E19-2012-146

V. S. Sivozhelezov[1], I. V. Bednyakov, T. P. Akishina,
R. V. Polozov, V. V. Ivanov

# ELECTROSTATICS AS A FACTOR OF BIOMOLECULAR RECOGNITION IN PROCESSES OF TRANSCRIPTION AND TRANSLATION

[1] Institute of Cell Biophysics, RAS, Pushchino

Сивожелезов В. С. и др. E19-2012-146
Электростатика как фактор биомолекулярного распознавания
в процессах транскрипции и трансляции

Электростатика ДНК, РНК и белков в основном определяет функции этих биомолекул в процессах транскрипции и трансляции, в которых генетическая информация преобразуется в аминокислотную последовательность белков, а затем реализуется в пространственной структуре и функции этих белков. В настоящей работе мы вычислили электростатические потенциалы промоторных участков ДНК, сигма-доменов РНК-полимеразы и транспортной РНК и определили связь между вычисленными потенциалами и функцией этих биомолекул в транскрипции и трансляции. При вычислении мы использовали распределенные среды, для которых биомолекулы являются естественной областью применения в силу как их сложности, так и большого гомологического разнообразия.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2012

Sivozhelezov V. S. et al. E19-2012-146
Electrostatics as a Factor of Biomolecular Recognition
in Processes of Transcription and Translation

Electrostatics of DNA, RNA and proteins is mainly determined by the function of these biomolecules in the processes of transcription and translation, in which the genetic information is converted into the amino acid sequence of proteins, and then implemented in the spatial structure and function of these proteins. Herein, we calculated the electrostatic potentials of the promoter regions of DNA, sigma domain of RNA polymerase and transport RNA, and determined the relationship between the calculated potentials and function of biomolecules in transcription and translation. We used distributed computational environment, for which biomolecules are the natural field of application due to their complexity and high homologous diversity.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2012

## 1. INTRODUCTION

Major cellular processes, such as gene transcription and translation, signal transduction, electron and ion transport, cell motion, a plethora of regulatory mechanisms, etc., are driven by evolutionarily selected processes of biomolecular recognition. Protein–protein and protein–nucleic acid recognition are the most complex and diverse of all recognition processes inherent in living cells. Properties of molecular complexes, such as nonlinearity, large dimensions (tens and hundreds of thousands of atoms are involved), enormous numbers of degrees of freedom, presence of characteristic collective motions, a wide range of spatial and temporal scales resulting in hierarchical nature of recognition, ensure their high biological specificity. The same properties make the physics of biomolecular recognition very difficult in terms of both problem setup and problem solution. The very dimensions of biomolecular systems are a part of the difficulty, since they are large in terms of atomic scale but their specificity prevents the use of macroscopic, solid-state approaches.

A large body of experimental data available of protein–protein and protein–nucleic acid interactions with well-defined conditions and parameter values provide a sufficient reference frame for computer modeling, within the entire temporal range important for biomolecular recognition. The arsenal of methods used in computer modeling, in principle, can also cover this entire range to provide exhaustive description and predictive power with respect to the experimental data. However, those methods very often fail to guarantee either accuracy or reliability. Thus, it is crucial to learn to model or simulate such systems with sufficient accuracy, and study their behavior under wide ranges of conditions, thus studying the system «as a whole». These continuously increasing requirements for accuracy, both in terms of parameter sets for molecular modeling (force fields) and in terms of methods and algorithms for calculating molecular structure and dynamics, urge for reformulation in both problem setups and computational implementations of solutions. It must be noted that the existing methods and algorithms, particularly those for modeling biomolecular recognition («docking» methods) ignore the hierarchical nature of recognition process and can therefore be misleading. This background has urged us to formulate the strategy described below:

- computational molecular modeling as the basic tool, necessary because of the above-mentioned complexity of biomolecular recognition and the necessity for atomic-scale modeling;
- selection and parameterization of biomolecular force fields necessary for accurate description of biomolecular interactions, particularly, electrostatic interactions that are particularly difficult to parameterize;
- development of sufficiently accurate and efficient methods and algorithms of recognition processes involving docking, equations of motion, statisti-

cal methods, database scanning methods, visualization, etc., including also software and hardware development;

- interpretation of modeling data and development of basic models of recognition that will result in an integrated description of recognition processes combining spatial and temporal hierarchical scales in a consistent manner.

Implementation of this strategy was started from calculations of electrostatic properties of biological molecules, considering both the theoretical prerequisites and the available body of experimental research proving the highest priority of electrostatic studies. The theoretical prerequisites are primarily the properties of electric fields determining their role in biology in general and in biomolecular recognition in particular. Firstly, electrostatic forces decay with distance much slower than other forces, so only they can account for long-distance recognition. Further, electric charge and potential have signs responsible for interaction specificity which is essential in biology. Besides, electrostatics serves as the driving force of recognition, at least at its earlier stages. The body of experimental data shows how these properties are manifested in biomolecular systems and render specificity to them. Indeed, electrostatic interactions play a key role in determining the mechanism of protein–protein complex formation, protein thermal stability, conformational transitions and dynamics of proteins. The specific distribution of electrostatic potential of a protein or a nucleic acid is essential during binding of the biopolymer with other biopolymers. Electrostatics underlies dependencies of thermodynamics and kinetics of protein interactions on ionic strength, pH, and point mutations of charged amino acid residues.

## 2. METHODS

Electrostatic calculations were performed by solving the nonlinear Poisson–Boltzmann equation that relates the electric potential with the charge distribution, protein partial charges taken from the AMBER force field [1], mobile solution charges approximated by Boltzmann distribution, with the dielectric constant assumed to be 2 inside the protein and 80 outside the protein. The electrolyte was assumed to be 1:1 ($z_1 = 1$, $z_2 = -1$) at physiological or subphysiological 50–150 mM concentrations. Solution is sought with finite difference multigrid method using a sequence of nested finite difference grids, the finest grid having up to $200 \times 200 \times 200$ points so that the interval between grid points is less than 1 Å. We have developed an algorithm of solving the nonlinear Poisson–Boltzmann equation allowing one to efficiently calculate electrostatic potentials for large objects such as proteins, nucleic acids, and their complexes [2]. Computation time in our implementation of the multigrid solution of the nonlinear Poisson–Boltzmann equation is proportional to $N$, where $N$ is the number of nodes in the grid. It

allows one to handle large molecular complexes such as ribosomal subunits, and long DNA fragments up to 1000 base pairs, including promoter sequences for both prokaryotic and eukaryotic species, using PC computers only, without the necessity to use professional workstations.

Visualization was performed using the software package MOLMOL [3], with our modifications concerning the distance at which the potential is mapped at the distance of Bjorrum length (7 Å) from the nuclei at which the interaction energy of two elementary charges equals the energy of thermal motion. Additionally, for DNA, the potential was visualized at the surface called the «electrophoretic sliding surface» (15 Å from the cylindrical axis of DNA).

The objects of investigation adopted herein are the biomolecules responsible for the most important and universal cellular processes of transcription (DNA-based RNA synthesis) and translation (RNA-based protein synthesis), i. e., proteins and DNAs of transcription, and proteins and RNAs of translation.

## 3. RESULTS

**3.1. Electrostatics and Curvature of Nucleic Acid Surfaces.** We calculated electrostatic potential distributions around nucleic acid fragments of specific shapes. The electrostatic potential values responsible for the earliest stage of nucleic acid–protein recognition were found to be controlled by charges on the phosphate groups alone. Significant anisotropy was observed for electrostatic potential of tRNAphe as well as its t-loop fragment, for which specific shapes of their molecular surfaces were responsible. We found this anisotropy to be caused by curvature of the nucleic acid shapes including hairpins, cruciforms, and loops in nucleic acids and charges of phosphate backbone, rather than charges of nitrous bases. The results prompted us to propose a hypothesis that dynamic control is possible in genome functioning mediated by electrostatic interactions involving transient nucleic acid structures of specific shapes, with phosphate groups being the sources of the necessary electric fields.

**3.2. Electrostatic Potential of tRNAs.** Distributions of phosphate backbone produced electrostatic potentials around several tRNAs were calculated by solving the nonlinear Poisson–Boltzmann equation, for tRNAs both free and bound to the proteins involved in translation: aminoacyl-tRNA synthetase and elongation factor EF-TU [4]. Comparison between the tRNAs allowed us to identify several regions of strong negative potential related to typical structural patterns of tRNA and invariant throughout the tRNAs. These patterns were found to be conserved upon binding of tRNAs to proteins, but electric potentials in the invariant patches and areas occupied by these patches depended on the particular tRNA-binding protein. Comparison of the calculated pK shifts of fluorescently labeled tRNA based on tRNA electric potentials with experimentally observed pK shifts of fluorescent

labeled tRNAs indicates that the total charge of tRNA is within the interval $40\,q$ to $70\,q$. This large charge leads to high absolute values of electric potential around tRNAs that allows one to propose a mechanism of electric charge switching on the corresponding synthetase. Due to its strong negative charge, tRNA increases the proton concentration in its nearest neighborhood, thus inducing positive charges on histidine residues of the synthetase during the early stage of protein–tRNA recognition.

**3.3. Sigma2 Domain of RNA Polymerase.** RNA polymerase (RNAP) is involved in processing and control of genetic information in all biological species. Regulation of transcription is largely performed by the sigma subunit of RNA polymerase, which controls transcription by recognizing 10 promoter elements of DNA. The earliest step of this recognition involves interaction of the promoter with the sigma2 domain of RNAP. We focus on the sigma2 domain. By calculating and visualizing electrostatic potential distribution of four different sigma2 domains, we identify the positive potential patch responsible for the RNAP interaction with 10 elements, which is conserved despite the differences in charged amino acid residue distribution along the sigma2 amino acid sequences. We found that those sigmas (two out of four) that are normally locked by antisigma factors and released only by external signal show very different electrostatic potentials at their contact regions with antisigma, which suggests different mechanisms of their interactions with the respective antisigmas. For one of the remaining two sigmas, which require no external antisigma factors and instead possess autoinhibition properties, electrostatic calculations suggest an autoinhibition/activation mechanism involving concerted movements of the acidic loop (belonging to large insertion present in sigma2 domains of many bacterial species) and the sigma1.1 domain. This mechanism cannot operate in the other autoinhibited sigma having a smaller insertion with no acidic loop in it. Comparative analysis performed herein indicates that similar functions in transcription control can be served by different means, underlining the complex interplay between evolutionary conservation and evolutionary diversity in developing such functions.

**3.4. Electrostatic Potentials of *E. coli* Promoters.** Origin, evolution, function, and regulation of promoter DNA are presently analyzed based on their sequences alone. This analysis is insufficient since it is the physicochemical properties of DNA that control the process of gene transcription and its regulation. In this work, an extension of analysis is performed based on physicochemical properties of specific DNA sequences. Classification of promoters and other functionally important genome fragments according to their sequence and physicochemical properties is a key factor for understanding gene transcription, replication, recombination, and their regulation. Electrostatic interactions comprise an essential component of those processes. Electrostatic potentials of *E. coli* promoters were calculated as well as periodic sequences [5]. Specific electrostatic characteristics of promoter DNA corresponding to features of their primary

structure provide a foundation of classification of promoters based on both their sequence and their electrostatic potentials as well as other physicochemical properties.

Interaction of DNA with polymerases and other proteins that play key roles in transcription and its regulation is one of the most important examples of molecular recognition where selective binding of protein to the particular DNA sequence occurs [6]. Specificity of binding can be evaluated in terms of energy, by the difference in free energies for binding the same protein to the specific and average nonspecific DNA site. This value varies from about 40 to over 80 kJ/mole [7], which is quite a large difference considering that only noncovalent forces are involved in protein–DNA binding.

Such a wide range of specificity led to formulating a model of protein–DNA recognition process involving at least three steps [8]. The first is nonspecific binding of a protein to DNA which is energetically driven by the electrostatic complementarity of the DNA and protein contacting surfaces [9]. The second step is one-dimensional diffusion of a protein along DNA chain, which accelerates association rates beyond their three-dimensional diffusion limits [10–12]. During this step, electrostatic interactions of proteins with DNA retain the protein in the immediate vicinity of DNA, thus providing the required reduction in dimension from three to one. The third step is formation of more extensive contacts between DNAs, which occurs when a protein locates its target site. Again, the specific interaction of a protein with its target DNA sequence involves electrostatic interactions [8], but other factors also contribute, e. g., the mutual surface fitting due to the DNA-induced protein refolding [13, 14] and protein-induced conformation changes of DNA [15, 16].

Thus, electrostatic interactions are of primary importance in the multistep process of the protein–DNA recognition. In the first step of that process which occurs approximately at the electrophoretic sliding surface of DNA, which is about 15 Å away from the DNA longitudinal axis [17], the electrostatic interaction is the only physical factor since Coulomb electrostatic forces decay with distance much lower than other forces like hydrogen bonding, London forces, etc.

Even more importantly, calculation of electrostatic potential distribution along DNA for long chains will open the road to analyzing correlations of DNA functional properties with physical properties of the DNA sequence, particularly, the electrostatic properties. Earlier, correlations were established between the properties and the sequences themselves, and classification of DNA sequences was performed using the well-known cluster analysis technique. Such a classification allows one to elucidate structure–function relationships [18]. The drawback of such a classification is that it has no explicit physical basis. In contrast, correlations of electrostatic properties with functions will allow one to establish such a basis. Besides, DNA electrostatic properties are already known to correlate with its sequence, but that was earlier established for short DNA chains only.

On the other hand, the sequences of coding and promoter regions of DNA correlate between DNAs of various biological species, which allows one to identify evolutionary relationships, again via classification by cluster analysis [19]. Once correlations are characterized between electrostatic properties of those regions, the corresponding evolutionary relationships will acquire the physical basis.

Since distributions of electrostatic potentials or fields have distinct geometrical shapes, the classification can be inferred via morphology methods (Procrustean, Minkowski, or other metrics). The most accurate calculation method of electrostatic potentials and energies available for macromolecular systems is numerically solving the Poisson–Boltzmann equation on a rectangular grid [20]. But this method was not used for long DNA sequences recognized by some DNA-binding proteins, because the number of grid points $N$ scales linearly with the DNA length, and computation time typically scales, at best, as $N$.

In this work we adopt a multigrid method of solving the Poisson–Boltzmann equation in which computation time typically scales as $\ln N$, which allows us to handle several hundred base pairs long DNA sequences, exemplified in this study by *E. coli* promoter regions, which are 411 base pairs long.

The horizontal axis in all the maps shown (Figs. 1–4) coincides with the DNA helix axis. The color scale represents the electrostatic potential in units of $k_BT/q$, which is thermal motion energy $k_BT$ per unit of electric charge $q$. In those units, red color was chosen to correspond to 1.3, blue to 0.8, and white to intermediate values. In this colour scheme, the visualized electrostatic potential values will span a range of 0.5 $k_BT/q$, so that ten unit charges, which are typically present in protein fragments interacting with DNA, will account for a difference of 5 $k_BT/q$, which is quite sufficient for the electrostatic steering that happens as the protein approaches the DNA surface. Ion concentrations (1:1 electrolyte was assumed) were 0.15 M, which is the physiological value.

DNA sequences of *E. coli* promoter regions were taken from [23] and [24]. The start point of transcription is located at the position 257, so the coding sequence starts further downstream, and the promoter region is upstream from that point.

Figure 1 presents the electrostatic potentials of periodic DNA: poly(A), poly(AT), poly(G), and poly(GC). As one can see from Fig. 1, this electrostatic potential is also periodic in nature. The fact that the periodicity does not appear perfect on the cylindrical surface is explained by the geometry of B form of DNA. One can also see that the potential of poly(AT) sequence is drastically different from the rest of periodic sequences. Particularly, the spots of both the blue (less negative) and the red (more negative) potential are smaller and much less intense, indicating that the potential of the poly(AT) DNA sequence deviates from its average value much less than for other periodic sequences. Also, the alternating blue and red bands appear more frequently the electrostatic potential of the poly(AT) sequence, indicating that the potential of poly(AT) is

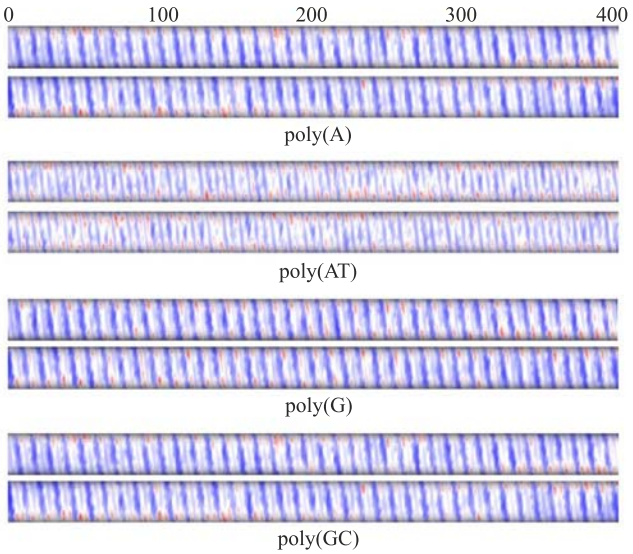


Fig. 1. Distribution of electrostatic potential around periodic DNA molecules. Each molecule is shown in two views differing by 180° rotation around the helix axis

of finer structure than for other periodic DNA sequences. Those distinguishing features of the poly(AT) electrostatic potential show that the electrostatic potential should strongly correlate with the presence of long (AT) runs in the sequence.

This feature itself shows that the calculations of the DNA electrostatic potential can contribute to the classification of DNA sequences in a manner similar to the analysis of the sequence itself, leading to expansion of the entire field of bioinformatics. Particularly, instead of building classifications based on the sequence alone, at least one physical property can be allowed for in building classifications, namely, the electrostatic potential.

Figure 2 shows the electrostatic potential of several promoter regions of *E. coli*, together with the adjacent coding regions. Qualitatively, the electrostatic potentials of these regions noticeably differ from the potentials of periodic sequences. The main difference is apparent presence of a strong dipolar component in the electrostatic potential across the DNA double helix. Indeed, the intense blue spots (less negative potential) are located well away from the intense red spots (more negative potential). In contrast, the periodic DNA sequences (Fig. 1) exhibit a more homogeneous distribution of the electrostatic potential across the double helix, visually more similar to the quadrupolar distribution.

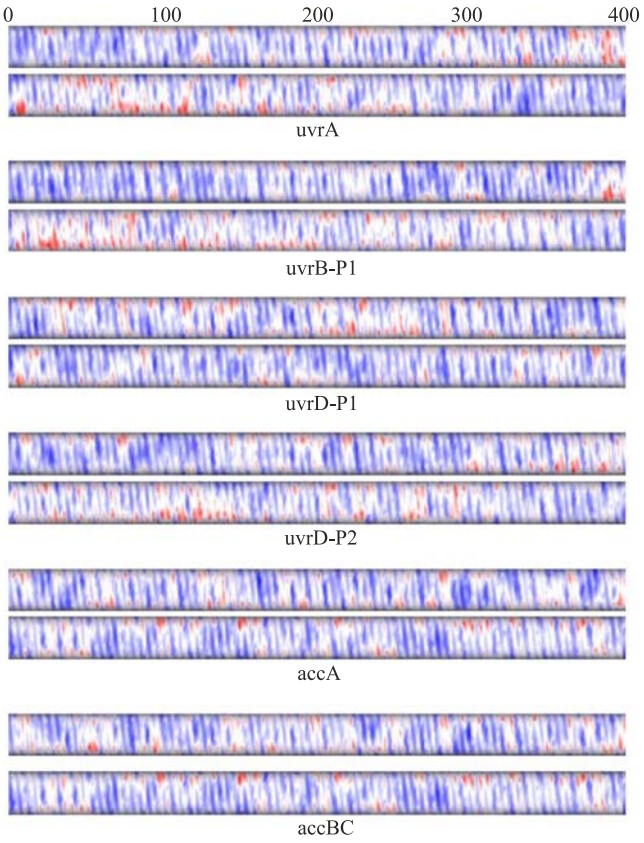


Fig. 2. Distribution of electrostatic potential around promoter DNAs of *E. coli*. Each promoter is shown in two views differing by 180° rotation around the helix axis

Of the six promoters shown, two top promoters, uvrA and uvrB-P1, show the maximal anisotropy of the electrostatic potential. Indeed, red and blue spots are larger and more intense than for the remaining four promoters. Two bottom promoters, accA and accBC, show the least anisotropy, and the middle two, uvrD-P1 and uvrD-P2, are intermediate in that respect. For all the six promoters, the direction of the dipole moment varies in a sequence-dependent manner.

The data obtained suggest that, first, the promoter and coding regions have electrostatic potential greatly differing from that of periodic sequences. Secondly, the electrostatic potential differs with the type of promoters, mostly in the asymmetry of the distribution of positive and negative patches of the electrostatic

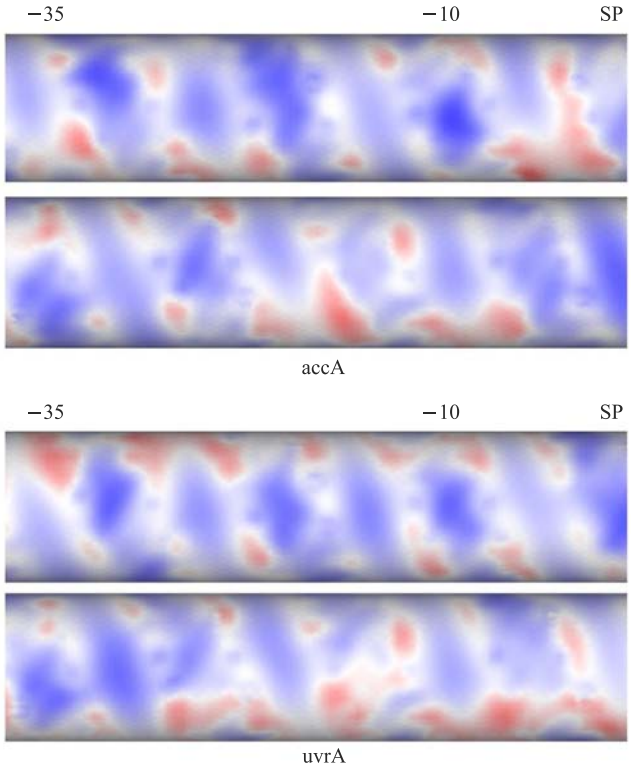


Fig. 3. Distribution of electrostatic potential around accA and uvrA promoter DNAs of *E. coli* from the –35 point to the transcription start point (denoted by SP) shown in two views differing by 180° rotation around the helix axis. The picture is scaled to show finer structure of the specified areas

potential. Finally, both the amplitude and the direction of the dipole moment across the DNA double helix change along the helix axis.

To show the finer structure in functionally important promoter areas (–35, –10, and starting point), the electrostatic potential distribution in those areas is presented for two promoters, accA and uvrA (Fig. 3), scaled to include those areas only. In the –35 area, a quasi-periodic potential distribution appears, in which red and blue spots are alternating. No explicit anisotropy of potential is observed in that region. In contrast, large areas of red and blue appear in the area from –10 to the starting point on the opposite sides of the cylinder, which suggests the anisotropy of the electrostatic potential in the vicinity of the starting point.

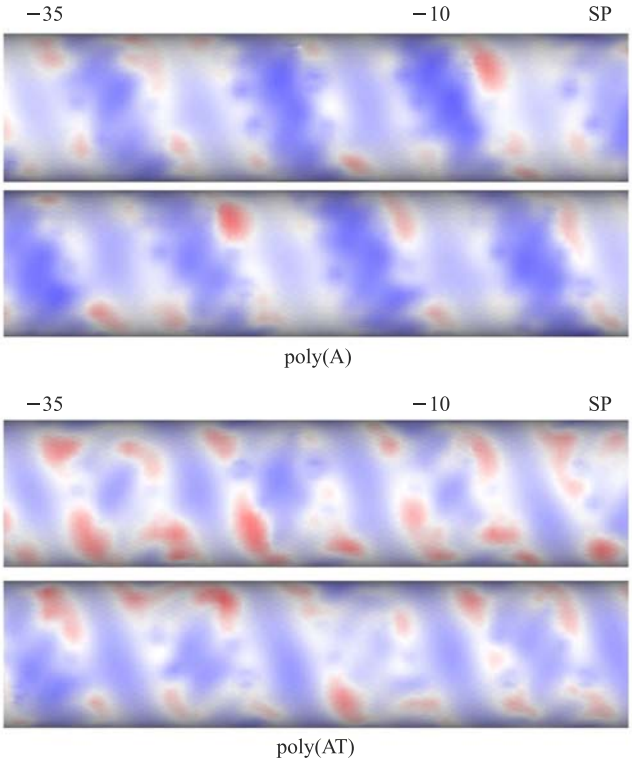


Fig. 4. Distribution of electrostatic potential around periodic DNA molecules poly(A) and poly(AT) in a fragment equal in length to fragments of promoter DNAs of *E. coli* from the –35 area to the transcription start point shown in two views differing by 180° rotation around the helix axis

The A/T tracks are known to occur more frequently in promoter sequences than in the full genome sequences, and to be distributed non-randomly in promoter sequences. In Fig. 4 the distribution of the electrostatic potential is presented around the periodic DNA sequences poly(A) and poly(AT) in a fragment equal in length to fragments of promoter DNA sequences of *E. coli* from the –35 area to the transcription start point. One can see that both the promoters accA and uvrA have the electrostatic potential distributions visually more similar to that of poly(AT) sequence than to that of poly(A). Thus, the electrostatic potential distribution of promoters appears to correlate with the contents and positions of A/T tracks along the promoter.

**3.5. Biological Significance of Promoter DNA Electrostatic Calculations.** Taken together, our data indicate that the electrostatic potential can improve the classification of DNA sequences by providing the physical basis. In such an

improved classification, the physical basis will be rendered for drawing structure–function relationships and evolutionary relationships between various DNA sequences, thus contributing to the development of bioinformatics.

Therefore, the entire body of data including the primary structure (sequence), secondary structure (geometry), and physical properties of specific DNA sequences will provide a unified basis for promoter classification, which is a key feature in understanding promoter functioning (transcription), their evolution and regulation.

## 4. CONCLUSIONS

Long and numerous DNA sequences present a «natural» field of application for distributed computations. Promoter sequences responsible for regulation of transcription of every gene require particular attention. Considering that classification of DNAs based on their electrostatic potential will require pairwise quantitative comparisons of those potentials, distributed computations are possibly the only choice because there are about 35,000 genes in mammalian genomes, and a promoter sequence is at least several hundred base pairs long for each of them.

## 5. APPENDIX

**5.1. Model Building, Dielectric Boundary and Charge Assignments.** All atom models of DNA fragments were constructed using the evaluation version of the HyperChem 7.01 package [21]. DNA was assumed to be in the B form. Charges were assigned to the center of each atom. The values of charges were taken from the AMBER force field [22]. Additional charges of –0.25 $q$ were assigned to O1 and O2 atoms of phosphate groups to allow for the well-known counterion condensation effect, which is retention of part of counterions near the charged atoms of the phosphate groups. Dielectric constants were taken to be 2 for the DNA interior and 80 elsewhere. Potential was visualized as a topological map on the surface of a cylinder with 15 Å radius centered at the longitudinal axis of DNA, about 5 Å away from DNA sugarphosphate backbone. Such a surface approximates the electrophoretic sliding surface of the DNA, at which the first stage of DNA–protein recognition is believed to occur.

**5.2. Solving the Poisson–Boltzmann Equation.** Calculations of electrostatic potential, $\varphi$, of DNA fragments were performed by solving the Poisson–Boltzmann equation, which describes the electrostatic potential in solvent around DNA molecule according to

$$-\bigtriangledown\left(\epsilon(\mathbf{r})\bigtriangledown\varphi(\mathbf{r})\right)=4\pi(\rho_0(\mathbf{r})+\rho_1(\varphi(\mathbf{r}))). \quad (1)$$

Here $\mathbf{r} = (x, y, z) \in R^3$, $\varphi$ is the sought electrostatic potential, $\epsilon$ is the dielectric permeability, and $\rho_0$ is the charge distribution of DNA described by

$$\rho_0(\mathbf{r}) = \sum_i e z_i \delta(|\mathbf{r} - \mathbf{r}_i|), \tag{2}$$

where $z_i$ is the charge of the $i$th atom of the molecule in units of elementary charge, $\mathbf{r}_i$ is the radius vector of the $i$th atom, $e$ is the elementary charge (the absolute value of the electron charge), $\delta$ is the Dirac delta function, and

$$\rho_1(\mathbf{r}) = \sum_i n_i e z_i \exp(e z_i \varphi / k_B T). \tag{3}$$

When the potential is small enough ($\varphi \ll k_B T/e$), (1) reduces to its linearized form

$$- \bigtriangledown (\epsilon(\mathbf{r}) \bigtriangledown \varphi(\mathbf{r})) + \kappa^2 \varphi = 4\pi \rho_0(\mathbf{r}), \tag{4}$$

where

$$\kappa^2 = 4\pi e^2 \sum_i n_i z_i^2 / k_B T \tag{5}$$

is the ions density, where $n_i$ is the concentration of ions of the $i$th kind, $z_i$ is the charge of ion of $i$th kind in units of the elementary charge, $k_B$ is the Boltzmann constant, and $T$ is the absolute temperature assumed to be 300 K.

Boundary condition for the potential $\varphi(\infty)$ is set using the Debye–Huckel approximation. For the purpose of numerical solution we restrict the infinite region to a big parallelepiped $\Gamma$, with the condition imposed on its surface, for $\mathbf{r} = (x, y, z) \in \Gamma$:

$$\varphi(\mathbf{r})|_\Gamma = \sum_i \frac{e z_i \exp(-\kappa |\mathbf{r} - \mathbf{r}_i|)}{|\mathbf{r} - \mathbf{r}_i|}. \tag{6}$$

The problems (4) with boundary condition (2) in the region $\Gamma$ are solved using the finite elements method. We solve the discrete linear system iteratively by the multigrid method (MM). The details of the MM solution algorithm are presented in Subsection 5.3.

In order to solve the nonlinear equation (1), we apply the iterations according to

$$- \bigtriangledown (\epsilon(\mathbf{r}) \bigtriangledown \varphi^{n+1}) + \alpha \varphi^{n+1} = 4\pi(\rho_0 + \rho_1(\varphi)) + \alpha \varphi^n, \tag{7}$$

where $\varphi^n$ is the approximation of solution corresponding to the $n$th iteration. Thus, the linear problem with unknown $\varphi^{n+1}$ has to be solved on each iteration. The solution to the problem (4) is used as an initial approximation for iterations (7).

**5.3. MM Solution Algorithm.** To solve the nonlinear equation (1), we apply the iterations by formula (7) where $\varphi^n$ is the solution approximation corresponding to the $n$th iteration.

To solve the problems (1), (2) with the boundary condition (6), we discretize the region $\Gamma$ with finite elements. The solution approximation is found in the finite dimensional space $S$ with the basis of finite element functions $\Phi^i(x, y, z)$:

$$\varphi = \sum_i u^i \Phi^i. \tag{8}$$

Applying the Galerkin approach [20] to (1), (7) and (8), we obtain a linear algebraic system of equations $Au = f$ from which coefficients $u_i$ can be found. Then we solve this system of linear algebraic equations iteratively by the multigrid method. The MM uses the sequence of nested finite element grids as follows:

$$h_{l-1} = 2h_l; \quad S_{l-1} \subset S_l; \quad l = 1, \ldots, L, \tag{9}$$

where $h$ is the grid step size, $S$ is the corresponding space of basis functions. The final solution should be found on the finest grid number $L$. It is performed by iterations using a set of auxiliary grids $l = 0, 1, \ldots, L-1$. On each iteration the problem is reduced to a smaller one on the grid $L - l$ for which the same algorithm is applied recursively until the grid number 0 is achieved. The grid 0 is the coarsest (with the biggest step size $h_0$). It contains a small number of unknowns. Thus, the linear system for that grid can be easily solved by any direct method, for instance, by Gauss elimination.

The algorithm of one MM iteration for the grid number $l$ is presented in Fig. 5.

Here the quantities with subscript $l$ are related to grid number $l$, the notation $u_0 = A^{-1} f_0$ means the direct solution procedure on the grid $l = 0$, $I$ is the interpolation operator that transfers functions from one grid to another, the *relax* is a simple iteration of Gauss–Seidel type to damp high-frequency residual components.

```
procedure mgm(l, u_l, f_l)
array u_l, f_l;
if l = 0 then u_0 = A^-1 f_0 else
begin integer j; array d,v;
u_l := relax(A_l, u_l, f_l);
d := I_l^l-1 (f_l - A_l u_l); v = 0;
for j :=l until γ do mgm(l - 1; u, d)
u_l := u_l + I_l^l-1 v;
u_l := relax(A_l, u_l, f_l);
end
```

Fig. 5. Algorithm of one MM iteration for the grid number $l$

The iteration process finishes when the residual norm satisfies the criteria:

$$\frac{\| f_L - A_L u_L \|}{\| f_L \|} \leqslant 10^{-6}. \qquad (10)$$

It usually takes 4–5 iterations to converge the iteration process.

**Acknowledgements.** The work was supported by the Russian Foundation for Basic Research, project 07-07-234.

## REFERENCES

1. *Cornell W. D. et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules // J. Am. Chem. Soc. 1995. V. 117. P. 5179–5197.

2. *Fedoseyev A. I. et al.* // Abstracts of the Intern. Conf. «Physique en Herbe92», Marseille, July 1992;
*Sivozhelezov V. S.* // Abstracts of the Intern. Conf. CSAM93, St. Petersburg, 1993. P. 121;
*Sivozhelezov V., Nicolini C.* // J. Theor. Biol. 2005. V. 234, No. 4. P. 47985.

3. *Koradi R., Billeter M., Wuthrich K.* MOLMOL: a Program for Display and Analysis of Macromolecular Structures // J. Mol. Graph. 1996. 14:515. P. 2932.

4. *Polozov R. V. et al.* Transfer RNAs: Electrostatic Patterns and Recognition by Synthetases and Elongation Factor EFTU // Biochemistry (bi0516733). 2005 (in press).

5. *Polozov R. V. et al.* On a Classification of *E. coli* Promoters According to Their Electrostatic Potentials // Part. Nucl., Lett. 2005. V. 2, No. 4(127). P. 82–90;
*Akishina T. P. et al.* Study of Electrostatic Potentials of DNA Promoters // XX Intern. Symp. on Nuclear Electronics & Computing (NEC'2005), Varna, Bulgaria, Sept. 12–18, 2005: Book of Abstracts. Dubna: JINR, 2005. P. 11.

6. *Romberg R. D.* // Trends Cell Biol. 1999. V. 9. P. M46–M49.

7. *Coleman R. A., Pugh B. F.* // J. Biol. Chem. 1995. V. 270. P. 13850–13859.

8. *Ohlendorf D. H., Matthew J. B.* // Adv. Biophys. 1985. V. 20. P. 137–151.

9. *Fogolari F. et al.* // J. Mol. Biol. 1997. V. 267. P. 368–381.

10. *Hsieh M., Brenowitz M.* // J. Biol. Chem. 1997. V. 272. P. 22092–22096.

11. *Jeltsch A. et al.* // EMBO J. 1996. V. 15. P. 5104–5111.

12. *Berkhout B., van Wamel J.* // J. Biol. Chem. 1996. V. 271. P. 1837–1840.

13. *Votavova H. et al.* // J. Biomol. Struct. Dyn. 1997. V. 15. P. 587–596.

14. *Carra J. H., Privalov P. L.* // Biochemistry. 1997. V. 36. P. 526–535.

15. *Guzikevich-Guerstein G., Shakked Z.* // Nature Struct. Biol. 1996. V. 3. P. 32–37.

16. *Strauss-Soukup J. K., Maher L. J., 3rd* // Biochemistry. 1998. V. 37. P. 1060–1066.

17. *Schellman J. A.* // Biopolymers. 1977. V. 16. P. 141–534.

18. *Parsons J. D.* // Comput. Appl. Biosci. 1995. V. 11. P. 603613.

19. *Guan X., Du L.* // Bioinformatics. 1998. V. 14. P. 783–788.

20. *Polozov R. V. et al.* // J. Biomol. Struct. Dyn. 1999. V. 16. P. 1135–1143.

21. http://www.hyper.com/products/description/hyper7.htm.

22. http://sigyn.compbio.ucsf.edu/amber/.

23. *Ozoline O. N. et al.* // Mol. Biol. 2002. V. 36. P. 682–688.

24. *Chasov V. V. et al.* // Biofizika. 2002. V. 47. P. 809–819.

Received on December 28, 2012.

Редактор *Е. И. Кравченко*

Подписано в печать 28.02.2013.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,12. Уч.-изд. л. 1,53. Тираж 205 экз. Заказ № 57929.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@jinr.ru
www.jinr.ru/publish/