Electrostatics as a Factor of Biomolecular Recognition in Processes of Transcription and Translation

T.P. Akishina[1], V.V. Ivanov[1], R.V. Polozov[2], V.S. Sivozhelezov[3,4]

[1]*Laboratory of Information Technologies, Joint Institute for Nuclear Research, 141980, Dubna, RUSSIA*

[2]*Institute for Theoretical and Experimental Biophysics, Moscow region, Puschino, 142290, RUSSIA*

[3]*Institute of Cell Biophysics, Moscow region, Puschino, 142290, RUSSIA*

[4]*Nanoworld Institute, University of Genoa, 30 Corso Europa, Genoa 16132 ITALY*

General. Major cellular processes such as gene transcription and translation, signal transduction, electron and ion transport, cell motion, a plethora of regulatory mechanisms and much more, 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 non-linearity, 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 both in terms of 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.

State of the art. 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 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 below-described **strategy for biomolecular recognition research:**

- 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, statistical 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, electrostatic 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.

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 charged 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 sub-physiological 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 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 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 eucaryotic 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).

Choice of objects of investigation. Primarily, these 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. The other direction is structural defects in proteins that are known to result in formation of improper protein aggregates and, consequently, to developing severe. Such proteins include steroid-metabolizing cytochromes P450 from which malfunctioning leads to atherosclerosis, and amyloid precursor proteins responsible for the so called conformational diseases, including severe neural degradation diseases like Alzheimer, Parkinson, prion diseases, etc. For all objects, we investigate structures of molecular complexes, parameters of kinetics and thermodynamics of their formation, dependence of those parameters on characteristics of the medium, and evolutionary aspect of complex formation.

RESULTS AND DISCUSSION

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.

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 basing 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 to propose a mechanism of electric charge switching on the corresponded 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.

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 element, 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 anti-sigma factors and released only by external signal, show very different electrostatic potentials at their contact regions with anti-sigma, which suggests different mechanisms of their interactions with the respective anti-sigmas. For one of the remaining two sigmas, which require no external anti-sigma 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 indicate 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.

Electrostatics of bacterial promoters. Origin, evolution, function, and regulation of promoter DNA are presently analyzed basing 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 basing 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.

CONCLUSIONS

Physicochemical basis of structure and function of biological systems is comprised by biomolecular recognition resulting in complex formation. In case of pathology, recognition is disturbed which causes unordered association termed aggregation (the latter including fibril formation as a particular case). The current opinion in biology is that basic principles underlying the widely diverse biological objects are quite similar and valid for various levels of organization, from recognition and aggregation of molecules and cells to populations of organisms, the problem of organism aggregation being central in theories of ecology and evolution. Therefore solving recognition and aggregation tasks in *vitro* as done herein is often necessary for solving similar tasks in *vivo*. Recognition and aggregation, when treated as physicochemical problems, are quite difficult to pursue due to the

inherent features of such problems such as their multicomponent nature, complexity of critical behavior, large distinctions in spatiotemporal scales, balance of several types of interactions, and large degree of anisotropy.

The novel theoretical and computational methods and tools being developed herein will allow more adequate description of processes of recognition and aggregation, and the ensuing applications, including studies of pathogenesis of fibril formation, development and construction of protein microarrays, and *de novo* protein design.

References

[1] Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M. Jr., Ferguson D.M., Spellmeyer D. C., Fox, T., Caldwell J. W., and Kollman P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules, J. Am. Chem. Soc. 1995 117, 5179-5197.

[2] Fedoseyev A.I., Purtov S.V., Sivozhelezov V.S. et al. Abstracts of the International Conference "Physique en Herbe-92", Marseille, July 1992; Sivozhelezov V.S. Abstr. Int. Conf. CSAM-93, St. Petersburg, 1993, p.121; Sivozhelezov V, Nicolini C. J Theor Biol. 2005 234(4):479-85.

[3] Koradi R., M. Billeter, and K. Wuthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51-5, 29-32.

[4] R.V. Polozov, M.Montrel, V.V. Ivanov, Yu. Melnikov, V.S. Sivozhelezov. Transfer RNAs: Electrostatic Patterns and Recognition by Synthetases and Elongation Factor EF-TU". Accepted by Biochemistry (bi0516733), 2005.

[5] R.V.Polozov, V.S.Sivozhelezov, V.V.Ivanov and Yu.B.Melnikov. On a Classification of *E.coli* Promoters According Their Electrostatic Potentials, Particles & Nuclei, Letters, 2005, Vol. 2, No. 4(127), pp. 82-90; T.P.Akishina, V.V.Ivanov, R.V.Polozov and V.S.Sivozhelezov. Study of Electrostatic Potentials of DNA Promoters, In: XX International Symposium on Nuclear Electronics & Computing (NEC'2005), Varna, Bulgaria, September 12-18, 2005: Book of abstracts, Dubna, JINR, 2005, p. 11.