-dimensional (2D) and three-dimensional (3D) maps of free energy levels (Poincare sections), and also autocorrelation functions of dihedral angles φ, ψ and χ.

METHODS OF INVESTIGATION

Searching AFP_{14-20}-like motifs. In order to reveal short amino acid sequences similar to LDSYQCT in physiologically active proteins the method of local alignment (searching for sites of local similarity) using programs FASTA (University of Virginia) [17-19] and BLAST (NCBI, National Center for Biotechnology Information) [20] were used.

In case of the FASTA program amino acid sequences of physiologically active proteins were extracted from the UniProtKB/Swiss-Prot database, which is the most comprehensive annotated database that contains curated protein information of protein sequence, function, classification, and cross-references. The algorithm GLFASTA (global/local protein–protein search) (version 35.04) with default parameters (amino acid substitution matrix Blosum50, open – 12, ext – 2) was used. The GLFASTA algorithm is the best suitable for local alignment of proteins to search for similar short peptide motifs [21].

In case of the BLAST program (version 2.2.23+) algorithms Blastp (protein–protein BLAST) and PSI-BLAST (Position Specific Iterated BLAST), which are the most proper to goals of the present work, were used [20]. Here non-redundant protein sequence databases including GenBank, PDB, UniProtKB/Swiss-Prot, etc. [15, 16, 22] were searched. The following parameters were used: max target sequences – 100 (500 for PSI-BLAST), substitution matrix – BLOSUM62, expect threshold – 10 (for PSI-BLAST 0.005), word size – 3.

Structural and functional characterization of proteins, in which AFP_{14-20}-like motifs were revealed, was performed using the annotated database UniProtKB/Swiss-Prot. The annotations contain information regarding: possible names of a protein; taxonomic characteristics of the biological species from which it was isolated; gene ontology data; amino acid sequence and structural domains of the protein; and also known functions and biological processes in which it participates [15]. Functional classification of proteins that contain AFP_{14-20}-like motifs was performed.

Construction of LDSYQCT peptide analogs. Analogs of the heptapeptide (14)LDSYQCT(20) were constructed using the HyperChem program for molecular modeling [23]. Point substitutions were made on the basis of data on analysis of amino acid substitutions in AFP_{14-20}-like motifs of physiologically active proteins, which are revealed using local alignment as described above and also taking into account differences in physicochemical properties of natural amino acids.

Earlier, we studied the peptides designated as P1 to P12 [11, 12]. The analogs studied in this work are designated as P13 to P25, and the initial peptide LDSYQCT is designated here, like in previous works, as P5. New analogs were obtained by the following amino acid substitutions: 1) leucine at position 14 – L14H, L14P, and L14T (peptides **P14**, **P22** and **P24**, correspondingly); 2) aspartic acid at position 15 – D15G (peptide **P21**); 3) serine at position 16 – S16A and S16K (peptides **P15** and **P16**, correspondingly); 4) glutamine at position 18, namely, Q18A and Q18R (peptides **P17** and **P20**, correspondingly); and 5) threonine at position 20 – T20R (**P18**) and T20V (**P19**).

Besides, two peptides that contain two or three substitutions simultaneously were studied. One of them is the peptide **P13** with amino acid sequence IMSYICS. It is similar to the heptapeptide AFP_{14-20} and represents the fragment of human AFP located in domain II (a.a. 284-290 in a full-length molecule and a.a. 266-272 in a mature molecule). The other one is the peptide **P23** obtained by double substitution, which is frequently observed in AFP_{14-20}-like motifs of physiologically active proteins: L14P and D15G. Amino acid sequences of the analogs are shown in Table 1.

In all peptides under study acidic amino acid (aspartic and glutamic acids) residues were in deprotonated (negatively charged), and basic amino acid residues (lysine, histidine and arginine) – in protonated (positively charged) form. To escape terminal effects, the *N*- and *C*-ends of the peptides were blocked by acetyl and *N*-methyl groups, correspondingly.

MD simulation protocol. Models of the molecules were studied in full-atomic approximation in the potential force field Amber [24, 25]. Its parameters were completed by experimental data [26], and quantum chemistry calculations performed by the GAMESS program package [27]. The restricted Hartree–Fock method with basis set 6-31GF was applied.

MD calculations were performed at trajectory length of 10 nsec in a periodic box with sizes of 100 × 100 × 100 Å. Temperature was given to be equal to 2000K to accelerate scanning of conformational space by representative point and reaching quasi-ergodicity of the trajectories [28]. Both Berendsen and collision thermostats were used to maintain constant temperature. Effects of the collision thermostat are based on interactions between atoms of the molecular system studied and the equilibrium ensemble of particles with a definite mass at a given temperature according to the elastic collision law [29, 30]. Usage of the collision thermostat provides statistically correct distribution of energy on freedom degrees [31]. Masses of thermostat particles were given to be equal to 18 Da and the mean frequency of collisions was equal to 55 $psec^{-1}$. Viscosity of the collision medium is close to that of water at normal conditions. Dielectric constant ε was equal to 1. The Verlet algorithm was used for numerical integration. Initial atomic velocities were chosen by a random number generator with Maxwell distribution.

Treatment of trajectories obtained by MD simulation. Rotations around single covalent bonds *N*–Cα, Cα–*C′* and Cα–*C*β in a polypeptide chain are described using dihedral angles φ, ψ and χ, correspondingly [32]. Rotation potentials of angles φ and ψ are characterized by a low energy barrier (about 1 kcal/mol), which provides almost free rotation around *N*–Cα and Cα–*C′* bonds. So, changes of angles φ and ψ make the main contribution to flexibility of a polypeptide chain.

2D and 3D maps of free energy levels (Poincare sections) were used to estimate probability densities of conformation realization by changing of the dihedral angles. The probability densities were determined with fixed values of two or three angular variables and averaging of other variables. The 2D and 3D distributions of probability densities of conformation realization for all the angular combinations were calculated according to formulae (1):

$$p(\alpha_n,\alpha_m,\ldots) = \int_{-\pi}^{\pi}\ldots\int_{-\pi}^{\pi} p(\alpha_1,\ldots,\alpha_i,\ldots,\alpha_N)\prod_{\substack{i=1\\ i\neq n,m\ldots}}^{N} d\alpha_i \,. \qquad (1)$$

Here α_n,α_m is a set of dynamic variables, $p(\alpha_1, \ldots, \alpha_i, \ldots\alpha_N)$ is a density of probability to discover the system at a given point of conformational space.

Dynamic behavior of an individual conformational degree of freedom of the molecule was evaluated using autocorrelation functions of dihedral angles of special types. Normalized autocorrelation functions were calculated according to formulae (2):

$$F_{xx} = \left\langle e^{i\varphi(t)}e^{-i\varphi(t+\tau)}\right\rangle - \left|\left\langle e^{i\varphi(t+\tau)}\right\rangle\right|^2. \qquad (2)$$

Here φ is dihedral angle value at times of t and $t + \tau$. Information for analysis is contained in the plot of dependence of the real part of the autocorrelation function (F_{xx}) on time. Autocorrelation functions allow to judge about characteristic times of motion and types of dynamic behaviour of dihedral angles.

RESULTS

Analysis of amino acid substitutions in AFP_{14-20}-like motifs. Comparison of two programs (FASTA and BLAST) for local alignment showed that the GLFASTA algorithm is the best suited for goals of the present work because it reveals full-length heptapeptide motifs. On the contrary, the Blastp and PSI-BLAST algorithms provide results as truncated peptide motifs. Structural and functional characterization of proteins that contain AFP_{14-20}-like motifs performed using the annotated UniProtKB/Swiss-Prot knowledgebase allowed classification of the proteins on the basis of their functions. Numbers of proteins that belong to a certain functional class were calculated as the percentage of the total number of proteins extracted from databases using the GLFASTA program.

Proteins that contain AFP_{14-20}-like motifs were determined to belong to the following functional classes: a) transcription factors (22% of proteins extracted); b) enzymes, predominantly oxidoreductases, and enzymes participating in biosynthesis of proteins and nucleic acids (40%); c) cell adhesion proteins and their receptors (14%); d) growth factors, their receptors, and intracellular effectors (10%); e) proteins that participate in transport and intracellular localization of other proteins (9%); f) anti- and proapoptotic proteins (5%). Interestingly, in proteins that contain EGF-like domains and their repeats (such as cell adhesion proteins and Notch proteins) AFP_{14-20}-like motifs were revealed mainly in these domains.

Moreover, the majority of the revealed proteins participate in regulation of embryonic and tumor development. For example, transcription factors regulate processes of cell proliferation, differentiation, and apoptosis, participate in regulation of the cell cycle, vascular-

Table 1. Analogs of the peptide LDSYQCT – biologically active site of human α-fetoprotein (a.a. 14-20)

Designation of peptides	Substitutions of amino acid residues	Amino acid sequences of peptides	Changes in physicochemical properties of amino acid residues at a particular position	Changes in conformational mobility for amino acid residues in analogs
1	2	3	4	5
P5	initial peptide	LDSYQCT	–	–
P13*	L14I	IMSYICS	hydrophobic aliphatic/ hydrophobic aliphatic	decreases for I14 and I18
	D15M		hydrophilic negatively charged/ hydrophobic aliphatic	increases for S16 and S20
	Q18I		hydrophilic uncharged/ hydrophobic aliphatic	
	T20S		hydrophilic uncharged/ hydrophilic uncharged	
P14	L14H	HDSYQCT	hydrophobic aliphatic/ hydrophilic positively charged	decreases at position of substitution (H14) and increases for S16
P15	S16A	LDAYQCT	hydrophilic uncharged/ hydrophobic aliphatic	changes (increases) only at position of substitution (A16)
P16	S16K	LDKYQCT	hydrophilic uncharged/ hydrophilic positively charged	decreases for K16, for other amino acid residues does not sufficiently change
P17	Q18A	LDSYACT	hydrophilic uncharged/ hydrophobic aliphatic	increases for S16 and at position of substitution (A18)
P18	T20R	LDSYQCR	hydrophilic uncharged/ hydrophilic positively charged	slightly increases for L14 and S16, does not change for other amino acid residues
P19	T20V	LDSYQCV	hydrophilic uncharged/ hydrophobic aliphatic	decreases only at position of substitution (V20)
P20	Q18R	LDSYRCT	hydrophilic uncharged/ hydrophilic positively charged	slightly increases for L14, does not change for others
P21	D15G	LGSYQCT	hydrophilic negatively charged/ lack of side chain	increases for all amino acid residues
P22	L14P	PDSYQCT	hydrophobic aliphatic/ hydrophobic cyclized	restricted for P14, does not change for other amino acid residues
P23	L14P,	PGSYQCT	hydrophobic aliphatic/ hydrophobic cyclized	restricted for P14 and increases for other amino acid residues
	D15G		hydrophilic negatively charged/ lack of side chain	
P24	L14T	TDSYQCT	hydrophobic aliphatic/ hydrophilic uncharged	does not change for all amino acid residues

Table 1. (Contd.)

1	2	3	4	5
P25**	S16K,	LDKYACN	hydrophilic uncharged/ hydrophilic positively charged	increases for L14 and A18
	Q18A,		hydrophilic uncharged/ hydrophobic aliphatic	decreases for K16, does not change for N20
	T20N		hydrophilic uncharged/ hydrophilic uncharged	

Notes: a.a., amino acid residue; numeration for amino acid residues is given for the mature polypeptide chain.
* **P13** is an analog of LDSYQCT peptide located in domain II of human AFP (a.a. 266-272).
** **P25** is a part (a.a. 26-32) of the receptor-binding site of epidermal growth factor (EGF).

ization, and morphogenesis, and provide for correct body formation and segmentation. Enzymes that contain AFP_{14-20}-like motifs are predominantly oxidoreductases or participate in synthesis of nucleic acids and proteins. Oxidoreductases may prevent damage that takes place in response to oxidative stress during embryonic development. Synthesis of nucleic acids and proteins is crucial in cell division and differentiation. Table 2 contains the most characteristic representatives of each functional class of AFP_{14-20}-like motif containing proteins.

We also analyzed amino acid substitutions in AFP_{14-20}-like motifs of 300 proteins extracted by local alignment using the GLFASTA algorithm (Table 3). Cysteine, tyrosine, and aspartic acid residues were shown to be the most conserved in AFP_{14-20}-like motifs. They are very or relatively rarely subjected to substitutions. The cysteine residue is conserved in 99.0% of 300 AFP_{14-20}-like motifs extracted, it being substituted only in three cases. The tyrosine residue is preserved in 78.3% of the motifs and is substituted predominantly by phenylalanine (about 6%) and serine (about 4%) residues. Seemingly, at this position an aromatic benzene ring and/or hydroxyl group is of importance. The aspartic acid residue is preserved in 77.7% of the motifs and is most often substituted to glutamic acid (7%) or lysine (6%). This indicates the important role of a charged (predominantly negatively charged) functional group at this position.

Leucine (64%), serine (53%), and glutamine (49%) residues are less conserved in the extracted AFP_{14-20}-like motifs. Leucine is the most often substituted by hydrophobic amino acid residues – isoleucine, valine, phenylalanine, and methionine. Serine is the most often substituted by threonine and lysine residues and also to negatively charged amino acid residues (D and E) and their amides (N and Q). In general, amino acid residues that are able to participate in formation of hydrogen bonds (S, T, Q, N, K, R and H) comprise 79% of the amino acid residues at this position. The glutamine residue is relatively variable and is subjected to substitutions, mainly to hydrophilic and charged (both negatively and positively) amino acid residues. However, amino acid residues that participate in formation of hydrogen bonds are predominantly (81%) located at this position.

The C-terminal residue (threonine) is the most variable one in AFP_{14-20}-like motifs and can be substituted by any amino acid residue except H and W. Among the amino acid residues to which it is mostly substituted are those that sufficiently differ from each other by their physicochemical properties (serine, asparagine, glutamic acid, valine, etc.). So, it can be concluded that physicochemical properties of the amino acid residue at this position are not important.

Analysis of 2D and 3D maps of free energy levels (Poincare sections). Molecules of proteins and peptides represent a cooperative system composed of amino acid residues in which energy of rotation around one covalent bond depends on rotation around neighboring bonds [33]. However, due to peculiarities of the peptide bond (sp^2-hybridization and *trans*-conformation) that create restrictions for rotation around it, two neighboring amino acid residues may change conformation almost independently, while rotations around covalent bonds inside the same residue are interconnected. Potential energy changes during rotation around neighboring covalent bonds allow determining sterically permitted or prohibited combinations of dihedral angles that are used to describe these rotations. Charting of potential energy provides the conformational map that contains areas of permitted or prohibited values of the dihedral angles plotted on the absciss and ordinate axes.

The Ramachandran map (map of potential energy levels in coordinates (φ, ψ)) is usually used to evaluate conformational changes in molecules of proteins and peptides [32]. Unlike the Ramachandran map, two-dimensional (2D) and three-dimensional (3D) maps of free energy level that we use in this work take into account contribution of entropy factor into stabilization of conformations. The 2D and 3D maps demonstrate probabil-

Table 2. Functional classes of physiologically active proteins that contain AFP_{14-20}-like motifs

Protein name and polypeptide chain length	Amino acid sequence of AFP_{14-20}-like motif	Numbers of amino acid residues of AFP_{14-20}-like motifs (biological species)	Function of protein	Accession code in UniProtKB/ Swiss-Prot database
1	2	3	4	5
		Transcription factors		
Yemanuclein-α (1002 a.a.)	LDDYQCT	846-851 *D. melanog.*	DNA-binding protein	P25992
Nucleus accumbens-associated protein 1 (514 a.a.)	LDSVQCT	172-178 mice	repression of transcription	Q7TSZ8
Host range factor (218 a.a.)	VDSYKCT	14-20 *Alpha-bacilovirus*	regulation of virus replication in cell lines	Q90165
Neurogenic locus notch homolog protein 1 (2531 a.a.)	VDSYTCT	964-970 human	transcription activation; regulation of cell proliferation, differentiation and apoptosis	P46531
Neurogenic locus notch homolog protein 4 (1964 a.a.)	LGSYQCL	212-218 mice	negative regulation of cell differentiation	P31695
Transcription elongation factor B polypeptide 3. Elongin-A (434 a.a.)	ADTYQCV	42-48 *C. elegans*	factor of transcription elongation	Q09413
Zinc finger homeobox protein 3 (3726 a.a.)	GDSYQCK	982-988 mice	repression of transcription; muscle formation	Q61329
		Enzymes		
Genome polyprotein (3140 a.a.)	LDNYKCI	1333-1339 *Plum pox potyvirus*	RNA-polymerase, helicase	Q84934
Ubiquitin-like modifier-activating enzyme (1052 a.a.)	LDKYQCV LEKYQCV	151-157	modification (ubiquitination) of proteins	A0AVT1 (human) Q8C7R4 (mice)
Stress response protein yHaX (288 a.a.)	LESYQCN	96-102 *Bacillus subtilis*	hydrolase; stress-response	O07539
Anaredoxin (76 a.a.)	LESYQCM	19-25 *Nostoc sp.*	endonuclease, oxidoreductase	Q44141
Phenylalanyl-tRNA synthetase, β-chain (799 a.a.)	LDSYTCL LDPYSCS	365-371 769-775 *Leptospira biflexa*	amino acyl t-RNA-synthase; protein synthesis	B0SAR6
NADH-quinone oxidoreductase, subunit A (118 a.a.)	LDTYECG	39-45 *Roseiflexus castenholzii*	oxidation-reduction reactions	A7NL04
CTP synthase; UTP-ammonia ligase (533 a.a.)	LGSYQCR	433-439 *Methanobacter thermoautotr.*	pyrimidine nucleotide biosynthesis	O26519
Dicer-like protein 1, ribonuclease (1548 a.a.)	LDAYKCD	1525-1531 *Cryphonectria parasitica*	RNA processing, splitting of RNA to RNAi; antiviral protection	Q2VF19

Table 2. (Contd.)

1	2	3	4	5
Cytosine-specific methylase HgiCI (420 a.a.)	LDKYNCS	212-218 *Herpetosiphon aurantiacus*	DNA methylation on cytosine, protection from endonuclease	P25263
DNA repair protein mutS (844 a.a.)	LDKYRCV	420-426 *Streptoc. pneumonia*	reparation of DNA	C1CTY2
Cell-adhesion proteins and their receptors				
Laminin, α-2 subunit (3122 a.a.)	LDDYRCT	1489-1495 human	regulation of cell adhesion; contains 17 EGF-like domains	P24043
Transmembrane cell adhesion receptor mu (3767 a.a.)	MDSYECD	2282-2288 *C. elegans*	regulation of cell adhesion and growth; contains 52 EGF-like domains	P34576
Protocadherin Fat 4, tumor suppressor Fat (4981 a.a.)	LDSFHCS	3646-3652 mice 3644-3650 human	regulation of cell adhesion and polarization; contains 6 EGF-like domains	Q2PZL6 Q6V0I7
Delta-like protein 1; Drosophila Delta homolog 1 (714 a.a.)	EDSYSCT	346-352 rat	cell adhesion regulation; contains 8 EGF-like domains	P97677
Agrin (2073 a.a.)	LESYECA	1853-1859 chicken	discovered in embryonic brain and in tumor tissues	P31696
Growth factors, their receptors and possible intracellular effectors				
Pro-epidermal growth factor (1217 a.a.)	LDSYTCN	1002-1008 mice	stimulation of epithelial and endothelial cell growth	P01132
Cell surface receptor daf-1; abnormal dauer formation protein 1 (669 a.a.)	FESYQCA	484-490 *C. elegans*	stress-response; member of TGF-β family	P20792
Dorsal-ventral patterning protein tolloid (1067 a.a.)	LGSYQCG	609-615 *D. melanog.*	TGF-β-mediated signaling	P25723
Transport and localization of proteins				
Vacuolar protein sorting associated protein 13C (3748 a.a.)	LDSYRCQ	2578-2584 mice 2583-2589 human	protein localization	Q8BX70 Q709C8
Autophagy-related protein 7 (603 a.a.)	LDDYKCV	565-571 *S. cerevisiae*	ubiquitin-associated transportation of proteins	Q6CXW3
Dynamin-binding protein; scaffold protein (1577 a.a.)	LDFYNCT	1129-1135 human 1132-1138 mice	protein–protein interactions; transportation of proteins from Golgi apparatus to cell membrane	Q6XZF7
Anti- and pro-apoptotic proteins				
Protein B, homolog of protein fem-1 (627 a.a.)	MDNYECN	437-443 rat 437-443 human	death receptor associated protein; mediates apoptosis	P0C6P7 Q9UK73

Table 3. Calculations of amino acid substitutions in AFP_{14-20}-like motifs of physiologically active proteins

Peptide / All amino acid residues	L (14)	D (15)	S (16)	Y (17)	Q (18)	C (19)	T (20)
G	2 (0.67%)	6 (2.00%)	10 (3.33%)	0	6 (2.00%)	0	7 (2.33%)
A	2 (0.67%)	3 (1.00%)	4 (1.33%)	4 (1.33%)	2 (0.67%)	0	10 (3.33%)
L	192 (64.0%)	0	7 (2.33%)	2 (0.67%)	3 (1.00%)	0	9 (3.00%)
I	38 (12.67%)	0	3 (1.00%)	3 (1.00%)	0	0	11 (3.67%)
V	15 (5.00%)	0	5 (1.67%)	4 (1.33%)	8 (2.67%)	0	23 (7.67%)
P	0	3 (1.00%)	3 (1.00%)	1 (0.33%)	3 (1.00%)	0	3 (1.00%)
M	12 (4.00%)	0	0	1 (0.33%)	0	0	2 (0.67%)
C	2 (0.67%)	0	1 (0.33%)	1 (0.33%)	1 (0.33%)	297 (99.0%)	1 (0.33%)
S	0	1 (0.33%)	159 (53.0%)	11 (3.67%)	14 (4.67%)	1 (0.33%)	35 (11.67%)
T	0	3 (1.00%)	35 (11.67%)	0	9 (3.00%)	1 (0.33%)	98 (32.67%)
F	13 (4.33%)	0	1 (0.33%)	20 (6.67%)	0	0	3 (1.00%)
Y	1 (0.33%)	0	1 (0.33%)	235 (78.3%)	2 (0.67%)	0	9 (3.00%)
D	1 (0.33%)	233 (77.7%)	17 (5.67%)	0	7 (2.33%)	0	16 (5.33%)
N	0	9 (3.00%)	14 (4.67%)	0	16 (5.33%)	0	32 (10.67%)
E	3 (1.00%)	20 (6.67%)	13 (4.33%)	4 (1.33%)	27 (9.00%)	0	30 (10.0%)
Q	8 (2.67%)	4 (1.33%)	10 (3.33%)	3 (1.00%)	148 (49.3%)	1 (0.33%)	5 (1.67%)
K	3 (1.00%)	18 (6.00%)	11 (3.67%)	3 (1.00%)	23 (7.67%)	0	4 (1.33%)
R	4 (1.33%)	0	6 (2.00%)	5 (1.67%)	22 (7.33%)	0	5 (1.67%)
H	0	1 (0.33%)	2 (0.67%)	3 (1.00%)	12 (4.0%)	0	0
W	4 (1.33%)	0	0	2 (0.67%)	0	0	0

Note: Percentage of substitutions of amino acid residues is given in brackets.

ity densities of conformation with change in values of dihedral angles φ, ψ, and χ. The darkest areas on the map correspond to minima of free energy level with highest probability density of conformation.

Figure 1 shows 2D maps of free energy changes for amino acid residues in the heptapeptide LDSYQCT (peptide **P5**) and its 10 analogs (peptides **P14**-**P22**, **P24**) and also in three more peptides of which the first is the fragment of human AFP domain II (a.a. 266-272 in the mature molecule) with sequence IMSYICS (**P13**), the second being the peptide **P23** obtained by double amino acid substitution L14P and D15G, and the third being the peptide fragment with sequence LDKYACN (**P25**) which is the part (a.a. 26-32) of receptor-binding site of epidermal growth factor (EGF).

Let us first consider 2D maps of free energy change for amino acid residues in the peptides under study in coordinates (φ, ψ). Leucine residue (L14) in the maternal peptide **P5** shows an extended contour along the ψ-coordinate with two local minima of free energy (Fig. 1a). One deep minimum is located at φ = –135, ψ = –30 (on the edge of right-handed 3_{10} helix), and the other (not deep) at φ = –135, ψ = 150 (corresponds to β-structure). Thus, residue L14 is characterized by a relatively limited set of probable conformations, and this is confirmed by the 3D map of this residue (Fig. 2a).

The 2D map of the serine residue (S16) in peptide **P5** shows only one deep minimum at φ = –60, ψ = –30 (corresponds to right-handed α-helix). Additional loci arise in 2D maps of the glutamine (Q18) and threonine (T20)

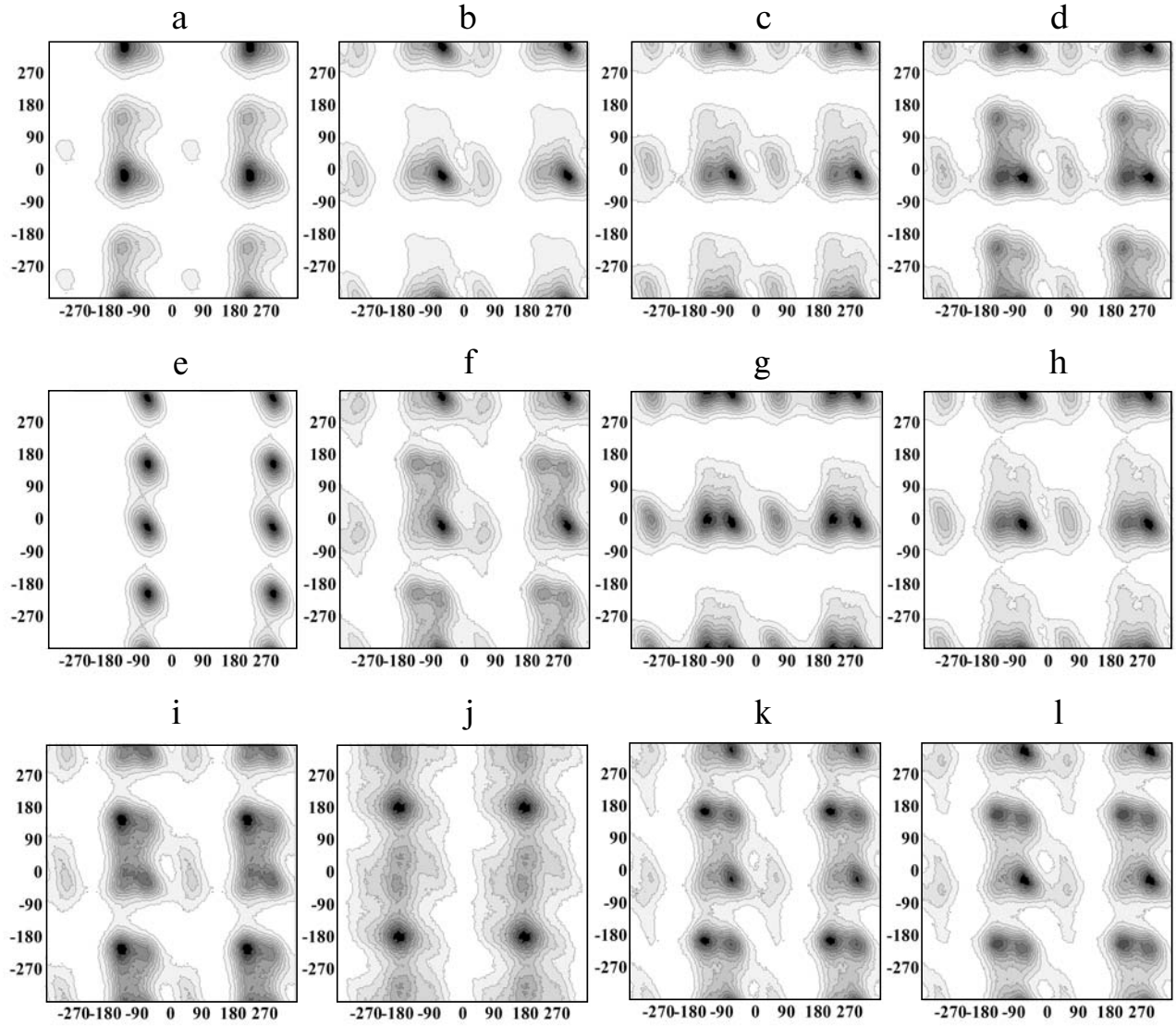

Fig. 1. Two-dimensional (2D) maps of free energy levels (Poincare sections) in coordinates (φ, ψ) for amino acid residues L14 (a, i), P14 (e), G15 (j), S16 (b, f, k), A16 (g), Q18 (c), A18 (h), T20 (d), and S20 (l) in the heptapeptide AFP_{14-20} (**P5**) (a-d) and its analogs **P22** (e), **P14** (f), **P15** (g), **P17** (h), **P21** (i, j, k), and **P13** (l).

residues. These are two loci at φ = 60 and ψ = 30 (left-handed 3_{10} helix) and at φ = –135, ψ = –30 (right-handed 3_{10} helix) for Q18; and one locus at φ = –150, ψ = 150 (corresponds to β-structure) for T20. In this case the areas of free energy minima for T20 are linked by a band (Fig. 1, b-d), and this indicates existence of correlated changes in dihedral angles that have collective degree of freedom. The increased set of probable conformations for T20 in comparison to other residues (including S16) is confirmed by the 3D maps of this residue (Fig. 2d).

In three analogs of peptide AFP_{14-20} – **P14**, **P22**, and **P24** – the leucine residue at position 14 was substituted by histidine, proline, or threonine (L14H, L14P, and L14T, respectively). One more peptide (**P23**) contains simultaneously two substitutions – L14P and D15G.

In peptide **P14** the mobility of the residue located at the substituted position (H14) is decreased. Thus, here the histidine residue demonstrates limited mobility in comparison with that in the maternal peptide **P5** (Fig. 2e). However, 2D map of the serine residue (S16) in peptide **P14** demonstrates an increased set of probable conformations in comparison with S16 in the maternal peptide **P5** (Fig. 1f). This may be a result of the presence of two neighboring oppositely charged amino acid residues – H14 and D15. Seemingly, electrostatic attraction of aspartic acid by histidine decreases steric limitations for the adjacent serine residue. The pattern of distribution of free energy levels for other residues in peptide **P14** does not significantly change in comparison with peptide **P5**.

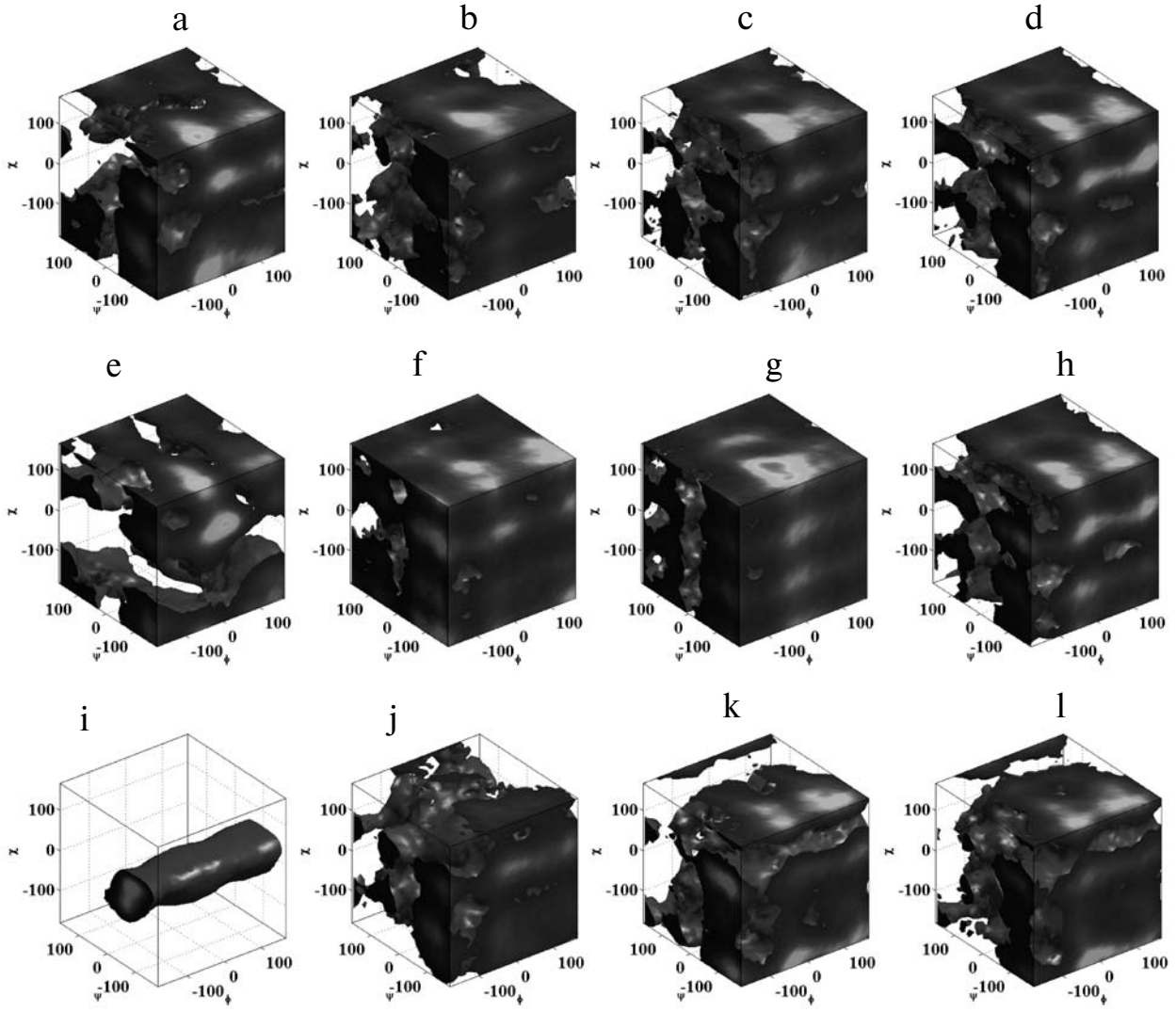

Fig. 2. Three-dimensional (3D) maps of free energy levels (Poincare sections) for amino acid residues L14 (a, e), P14 (i), S16 (b, f), K16 (j), Q18 (c), A18 (g), I18 (k), T20 (d, h), and V20 (l) in the heptapeptide **P5** (a-d) and its analogs **P14** (e), **P22** (i), **P23** (f), **P24** (h), **P16** (j), **P18** (g), **P13** (k), and **P19** (l).

Analog **P22** differs from the maternal peptide by the presence at position 14 of proline residue instead of leucine (substitution L14P). The proline residue in this peptide demonstrates significant limitation in conformational changes (Fig. 1e): the region of free energy minimum in its 2D map shows a narrow contour with two loci, i.e. at $\varphi = -60$, $\psi = -30$ (right-handed α-helix) and at $\varphi = -60$, $\psi = 150$ (polyproline helix). Restrictions in proline residue motion are also confirmed by its 3D map (Fig. 2i). However, 2D and 3D maps of residues S16, Q18, and T20 practically do not change in comparison with the corresponding maps of these residues in the maternal peptide **P5**.

The presence of the glycine residue that follows the proline residue in peptide **P23** (Table 1) does not change the 2D and 3D maps (Figs. 1e and 2i) of the latter in comparison with those of analog **P22** (except decreasing the depth of the locus at $\varphi = -60$, $\psi = 150$). However, serine residue S16 demonstrates an increased set of probable conformations—its 2D and 3D maps (Fig. 2f) are similar to those for S16 in peptide **P14**. This is a result of absence of a side chain in the neighboring glycine residue at position 15, and this decreases steric restrictions for S16. Conformational mobility of Q18 is increased in comparison with that in the maternal peptide **P5**. However, the behavior of the threonine residue T20 is not sufficiently altered in comparison with that in peptide **P5**.

Substitution of leucine by threonine (L14T) was made in analog **P24**. This does not significantly change the pattern of distribution of free energy levels for T14 in comparison with L14 in the maternal peptide **P5**. The 2D and 3D maps of other residues are also not significantly

changed in comparison with those in peptide **P5**. Interestingly, there are two threonine residues (T14 and T20) in peptide **P24** with larger set of probable conformations for T20 (Fig. 2h). This may be explained by the presence in the neighborhood of T14 of aspartic acid residue D15 that decreases the set of probable conformations for T14.

To prepare analog **P21** the aspartic acid residue at position 15 was substituted by glycine (D15G). This substitution leads to remarkable changes in the 2D and 3D maps of free energy levels for all residues in this peptide. For residue L14 (Fig. 1i) that precedes glycine residue G15, the set of probable conformations are increased with shifting of locus of free energy minimum to $\varphi = -150$, $\psi = 150$ (corresponds to β-structure). The 2D map of G15 is almost fully occupied, and this shows the existence of a large set of probable conformations that is characteristic for this amino acid with no side chain (Fig. 1j). Substitution D15G leads also to increased set of probable conformations for residue S16 (Fig. 1k) in analog **P21** in comparison with S16 in the maternal peptide LDSYQCT (**P5**). Sets of probable conformations are also increased for residues Q18 and T20 (although to less extent) in **P21** in comparison with those in **P5**.

The serine residue at position 16 was subjected to substitutions S16A and S16K (peptides **P15** and **P16**, respectively). The substitution S16A in peptide **P15** alters the contour of free energy minimum for A16 and gives rise to deep new loci (Fig. 1g) at $\varphi = -60$, $\psi = -30$ and $\varphi = -135$, $\psi = -30$, which are linked by isthmus that suggests the possibility of conformational transitions. Moreover, increasing conformational mobility of A16 is observed in 3D maps. The sets of probable conformations are not notably changed for the other amino acid residues.

Substitution S16K (peptide **P16**) changes the 2D and 3D maps of amino acid residues predominantly at the position of substitution. Appearance of an additional deep locus on the 2D map of leucine residue L14 at $\varphi = -60$, $\psi = -30$ (corresponds to right-handed α-helix) is notable, although the 3D map of this residue does not show significant changes in comparison with that in peptide **P5**. Lysine residue K16 shows restrictions in conformational mobility in comparison with S16 in the maternal peptide **P5**, and this is confirmed by its 3D map (Fig. 2j). This circumstance may be explained by the existence of the large side chain in the lysine residue. Besides, K16 slightly increases conformational mobility of glutamine residue Q18. This may be explained by electrostatic attraction between neighboring oppositely charged residues D15 and K16 in peptide **P16**, as well as between H14 and D15 in peptide **P14**.

The **P17** and **P20** analogs differ from the maternal peptide by point substitutions of glutamine residue at position 18 to alanine and arginine (Q18A and Q18R, respectively). In peptide **P17** the substitution Q18A does not cause significant changes in the sets of probable conformations for residues L14 and T20 but significantly increases that for serine (S16) and alanine (A18) residues (Fig. 2g). Comparison of the two alanine residues – A16 in peptide **P15** (LDAYQCT) and A18 in peptide **P17** (LDSYACT) – shows that conformational mobility of A16 (Fig. 1g) is smaller than that of A18 (Fig. 1h). This might be explained by steric constraints for A16 located between two amino acid residues (D15 and Y17) with large side chains.

Increasing depth of the locus at $\varphi = -135$, $\psi = 150$ (corresponds to β-structure) for residue L14 in peptide **P20** that contains the substitution Q18R was observed. However, conformational mobility of residues S16 and T20 and, more interestingly, of the residue at the position of substitution (R18) is not significantly changed. Such a behavior of the residues in peptides **P17** and **P20** may be explained by formation of hydrogen bonds between amino acid residues S16 and Q18 in the maternal peptide **P5**. Substitution of Q18A leads to disruption of this interaction and, consequently, to increasing of mobility of S16 in peptide **P17**. Arginine residue R18 in peptide **P20** may also participate in formation of hydrogen bond with S16, and as a result the conformational behavior of residues at position 18 in the maternal peptide **P5** and its analog **P20** is similar. Thus, point substitutions made at position 18 suggest functional significance of residue Q18 – participation in formation of intramolecular hydrogen bonds (most probably with S16).

Analogs **P18** and **P19** were obtained by point substitution of the threonine residue at position 20 to arginine or valine – T20R and T20V, respectively. In peptide **P18** the set of probable conformations is slightly increased for residues L14 and S16 with increasing depth of the locus at $\varphi = -135$, $\psi = 150$ (corresponds to β-structure) for L14, and this is also confirmed by 3D maps of these residues. Conformational mobility is not changed for residues Q18 and R20.

In peptide **P19** amino acid residues L14, S16, and Q18 demonstrate the same pattern of free energy levels as in the initial peptide. The valine residue at position 20 (V20) is characterized by a smaller set of probable conformations than that for T20, and this is explained by existence of a large side chain (Fig. 2l). Absence of notable changes in conformational and dynamic properties of amino acid residues in peptides **P18** and **P19** (except for the position of substitution) shows that threonine (T20) in the maternal peptide LDSYQCT seemingly, does not participate in intramolecular interactions.

We have also studied two peptides that represent fragments of natural proteins – domain II of AFP (IMSYICS, peptide **P13**) and a part of the receptor-binding site of EGF (LDKYACN, peptide **P25**). The first peptide has two isoleucine and two serine residues. Change in location of energy minima and increasing depth of the locus at $\varphi = -135$, $\psi = 150$ are observed in 2D maps of

the isoleucine residue at position 14 (I14) in comparison with that of L14 in the maternal peptide **P5**. However, the 3D map of I14 shows, on the contrary, a slight decrease in its conformational mobility in comparison with L14. Another isoleucine residue is located at position 18 instead of glutamine (substitution Q18I). The difference between Q18 and I18 is the presence of a deep locus at $\varphi = -150$, $\psi = 150$ for the latter. However, decrease in the set of probable conformations is observed in the 3D map of I18 (Fig. 2k) in comparison with Q18 in the initial peptide **P5**. Comparison of the two isoleucine residues in peptide **P13** shows that there is no significant difference in their conformational mobility.

Two deep new loci at $\varphi = -60$, $\psi = 150$ and $\varphi = -135$, $\psi = 150$ appear in 2D maps of serine residue S16 in peptide **P13** in comparison with S16 in peptide **P5**. Increase in conformational mobility of S16 in peptide **P13** is observed in the 2D maps in coordinates (ψ,χ) and in the 3D maps of this residue. This may be explained by lack of intramolecular interactions in which S16 could participate, as it takes place in the peptide **P5** (formation of hydrogen bonds with Q18). Also *C*-terminal amino acid in peptide **P13** is serine (S20), but not threonine (T20) as in peptide **P5**. The residue S20 is characterized by an increase in the set of probable conformations in comparison with T20, that is, in general, characteristic for the serine residue (Fig. 1l). This is observed in the 2D maps in coordinates (φ,ψ) and (ψ,χ) as well as in the 3D maps for this residue. Comparison of the two serine residues in peptide **P13** shows that there is no significant difference in their conformational mobility, despite the difference in the microenvironment (neighborhood to different amino acid residues) and location at *C*-terminus as for S20.

The EGF-derived peptide LDKYACN (**P25**) differs from peptide **P5** by three substitutions: S16K, Q18A, and T20N. This triple substitution causes the following changes in conformational properties of amino acid residues in comparison with the peptide **P5**. The 2D map of L14 in peptide **P25** has along with the locus of right-handed 3_{10} helix a deep new locus at $\varphi = -60$, $\psi = -30$, and this indicates the possibility to form right-handed α-helix. The 3D map of this residue also demonstrates a slight increase in conformational mobility in comparison with that in peptide **P5**. The lysine residue K16 is characterized by a decrease in conformational mobility in comparison with S16. The 2D and 3D maps of residue K16 in two peptides (**P25** and **P16**) containing the lysine residue immediately after the aspartic acid residue (D15) demonstrate significant similarity in probability densities of conformation. This may be explained by mutual impact and possible electrostatic attraction between oppositely charged side chains of lysine and aspartic acid residues in both peptides.

Alanine residue A18, on the contrary, is characterized by a significant increase in the set of probable conformations (by appearance in the 2D map of a deep new locus at $\varphi = -150$, $\psi = 150$ and almost full rearrangement of the 3D map) in comparison with Q18 in **P5**. Comparison of the two alanine residues A18 in different peptides – LDSYACT (**P17**) and LDKYACN (**P25**) – shows that the 2D maps (Fig. 1h) of this residue in coordinates (φ,ψ) and (ψ,χ), as well as its 3D map are less occupied in peptide **P17**. This may be explained by electrostatic attraction of D15 to K16 in peptide **P25**, which provides increasing mobility of adjacent residues. Interestingly, conformational properties of residues at position 20 in both peptides (T20 and N20) are similar to those for the threonine residue (T20) in the maternal peptide **P5**.

Analysis of autocorrelation function plots. Autocorrelation functions calculated using the formulae (2) contain information about dynamic parameters of amino acid residues in the peptides under study: 1) characteristic time of quenching (τ), and 2) residual correlation. Characteristic time of quenching of autocorrelation function is the time of conformational transition when a dihedral angle value is changed; the residual correlation contains information about restrictions in conformational mobility of residues in the potential well.

Plots of dependence of real part of autocorrelation function on time for amino acid residues in the maternal peptide **P5** and its analogs are shown in Fig. 3. It is seen that the value of residual correlation depends on the type of a residue at a definite position and its microenvironment and also on type of dihedral angle.

In variation of dihedral angle φ there are relatively high values of residual correlation (from 0.2 to 0.4) at low time of quenching – from 5 to 10 psec in amino acid residues at position 14 (Fig. 3a). The lowest value of residual correlation (0.2) at $\tau = 5$ psec is observed for L14 in peptide **P21**, in which the glycine residue (G15) is located immediately after the leucine residue. This parameter is relatively low (0.25) also for L14 in peptide **P20** that may be explained by electrostatic attraction between D15 and R18. High values of residual correlation (0.35-0.40) are observed for H14 in peptide **P14**, for L14 in peptides **P17**, **P19** and **P5**, and also for T14 in peptide **P24**. Only in the case of the proline residue in peptides **P22** and **P23** the value of residual correlation reaches 0.75 at very low characteristic time of quenching (about 2 psec).

The serine residue at position 16 (S16) in the most peptides (including **P5**) shows values of residual correlation close to 0.15-0.20 at τ = 15-20 psec. Decrease in time of quenching (to 10 psec) is observed for S16 following glycine residue (G15) in peptides **P21** and **P23**. In peptide **P15** (LDAYQCT) there is an alanine residue at position 16 (A16) that shows the smallest value of residual correlation (about 0.1) at $\tau = 30$ psec (Fig. 3b). Peptides **P16** (LDKYQCT) and **P25** (LDKYACN) contain a lysine residue (K16) that shows the greatest value of residual correlation (about 0.35 in **P16** and 0.25 in **P25**); the characteristic time of quenching for K16 in both peptides is 30 psec.

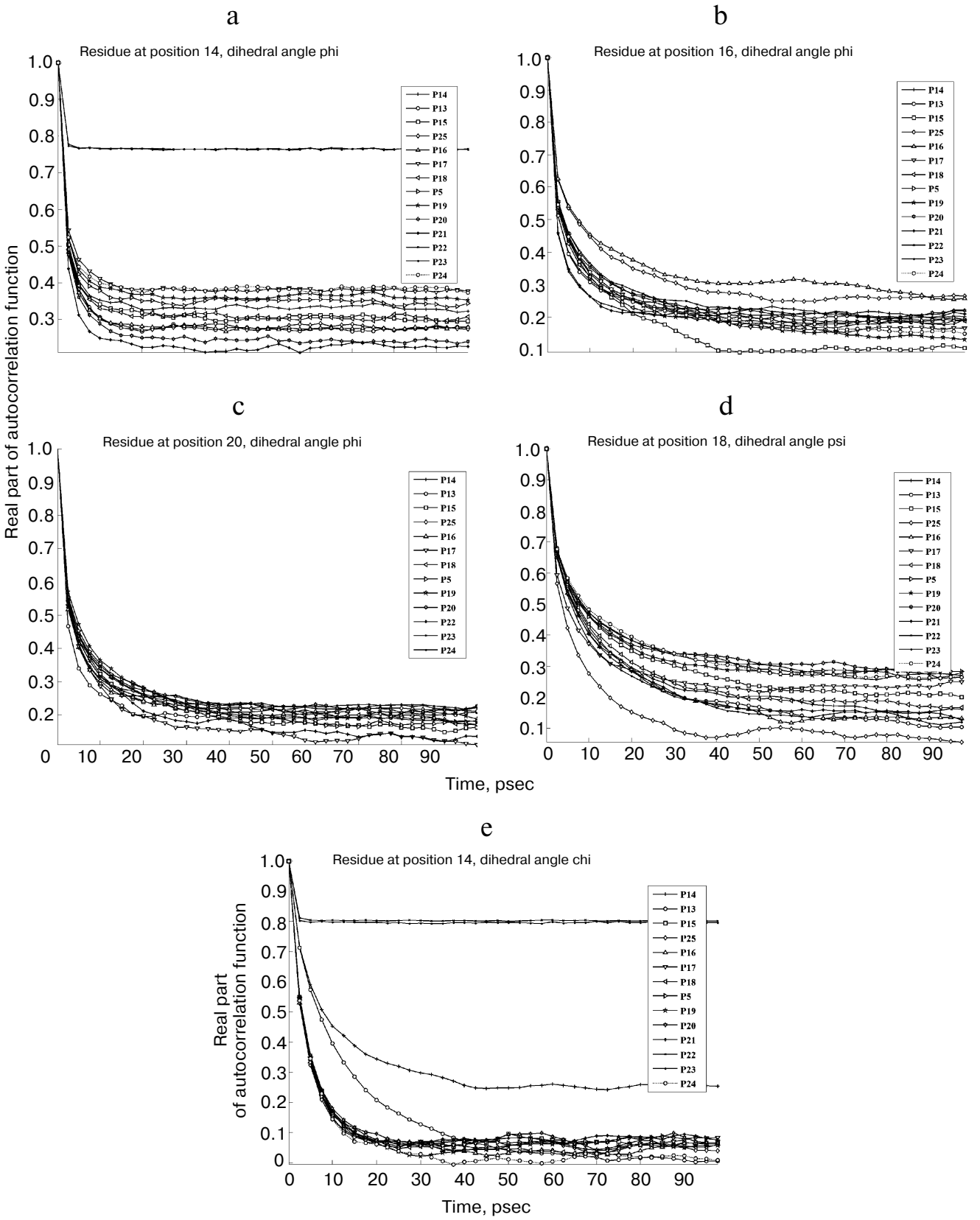

Fig. 3. Time dependence of real part of autocorrelation function for dihedral angles φ (a, b, c), ψ (d), and χ (e) for amino acid residues at positions 14 (a, e), 16 (b), 18 (d) and 20 (c).

At position 18 substitutions were made in peptides **P17** (Q18A), **P20** (Q18R), and **P25** (Q18A); in other peptides the glutamine residue was conserved. Values of residual correlation for amino acid residues at this position 18 vary from 0.15 to 0.35 at τ = 10-20 psec. The greatest value of residual correlation (0.35) is observed for residue I18 in peptide **P13** (IMSYICS), and the smallest (0.15) for residue Q18 in peptide **P23**. The latter may be explained by expansion of influence of glycine residue G15. Low values of dynamic parameters are also observed for alanine A18 in peptides **P17** (LDSYACT) and **P25** (LDKYACN). For amino acid residues Q18 in the maternal peptide **P5** and R18 in its analog **P20** mean values of dynamic parameters (residual correlation of 0.25 at τ = 20 psec) are characteristic.

The substitutions T20S, T20R, T20V and T20N were made at position 20 in peptides **P13**, **P18**, **P19**, and **P25**, respectively. In other peptides the threonine residue was conserved. Despite the differences in physicochemical properties, plots of autocorrelation functions for all amino acid residues at this position demonstrate a narrow range of variations of residual correlation (from 0.10 to 0.25) at τ = 10-20 psec. The smallest residual correlation value is observed for T20 in peptides **P17** and **P22** (Fig. 3c), and the smallest value of τ = 5 psec is for S20 in peptide **P13**. The greatest values of residual correlation (0.25) for this position are observed for T20 in peptides LDSYRCT (**P20**) and PGSYQCT (**P23**).

In variation of dihedral angle ψ for amino acid residues at position 14, histidine residue H14 demonstrates the greatest value of residual correlation (approaching to 0.25) in peptide **P14** (HDSYQCT). Other amino acid residues show values of residual correlation about 0.1-0.2 at τ = 20-30 psec. Only proline residue P14 in peptides **P22** (PDSYQCT) and **P23** (PGSYQCT) demonstrates a characteristic curve with high value of residual correlation, which slowly (with τ = 80 and 40 psec, correspondingly) approaches zero. Interestingly, the glycine residue (G15) decreases characteristic time of quenching for amino acid residues located before it (proline and leucine in peptides **P21** and **P23**, correspondingly).

Amino acid residues at position 16 demonstrate a wide range of residual correlation values with the greatest value (0.4) for A16 in peptide **P15**. The smallest value of residual correlation (close to 0.05) is observed for S16, which is located immediately after glycine residue G15 in peptides **P21** and **P23**, and also for S16 in peptide **P13**. The residual correlation value approaching 0.3 at τ = 30 psec is observed for S16 in maternal peptide **P5**.

For residues at position 18 values of residual correlation vary from 0.1 (for A18 in peptide **P25**) to 0.3 (for Q18 and R18 in peptides **P19** and **P20**, respectively). For the latter two peptides residual correlation values for amino acid residues at this position are comparable with those in maternal peptide **P5** (Fig. 3d). Characteristic time of quenching varies, mainly, from 20 to 30 psec.

Residues at position 20 show residual correlation values from zero (for S20, T20, and N20 in peptides **P13**, **P20**, and **P25**, respectively) to 0.2 (for R20 and V20 in peptides **P18** and **P19**, respectively). Characteristic time of quenching varies from 5 to 15 psec.

In variation of dihedral angle χ residues at position 14 demonstrate a narrow range of residual correlation values (from 0 to 0.1) at a quite low τ (about 10 psec). Exceptions are peptides **P14**, in which H14 show dynamic parameter of 0.25 at τ = 20 psec (Fig. 3e), and peptides **P22** and **P23**, in which the proline residue shows very high value of residual correlation (about 0.8). Also, isoleucine I14 in peptide **P13** shows relatively high value of τ = 20 psec.

Amongst amino acid residues at position 16, alanine residue A16 in peptide **P15** with both dynamic parameters close to zero should be pointed out. In other peptides the serine residue located at this position demonstrates residual correlation value close to 0.05 (in peptide **P5** close to zero) and τ = 5 psec. In peptides **P16** and **P25** there is lysine residue K16 at this position with high values of residual correlation and characteristic time of quenching (0.15-0.20 and 20 psec, respectively).

For amino acid residues at position 18 in most peptides low values of dynamic parameters close to 0.05 at τ = 10 psec are characteristic, with exception of peptides **P13**, **P17**, **P20**, and **P25**. In peptide **P13** isoleucine residue I18 has relatively high τ = 40 psec and low residual correlation value approaching 0.05. In peptide **P20** amino acid residue R18 shows residual correlation value about 0.1 at τ = 15 psec. In peptides **P17** and **P25** the alanine residue located at this position demonstrates again both dynamic parameters close to zero.

For all amino acid residues at position 20 residual correlation values vary from 0 to 0.1 at τ = 10-15 psec with the exception of S20 in peptide **P13** and V20 in peptide **P19** that have characteristic times of quenching equal to 5 and 30 psec, respectively.

DISCUSSION

The goal of this work was to elucidate the role of individual amino acid residues in intramolecular interactions that arise in heptapeptide AFP_{14-20} by comparing its conformational and dynamic properties with those of the peptide analogs obtained by point amino acid substitutions. The analogs were constructed taking into account not only the differences in physicochemical properties of amino acids (Table 1), but also results of analysis of amino acid substitutions in AFP_{14-20}-like motifs that we revealed in physiologically active proteins using the method of local alignment.

Notably, the overwhelming majority of proteins that were revealed to contain AFP_{14-20}-like motifs participate in regulation of cell proliferation, differentiation, and

apoptosis during embryonic development or tumor growth (Table 2). This makes it possible to suggest that AFP_{14-20}-like heptapeptide motifs represent the functionally important site of regulatory proteins responsible for their participation in determining of cell fate during embryo- and carcinogenesis.

Analysis of amino acid substitutions in the revealed AFP_{14-20}-like motifs showed that residues of aspartic acid, tyrosine and cysteine (at positions 15, 17 and 19, correspondingly) are the most conserved in the heptapeptide LDSYQCT under study (Table 3). The influence of point substitutions of these amino acid residues on conformational and dynamic properties of the heptapeptide AFP_{14-20} were studied and discussed earlier [11, 12]. In the present work we studied other analogs obtained by point amino acid residue substitutions at positions 14, 16, 18, and 20, which are the most variable ones. These are residues of leucine, serine, glutamine, and threonine. Analysis of these substitutions in physiologically active proteins showed predominantly hydrophobic character [34] of amino acid residues at position 14 (94%). Hydrophilic amino acid residues with polar groups able to participate in formation of hydrogen bonds (S, T, Q, N, K, R, and H) dominate at positions 16 and 18 (79 and 81%, respectively). We proposed intramolecular hydrogen bonding between S16 and Q18 and participation of leucine residue L14 in (intermolecular) hydrophobic interactions. To verify this hypothesis, we study conformational and dynamic properties of the heptapeptide AFP_{14-20} analogs obtained by point substitutions of the above mentioned amino acid residues by molecular dynamics simulation.

We showed that in peptide **P24** point substitution of the hydrophobic leucine residue by hydrophilic threonine residue (L14T) does not lead to changes in conformational mobility of any amino acid residue in this peptide. No remarkable differences in conformational and dynamic properties of amino acid residues in peptides **P5** and **P24**, despite the differences in physicochemical properties of threonine and leucine, may be explained by lack of their participation in intramolecular interactions. However, substitution of leucine to histidine (L14H) in peptide **P14** leads to decreasing of conformational mobility of the residue at the position of substitution (H14) and increasing of that for serine (S16). This may be explained by electrostatic attraction between histidine (H14) and aspartic acid (D15) in this analog, and the resulting decrease of steric restrictions for adjacent serine residue S16.

Residues S16 and Q18 in the maternal peptide AFP_{14-20} are characterized by limitation in conformational mobility in comparison with those in peptides in which substitutions of these amino acid residues were made. This may be explained by interaction of these amino acid residues with each other by formation of hydrogen bond. Substitution of glutamine residue to alanine (Q18A) in peptide **P17** leads to disruption of this interaction and, as a consequence, to increasing of mobility of S16. In peptide **P20** arginine residue R18 may also participate in formation of hydrogen bond with S16, and, as a consequence, similarity of conformational behavior of amino acid residues at position 18 in the initial peptide **P5** and its analog **P20** is observed.

Increasing of conformational mobility of S16 is well demonstrated by the 2D and 3D maps of this residue in peptide **P13**, in which the glutamine residue is substituted to isoleucine (Q18I). This may also be explained by lack of intramolecular interactions, in which S16 could participate, as occurs in the initial peptide **P5**. Substitution of hydrophilic serine residue to lysine with large [35] positively charged side chain (S16K) in peptide **P16** showed that conformational mobility of the lysine is remarkably lower than that of serine. However, conformational mobility of glutamine residue Q18 in this peptide is not changed, and this indicates that a hydrogen bond between K16 and Q18 is preserved.

Analysis of the 2D and 3D maps of free energy levels showed that the greatest changes in the peptides under study are induced by substitution of any amino acid residue to glycine. In peptides **P21** and **P23**, in which the substitution D15G was made, changes in conformational and dynamic properties of almost all residues are observed. Here glycine residue (G15) decreases characteristic time of quenching for residues located before it (proline and leucine in peptides **P21** and **P23**, respectively). It was also shown that a proline residue located at *N*-terminus of the peptides does not significantly influence conformational and dynamic properties of the residues following it.

Results obtained by analysis of the 2D and 3D maps of free energy levels are confirmed by plots of autocorrelation functions that show dependence of values of dihedral angles φ, ψ, and χ on time. For example, plots of autocorrelation functions for dihedral angle φ show that values of residual correlation for leucine (L14) are decreased, i.e. conformational mobility is increased in peptides with electrostatic attraction between side chains of residues located after L14 – between D15 and K16 in peptides **P16** and **P25**, and also between D15 and R18 in peptide **P20**.

The greatest conformational mobility (except glycine) is characteristic for alanine residue, independently of the position in the polypeptide chain, as it may be seen in peptides **P15**, **P17**, and **P25**. This is provided by the small side chain with the lowest reactivity in alanine [35]. As for the majority of other amino acid residues their conformational mobility may sufficiently vary depending on the residue position in the polypeptide chain and the amino acid sequence of the peptide. For example, values of residual correlation for serine residue at position 16 (S16) with change in dihedral angle φ may vary from 0.05 to 0.30 in peptides with different amino acid sequences.

In peptides **P21** and **P23** serine residue S16 that follows glycine G15 shows the lowest value of residual correlation (close to 0.05), and, consequently, the greatest conformational mobility that is explained by lack of steric restrictions during rotation around covalent bonds due to lack of a side chain in glycine. The same value of residual correlation is observed for S16 in peptide **P13**, in which the serine residue has no possibility to participate in intramolecular interactions. The value of residual correlation for residue S16 in the maternal peptide **P5** is relatively high and close to 0.30 (conformational mobility is relatively low), and this may be explained by formation of a hydrogen bond between S16 and Q18.

Thus, the data obtained in this work show that changes in conformational mobility of amino acid residues in analogs of heptapeptide AFP_{14-20} are linked with disruptions or arising of intramolecular interactions, that, in turn, influence existence of steric restrictions during rotation around covalent bonds. In heptapeptide AFP_{14-20}, seemingly, a hydrogen bond between S16 and Q18 arises, but threonine residue T20 does not participate in intramolecular interactions, despite the existence of the hydroxyl group that is also able to form hydrogen bonds.

Our data show that substitution of any amino acid residue to a residue that differs in physicochemical properties may not lead to notable changes in conformational and dynamic properties of a peptide if intramolecular interactions are not changed. Analysis of amino acid substitutions in the AFP_{14-20}-like motifs of regulatory proteins showed that these substitutions occur in the manner to preserve existing intramolecular interactions. However, not all functional groups of amino acid side chains that have the ability to participate in intramolecular interactions realize this ability.

Thus, MD calculations allow revealing of amino acid residues in peptides (and proteins) that realize their ability to participate in intra- and intermolecular interactions and discovering mechanisms underlying conformational changes.

REFERENCES

1. Terentiev, A. A., and Moldogazieva, N. T. (2006) *Biochemistry (Moscow)*, **71**, 120-132.
2. Terentiev, A. A. (1997) *Vestnik RGMU*, **1**, 76-79.
3. Terentiev, A. A., Simonova, A. V., Arshinova, S. S., Kulakov, V. V., Bakhus, G. O., Pavlichenkov, A. V., Kudryavtseva, E. V., Bespalova, Zh. D., Ovchinnikov, M. V., Moldogazieva, N. T., and Tagirova, A. K. (2004) *Immunologiya*, **5**, 279-281.
4. Terentiev, A. A., Moldogazieva, N. T., Tagirova, A. K., and Kazimirskaya, V. A. (2004) *Russ. J. Immunol.*, **9**, 54.
5. Kazimirsky, A. N., Moldogazieva, N. T., Tagirova, A. N., Poryadin, G. V., Salmasi, J. M., Alexandrova, I. A., and Terentiev, A. A. (2006) *Tumor Biol.*, **27**, Suppl. 2, 35.
6. Zaretsky, J. Z., and Wreschner, D. H. (2008) *Translat. Oncogenomics*, **3**, 99-136.
7. Moldogazieva, N. T., and Terentiev, A. A. (2006) *Uspekhi Biol. Khim.*, **46**, 99-148.
8. Terentiev, A. A., and Moldogazieva, N. T. (2007) *Biochemistry (Moscow)*, **9**, 920-935.
9. Neduva, V., and Russel, R. B. (2005) *FEBS Lett.*, **579**, 3342-3345.
10. Terentiev, A. A., and Moldogazieva, N. T. (2007) in *Tumor Markers Research Focus* (Chang, D. H., ed.) Nova Science Publishers, N. Y., pp. 163-176.
11. Moldogazieva, N. T., Terentiev, A. A., Kazimirsky, A. N., Antonov, M. Yu., and Shaitan, K. V. (2007) *Biochemistry (Moscow)*, **5**, 529-539.
12. Moldogazieva, N. T., Shaitan, K. V., Tereshkina, K. B., Antonov, M. Yu., and Terentiev, A. A. (2007) *Biofizika*, **4**, 611-624.
13. Gohlke, H., Kiel, C., and Case, D. A. (2003) *J. Mol. Biol.*, **330**, 891-913.
14. Case, D. A., Cheatham, III, T. E., Darden, T., Golhke, H., Lou, R., and Merz, K. M., Jr. (2005) *J. Comput. Chem.*, **26**, 1668-1688.
15. Boeckmann, B., Blatter, M.-C., Famiglietti, L., Hinz, U., Lane, L., Roechert, B., and Bairoch, A. (2005) *Comptes Rendus Biologies*, **328**, 882-899.
16. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2008) *Nucleic Acid Res.* (Database issue), D25-30.
17. Lipman, D. J., and Pearson, W. R. (1985) *Science*, **227**, 1435-1441.
18. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444-2448.
19. Pearson, W. R. (1990) *Methods Enzymol.*, **183**, 63-98.
20. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., and Zhang, Z. (1997) *Nucleic Acids Res.*, **25**, 3389-3402.
21. Smith, T. F., and Waterman, M. S. (1981) *J. Mol. Biol.*, **147**, 195-197.
22. Berman, H. M., Henrick, K., and Nakamura, H. (2003) *Nat. Struct. Biol.*, **10**, 980-989.
23. Froimowitz, M. (1993) *Biotechniques*, **14**, 1010-1013.
24. Pearlman, D. A., Case, D. A., Caldwell, J. W., Seibel, G. L., and Singh, U. C. (1991) *Amber 4.0*, University of California, San Francisco.
25. Cornell, W. D., Cieplak, P., Bayly, C., Gould, I. R., and Merz, K. M. (1995) *J. Am. Chem. Soc.*, **117**, 5179-5188.
26. Dashevsky, V. G. (1974) *Conformation of Organic Molecules* [in Russian], Khimiya, Moscow.
27. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., and Gordon, M. S. (1993) *J. Comput. Chem.*, **14**, 1347-1352.
28. Westerhoff, H., and van Dam, K. (1992) *Thermodynamics and Regulation of Transformations of Free Energy in Biosystems* [Russian translation], Mir, Moscow.
29. Lemak, A. S., and Balabaev, N. K. (1995) *Mol. Simulation*, **15**, 223-231.
30. Lemak, A. S., and Balabaev, N. K. (1996) *J. Comput. Chem.*, **17**, 1685-1695.
31. Golo, V. L., and Shaitan, K. V. (2002) *Biofizika*, **47**, 611-617.
32. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.*, **7**, 75-102.
33. Gowariker, V. R., Viswanathan, N. V., and Shreedhar, J. (1986) *Polymer Science*, New Age International Ltd., New Dehli.
34. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.*, **157**, 105-132.
35. Samanta, U., Bahadur, R. P., and Chakrabarti, P. (2002) *Protein Eng.*, **15**, 659-667.