

Joint Effect of Histone H1 Amino Acid Sequence and DNA Nucleotide Sequence on the Structure of Chromatosomes: Analysis by Molecular Modeling Methods

T. K. Gorkovets^a, G. A. Armeev^a, K. V. Shaitan^a, and A. K. Shaytan^{a, *}

^a Department of Biology, Moscow State University, Moscow, 119234 Russia

*e-mail: alex@intbio.org

Received January 21, 2018; in final form, March 15, 2018

Abstract—A chromatosome consisting of a nucleosome core, linker DNA, and linker histone (LH) is an important structural element of chromatin and plays a role in replication and transcription regulation. There are two experimentally confirmed modes of LH binding to the nucleosome and linker DNA that differ in their geometry: on-dyad and off-dyad binding. It was shown that the LH amino acid sequence influences the type of histone binding and the conformation of the chromatosome. However, the geometry of linker DNA bound with LH also changes. Thus, the mutual influence of these factors and the molecular basis determining the type of LH binding to nucleosomes remain unclear. In this study, molecular modeling methods, including homology modeling, atom–atom interaction analysis, and DNA deformation energy analysis, were applied to study the joint effect of the LH amino acid sequence and the DNA nucleotide sequence on the configuration of the chromatosome. The known crystal and NMR structures of the chromatosome for atom–atom interactions of LH and DNA, as well as the energy of DNA deformation in these structures for various DNA sequences, were analyzed. For various LH H1 variants, the analysis was carried out using homology modeling methods. Sequence-dependent differences in the bending energy of the linker DNA for two different conformations of the chromatosome were found, and nucleotide sequences preferred for these structures were proposed. As a result of the analysis, it was shown that the DNA nucleotide sequence, along with the LH amino acid sequence, influences the type of binding to the nucleosome. Hypotheses for experimental verification have been formulated, according to which the type of LH binding can change with different DNA nucleotide sequences.

Keywords: chromatosome, nucleosome, chromatin, linker histone, linker DNA, DNA flexibility, homology modeling.

DOI: 10.3103/S0096392518020025

INTRODUCTION

The structural units of chromatin are nucleosomes, DNA-protein complexes containing histones. The core of the nucleosome is formed by the octamer of histones H2A, H2B, H3, and H4, which is circled by a double DNA helix approximately 146 base pairs long [1]. The nucleosome has a pseudosymmetry axis of the second order, also called the dyadic axis (Fig. 1a). This axis passes through the center of the nucleosomal DNA, called the dyad.

The next level of compactification is the chromatosome [2] formed by binding of the linker histone (LH) H1 to the nucleosome that includes linker DNA sections that go beyond the core of the nucleosome (Fig. 1a). Thus, the chromatosome contains the octamer of histones H2A, H2B, H3, and H4, nucleosomal DNA, LHs, and DNA linkers.

LHs are present in many eukaryotic organisms in several variants differing in the length of the amino

acid sequence and its composition [3]. Different types of LH can be expressed in different cells and tissues. As a rule, the simplest organisms have only one LH variant, whereas, for example, 11 variants are known in humans [4], some of which are expressed only in germ cells. It is also worth noting that the presence of one or another LH variant depends not only on the type of cell but also on the stage of the cell cycle. Hereinafter, the term “LH” will be used for all linker histones without taking into account their belonging to a particular species or type.

Most LHs contain approximately 200 amino acid residues. LHs consist of three domains: a short and unordered N-terminus, followed by a globular domain formed by 70–80 amino acid residues and having a conservative tertiary structure, and a long C-terminal domain of approximately 100 amino acid residues. The C-terminus is unorganized and contains many lysine residues. According to experimental data [5], the globular domain of LH H1 binds to the nucleo-

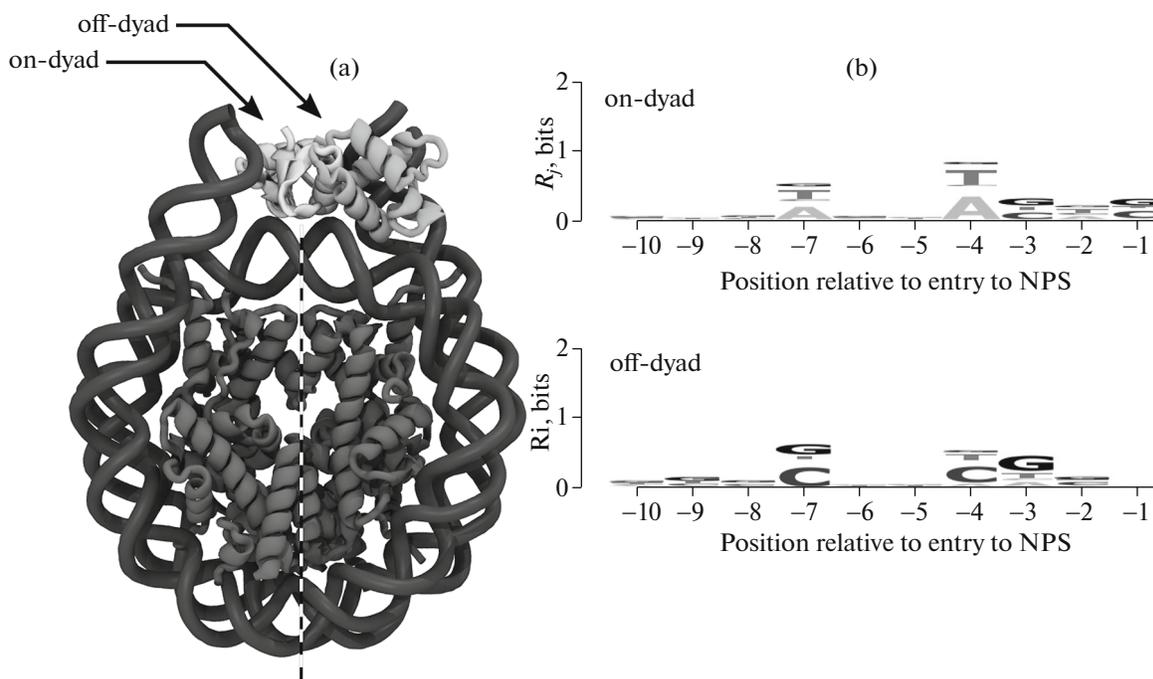


Fig. 1. (a) The view of the structure of the chromatosome. The protein and DNA are shown as representations of the secondary structure. Linker histones are indicated by shades of gray. The dyad axis is shown as a dotted line. (b) Visualization of the frequency of nucleotides for sequences tending to form an on-dyad (above) and off-dyad (bottom) structure. Visualization is in the form of sequence logos.

some together with the full-length LH H1. In view of the high degree of disorder in the terminal domains, there are no crystalline structures of full-length LH, but there are crystalline structures of the globular domain [6]. It is known that LH H1 contains a number of amino acid residues that play an important role in the formation of bonds with the nucleosome [7]. These residues are separated into two binding sites: the first is H25, R47, K69, K73, R74, and K85, which forms bonds with a large groove of DNA near the dyad of the nucleosome, and the second is R42, R94, and K97, which binds to the linker DNA.

Recently, the configuration of LH binding to the nucleosome has been discussed. Previously, there were various models of such binding, some of which even suggested the incorporation of LH between nucleosomal DNA loops and the histone octamer [8]. Such models did not find experimental confirmation [5, 9], in contrast to the models that suggested LH binding in the dyadic region of the nucleosome and linker regions of DNA.

The available experimental data [10–13] suggest two possible arrangements of LH in a chromatosome, conventionally called on-dyad and off-dyad. At the moment, the crystal structure of the chromatosome with the globular domain of LH H5 in *Gallus gallus* (H5 is the historical name of LH H1 in *G. gallus*) in the on-dyad configuration and the structure model of the chromatosome with the globular domain of LH H1 in

Drosophila melanogaster in the off-dyad configuration have been obtained [12]. It is important to note that the structure of the chromatosome in the off-dyad configuration was constructed with the help of molecular docking techniques based on the tetranucleosome structure (pdb code 1zbb) using nuclear magnetic resonance (NMR) data.

The globular domain of histone H5 in *G. gallus* demonstrated on-dyad binding to the nucleosome, whereas the globular domain of H1 in *D. melanogaster* showed off-dyad binding. Analysis of their sequences showed differences that probably play a role in the preference for a certain type of binding. Thus, the globular domain of H5 in *G. gallus* contains positively charged amino acid residues (R47, K55, R74, and K97) at positions corresponding to neutrally charged amino acid residues in the globular domain of H1 in *D. melanogaster* (L68, T76, S96, and A119) and neutrally charged amino acid residues (Q51, V80, and V87) at positions corresponding to positively charged residues of H1 in *D. melanogaster* (K72, K102, and K109) [12, 13]. In spite of the fact that some of the key residues do not form direct contacts with the nucleosome, when they were replaced in the globular domain of H5 by the amino acid residues characteristic of H1, LH showed a change in the type of binding from on-dyad to off-dyad [13].

Another factor that can be expected to influence the configuration of the chromatosome is the nucleo-

tide sequence of the DNA, both the linker and the one comprising the nucleosome. According to the nucleosome positioning study, it was suggested that LH preferably binds to AT-rich DNA regions [14].

Despite the fact that LH H1 showed on-dyad bindings in the resulting crystal structure of the chromatosome, according to the data obtained by molecular modeling [15, 16], the globular domain of LH H5 can also demonstrate the off-dyad type of binding. One explanation is the fact that the nucleotide sequences of DNA used in the modeling differed from the nucleotide sequence for which the crystal structure of the on-dyad conformation of the nucleosome complex with the globular H5 domain was obtained [12].

The above-described studies consider LH binding to a single nucleosome in vitro, whereas structures of a higher level of compactification are present in the cell: fibrils of 30 nm in diameter, which are a chain of nucleosomes and LH associated with linker DNA. According to the latest data of cryoelectron microscopy, LH in the structure of a 30 nm in diameter fibril formed by 12 nucleosomes demonstrates off-dyad binding, which may be due to the geometry of linker DNA in the fibril structure [10, 17]. Thus, it is impossible to exclude the effect of DNA caused by LH “recognizing” the specific DNA form upon binding [18].

Based on the foregoing, it is very likely that the type of LH binding to the nucleosome and linker DNA is affected not only by the amino acid sequence of histone itself but also by the nucleotide sequence of DNA as well as by the structure of the chromatin fibril. It is also known that, when a protein-DNA complex is formed, the strength and the type of binding are affected by both the direct interactions of protein with base pairs and the sugar-phosphate backbone and the geometry of DNA due to the energy of its deformation [18]. However, despite the presence of chromatosome structures in various configurations, the joint influence of these factors has not been studied previously.

In this paper, the joint effect of the amino acid sequence of LH and the nucleotide sequence of DNA on the type of the chromatosome configuration was studied using molecular modeling methods, namely, atom–atom interaction analysis, homology modeling, and DNA deformation energy analysis. The investigation made it possible to predict the nucleotide sequences of linker DNA that are most preferable for a particular type of binding and to formulate hypotheses for experimental verification.

MATERIALS AND METHODS

Analysis of amino acid sequences of various histone variants. In this work, amino acid sequences of LH H1 located in the HistoneDB database [4] were used. Alignment of available amino acid sequences was carried out in the Muscle program [19]. For further

analysis of the sequences, the Biopython toolkit was used [20].

Homology modeling. Homology simulation was carried out in the Modeller program [21, 22]. As the template proteins, the crystal structure of the chromatosome in the on-dyad configuration with pdb code 4qlc and the model of the chromatosome in the off-dyad configuration obtained on the basis of NMR data were used [12]. Using homology modeling, ten models were constructed for each investigated histone, from which the best structures for analysis of contacts were selected based on DOPEscore. Additional optimization of structures obtained by homology modeling was not carried out. The analysis of contacts was carried out using the Chimera software package [23].

DNA deformation energy calculation. Calculation of the dependence of the deformation energy of linker segments of DNA on its sequence was carried out in the space of generalized variables Tilt, Roll, Twist, Shift, Slide, and Rise; the transition from atomic coordinates to generalized ones was carried out using the 3DNA program [24]. The deformation energy of DNA sections was carried out in accordance with the following formula:

$$E = E_0 + \frac{1}{2} \sum_{i=1}^6 \sum_{j=1}^6 f_{ij} \Delta\theta_i \Delta\theta_j,$$

where E_0 is the minimum deformation energy (independent of the conformation, taken as 0), the summation is carried out over generalized variables for each neighboring nucleotide pairs, and f_{ij} are the stiffness coefficients for deviating the generalized variables from the equilibrium configuration. For calculations, a set of elastic coefficients and mean values for nucleotide pairs was used as described in [25].

The deformation energy was determined for all possible DNA sequences of each of the four DNA sections (two for each LH binding model) of 12 bp each. The DNA section next to which LH is located in the off-dyad model was called the entry into the nucleosome positioning sequence (NPS), and the opposite site was called the exit from NPS. For each variant of the sequence, the difference in the deformation energy of DNA between the on-dyad and off-dyad models was calculated. The resulting energy difference distribution was visualized using a graphical representation of conservatism of nucleotides: a sequence logo for 5% of structures with the largest (positive) energy difference (preferred for the on-dyad model) and 5% of structures with the lowest (negative) energy difference (preferred for the off-dyad model).

RESULTS AND DISCUSSION

The present study consisted of several complementary components. Based on the analysis of the experimental data, key LH residues were identified that contribute to one or another type of binding, on the basis

of which a classification model was constructed and applied to human LHs—H1.1, H1.2, H1.3, H1.4, H1.5, H1.0 (H1T°), TS H1.6 (H1T), TS H1.7 (H1T2, HANP1), OO H1.8 (H1oo), TS H1.9 (HILS1), H1.10, H1.11—as well as to LH H1 of *Xenopus laevis*. In this model, the probability of one or another type of binding for an arbitrary LH was estimated as follows: (1) the sequence of the investigated LH was aligned with the LH sequences with a known binding type; (2) the number of coincident types of amino acid residues was counted in key positions for binding. This model is predictive and does not allow fully estimating the accuracy of the proposed classification model until additional experimental data are accumulated.

For the experimentally determined structures of the chromatosome and structures with different variants of LH H1 constructed using homology modeling, contacts between LH and DNA were analyzed. The original analysis of the DNA conformation in various structures of the chromatosome and the analysis of the sequence space of linker DNA in terms of their bending energy and contacts with LH were carried out.

Determination of key LH residues influencing the type of binding. The amino acid sequence of LH H1 has a number of key positions whose mutations lead to a change in the type of binding [12]. Thus, based on the amino acid residues located in these positions, the expected type of binding can be determined for LH, which suggests the classification of LHs by the type of their binding to the nucleosome on the basis of amino acid sequences.

The experimentally determined structure of the chromatosome in the on-dyad and off-dyad configuration contains contacts directly with base pairs and linker DNA as well as nucleosomal DNA. These contacts are represented by both hydrogen bonds and Van der Waals interactions.

Some of the amino acid residues located at key positions (for example, amino acid residues corresponding to K55, Q51, and V87 of LH H5) do not form direct contacts with DNA in known structures. However, despite having no contacts with DNA, these residues can presumably interact with DNA by electrostatic potential, since they have different charges in histones H1 and H5.

LH classification based on the presence of key residues. Based on the proposed classification for three human LHs, the intended type of binding was defined as on-dyad; the type of binding cannot be uniquely determined for seven human histones, since they show similarity in key positions with both histone H5 of *G. gallus* (on-dyad type of binding) and histone H1 of *D. melanogaster* (off-dyad type of binding); and the supposed type of binding can be uniquely determined as off-dyad for one human histone. For LH of *X. laevis*, the binding type was defined as on-dyad.

DNA deformation energy in various configurations of the chromatosome. To determine the dependence of

the type of LH binding on the nucleotide sequence of the linker DNA, the difference in the deformation energies of the linker DNA sections between the on-dyad and off-dyad models was calculated for all possible variants of sequences. A Z-evaluation of the position of the sequence from the original models in the distribution of the calculated deformation energies ($Z = 0.75$) was also made, which suggests that the original sequence used for experimental obtaining of structures tends to form an on-dyad structure.

As can be seen from Fig. 1b, the majority of positions in nucleotide sequences of linker DNA sections are not significant for determining the DNA geometry in the framework of on-dyad and off-dyad models, except for nucleotides in positions -7 , -4 , and -3 . In the sequences preferred for the on-dyad model, A/T, A/T, and G/C are located in these positions, while C/G, C/T, and G/T are more preferable for the off-dyad model. The predominance of thymidines in the linker DNA was shown earlier by Cui and Zhurkin [14].

Also, the preferential (optimal) sequences of linker DNA (for which the bending energy maximally favors some conformation) for the on-dyad (CCGTC-CCGTC-PPN-ACGCCGGCGG) and off-dyad (GACGCCCGAC-PPN-GTGATGCTGC) models were found by the difference in deformation energies.

Analysis of the joint effect of the amino acid sequence of LH H1 and the nucleotide sequence of DNA. Based on the classification proposed earlier for various LH H1 variants, structural models of chromatosomes in the on-dyad and off-dyad configurations were constructed using homology modeling. In the models obtained, the analysis of contacts between LH H1 and DNA was carried out.

In both the models, there are contacts both between the amino acid residues of LH H1 and the sugar-phosphate backbone of DNA and directly with the nitrogenous bases, which supports the hypothesis of the effect of the nucleotide sequence of DNA on the type of binding between LH and the nucleosome due to the effect of direct readout of the DNA sequence by protein. Also, the number of these contacts varies depending on the LH variant for which the model is built. Contacts between LH and nitrogenous bases of DNA can be provided due to the formation of both hydrogen bonds and Van der Waals interactions.

For LH H1 of *D. melanogaster* and H5 of *G. gallus*, using homology simulations, models of chromatosomes were also constructed in configurations opposite to the experimental ones: the on-dyad configuration was used for LH H1 and the off-dyad configuration for LH H5. These models showed a decrease in the number of contacts between LH and DNA, which is consistent with the binding types identified experimentally.

Based on the calculation of the DNA deformation energy, models of the chromatosome were constructed

in the on-dyad and off-dyad configuration with their preferred (optimal) sequences determined by the difference in the DNA deformation energy, for which contacts between LH and DNA were also analyzed. The analysis showed that, when the initial nucleotide sequence in the experimental off-dyad and on-dyad structures is replaced for the optimal sequences determined in our work, which contribute to the initial type of binding, the number of contacts of LH H1 with base pairs either remains the same (for off-dyad binding, three contacts persist) or increases (for on-dyad binding, the number increases from two contacts to three). At the same time, a cross-comparison of the effect of optimal DNA sequences in on-dyad and off-dyad structures showed that structures with the optimal DNA sequences corresponding to their type of binding have the same or higher number of contacts between LH and base pairs than the structures in which the DNA sequence corresponding to the alternative type of binding is used. Thus, the optimal sequences of linker DNA for various conformations of the chromosome as proposed above can achieve their selectivity due to the mechanisms of indirect readout of the DNA sequence by protein and due to direct readout of the interaction of protein with base pairs.

Based on the foregoing, it can be assumed that the nucleotide sequence, as well as the geometry and DNA bending stiffness, is an important factor determining the configuration type of the chromosome. At the present time, this assumption indirectly agrees with the experimental data [10, 17]. It is likely that the amino acid sequence of LH is not a primary factor determining the type of binding of LH H1 to the nucleosome.

Thus, we can formulate a hypothesis that the same LH, depending on the nucleotide sequence of DNA, as well as on the geometry of linker DNA, can demonstrate different types of binding. Such pairs of DNA sequences are proposed in this paper. Later this hypothesis can be verified by estimating the distances between nucleotides by the Förster energy transfer efficiency measurement method (spFRET) using different nucleotide sequences for each LH.

ACKNOWLEDGMENTS

We are grateful to Y. Bai for providing the model of the chromosome. This work was carried out using the equipment of the Sharing Center of Ultrahigh-Performance Computing Resources of Moscow State University [26] with the financial support of the Russian Science Foundation (project no. 14-24-00031, agreement no. 14-24-00031-p).

REFERENCES

1. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J., Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature*, 1997, vol. 389, no. 6648, pp. 251–260.
2. Simpson, R.T., Structure of the chromosome, a chromatin particle containing 160 base pairs of DNA and all the histones, *Biochemistry*, 1978, vol. 17, no. 25, pp. 5524–5531.
3. Lyubitelev, A.V., Nikitin, D.V., Shaytan, A.K., Studitsky, V.M., and Kirpichnikov, M.P., Structure and functions of linker histones, *Biochemistry (Moscow)*, 2016, vol. 81, no. 3, pp. 213–223.
4. El Kennani, S., Adrait, A., Shaytan, A.K., Khochbin, S., Bruley, C., Panchenko, A.R., Landsman, D., Pflieger, D., and Govin, J., MS_HistoneDB, a manually curated resource for proteomic analysis of human and mouse histones, *Epigenet. Chromatin*, 2017, vol. 10, p. 2.
5. Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S., Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, vol. 107, no. 21, pp. 9620–9625.
6. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M., Crystal structure of globular domain of histone H5 and its implications for nucleosome binding, *Nature*, 1993, vol. 362, no. 6417, pp. 219–223.
7. Brown, D.T., Izard, T., and Misteli, T., Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo, *Nat. Struct. Mol. Biol.*, 2006, vol. 13, no. 3, pp. 250–255.
8. Pruss, D., Bartholomew, B., Persinger, J., Hayes, J., Arents, G., Moudrianakis, E.N., and Wolffe, A.P., An asymmetric model for the nucleosome: A binding site for linker histones inside the DNA gyres, *Science*, 1996, vol. 274, no. 5287, pp. 614–617.
9. Zhou, Y.B., Gerchman, S.E., Ramakrishnan, V., Travers, A., and Muyldermans, S., Position and orientation of the globular domain of linker histone H5 on the nucleosome, *Nature*, 1998, vol. 395, no. 6700, pp. 402–405.
10. Bednar, J., Garcia-Saez, I., Boopathi, R., et al., Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1, *Mol. Cell*, 2017, vol. 66, no. 3, pp. 384–397.
11. Zhou, B.-R., Feng, H., Kato, H., Dai, L., Yang, Y., Zhou, Y., and Bai, Y., Structural insights into the histone H1-nucleosome complex, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, vol. 110, no. 48, pp. 19390–19395.
12. Zhou, B.-R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S., and Bai, Y., Structural mechanisms of nucleosome recognition by linker histones, *Mol. Cell*, 2015, vol. 59, no. 4, pp. 628–638.
13. Zhou, B.-R., Feng, H., Ghirlando, R., Li, S., Schwieters, C.D., and Bai, Y., A small number of residues can determine if linker histones are bound on or off dyad in the chromosome, *J. Mol. Biol.*, 2016, vol. 428, no. 20, pp. 3948–3959.
14. Cui, F. and Zhurkin, V.B., Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA, *Nucleic Acid Res.*, 2009, vol. 37, no. 9, pp. 2818–2829.

15. Öztürk, M.A., Pachov, G.V., Wade, R.C., and Cojocaru, V., Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome, *Nucleic Acid Res.*, 2016, vol. 44, no. 14, pp. 6599–6613.
16. Pachov, G.V., Gabdoulline, R.R., and Wade, R.C., On the structure and dynamics of the complex of the nucleosome and the linker histone, *Nucleic Acid Res.*, 2011, vol. 39, no. 12, pp. 5255–5263.
17. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., and Li, G., Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units, *Science*, 2014, vol. 344, no. 6182, pp. 376–380.
18. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S., Origins of specificity in protein-DNA recognition, *Annu. Rev. Biochem.*, 2010, vol. 79, pp. 233–269.
19. Edgar, R.C., MUSCLE: Multiple sequence alignment with high accuracy and high throughput, *Nucleic Acid Res.*, 2004, vol. 32, no. 5, pp. 1792–1797.
20. Cock, P.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L., Biopython: Freely available Python tools for computational molecular biology and bioinformatics, *Bioinformatics*, 2009, vol. 25, no. 11, pp. 1422–1423.
21. Webb, B. and Sali, A., Comparative protein structure modeling using MODELLER, *Curr. Protoc. Bioinf.*, 2016, vol. 54, pp. 5.6.1–5.6.37.
22. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A., Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, vol. 29, pp. 291–325.
23. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E., UCSF Chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, vol. 25, no. 13, pp. 1605–1612.
24. Lu, X. and Olson, W.K., 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acid Res.*, 2003, vol. 31, no. 17, pp. 5108–5121.
25. Olson, W.K., Gorin, A.A., Lu, X.-J., Hock, L.M., and Zhurkin, V.B., DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, vol. 95, no. 19, pp. 11163–11168.
26. Voevodin, V., Zhumatii, S., Sobolev, S., Antonov, A., Bryzgalov, P., Nikitenko, D., Stefanov, K., and Voevodin, V., The practice of the supercomputer Lomonosov, *Otkrytye Sist., SUBD*, 2012, no. 7, pp. 36–39.

Translated by K. Lazarev