ELECTROSTATIC PROPERTIES OF *E.COLI* GENOME DNA

Kamzolova S.G., Sorokin A.A., Dzhelyadin T.R., Beskaravainy P.M.*, Osypov A.A.

Institute of Cell Biophysics of RAS, Pushchino, Moscow region, Russia
* Corresponding author: e-mail: bpm@icb.psn.ru; kamzolova@icb.psn.ru

Keywords: *E. coli genome, promoters, coding region, electrostatic potential distribution*

Summary

Motivation: Distribution of electrostatic potential around nucleotide sequences is one of fundamental characteristics of DNA contributing to its recognition by DNA-binding proteins. Analysis of electrostatic properties of natural DNAs had to await the development of appropriate calculation methods for long nucleotide sequences. A method recently proposed in our work (Kamzolova *et al*., 2000) satisfies the requirement thus opening a promising means for studying electrostatic properties complete genomes and their different regions. Here, the method was used for analysis of electrostatic potential of *E. coli* genome.

Results: Distribution of electrostatic potential of the complete sequence of *E. coli* genome was calculated. It is found that DNA is not a uniformly charged molecule. There are some local inhomogeneities in its electrostatic profile which correlate with promoter sequences. These characteristic variations of electrostatic potential of DNA may be involved in RNA-polymerase-DNA recognition.

Availability: Electrostatic potential distribution analysis software is available at request to academic users (lptolik@icb.psn.ru).

Introduction

One of fundamental physico-chemical characteristics of any macromolecule, which affects its interaction with different ligands, is the distribution of electrostatic potential around it. DNA is a highly charged polyelectrolite and therefore its electrostatic potential may be one of the main features recognized by DNA-binding proteins. Really, electrostatic interactions between promoter DNA and *E. coli* RNA-polymerase ($E\sigma^{70}$) has been recently shown to be of considerable importance in regulating promoter function (Kamzolova *et al*., 2000). Electrostatic characteristics of promoter DNA have been suggested to be a new promoter determinant marked by its relative independence from promoter nucleotide sequence. The important role of electrostatic interactions in the multi-step process of protein-DNA recognition has also been shown for some other DNA-binding proteins (Misra *et al*., 1994; Fogolary *et al*., 1997; Labeots, Weiss, 1997) It is not surprising, then, that theoretical analysis of electrostatic potential distribution around DNA molecule and its fragments containing protein-binding sites is one of the pressing areas in modern research of protein-DNA recognition coding. Recently we have proposed a simplified method for calculation of electrostatic potential distribution for long nucleotide sequences (Kamzolova *et al*., 2000; Sorokin, 2001). Here the method was used for theoretical analysis of electrostatic properties of the complete sequence of E. coli genome. As an example to illustrate a functional meaning of the information hidden away in the electrostatic map of the genome, a comparative analysis of electrostatic patterns of promoter and nonpromoter DNA sites was made.

Methods

The complete sequence of *E. coli* K-12 genome was taken from GenBank (accession number U00096). The electrostatic potential around double-helical DNA molecule was calculated by the Coulombic method (Kamzolova *et al*., 2000) using the computer program of Sorokin A. (Sorokin, 2001).

Results and Discussion

When considering the prospects in studying electrostatic properties of a complete molecule of natural DNAs it should be emphasized that until recent years such analysis has been hampered due to considerable difficulties of theoretical calculations of electrostatic potentials for long DNA fragments. An accurate calculation can be carried out only for short DNA sequence not more then 30–40 nucleotide pairs using nonlinear Poisson – Boltzman equation (Jayaram *et al*., 1989). Recently we have proposed a simplified method for calculation of electrostatic potential distribution based on Coulomb's law which can be used for long fragments of double – helical DNA including complete genomes (Kamzolova *et al*., 2000; Sorokin, 2001). Though the method cannot be applied for rigorous treatment, it is quite suitable for qualitative analysis of electrostatic map of a natural DNA molecule and for comparative studies of its different parts.

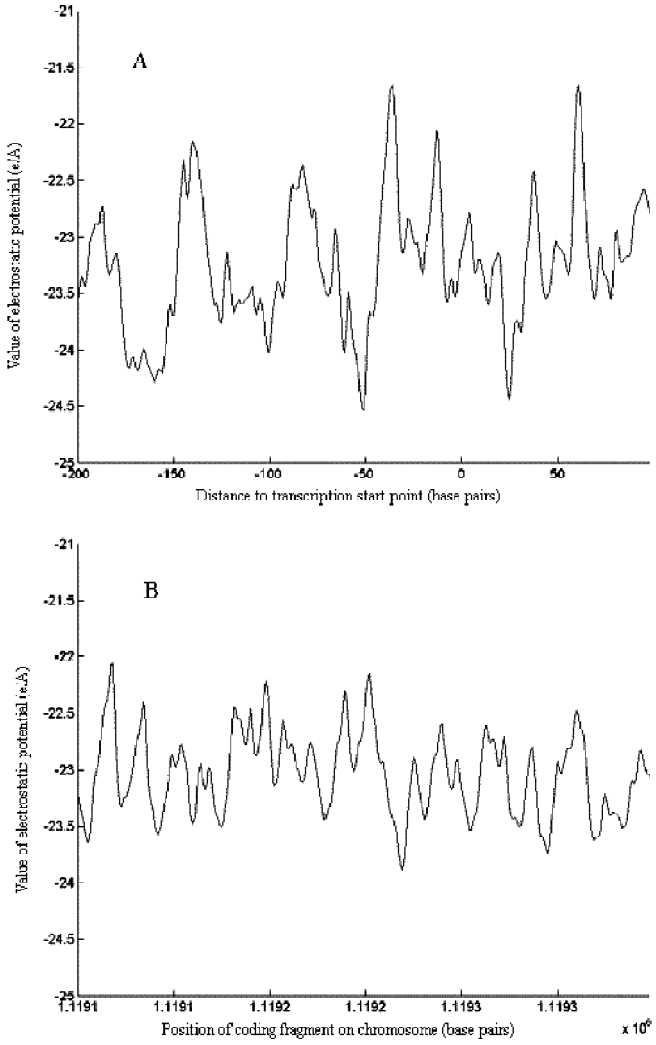

Fig. Electrostatic potential distribution for promoter (A) and coding region (B) in *E. coli* genome.

Using this method, calculation of electrostatic potential distribution was performed for complete nucleotide sequence of *E. coli* genome containing 4639221 base pair. The results obtained representing the profile of electrostatic potential distribution around the complete genome are in our database (kamzolova@icb.psn.ru)

The possibility of extracting some functional information from the electrostatic map of *E. coli* genome can be demonstrated by the example provided by a large-scale analysis of electrostatic patterns of promoter and nonpromoter DNA sites. Electrostatic profiles of 359 promoters identified in *E. coli* genome as well as of their nearby coding sequences were analyzed by the presence of peaks and valleys as well as by their arrangement and values. Figure shows some representative examples of electrostatic patterns for promoters (A) or DNA coding regions (B). It is found that coding regions are characterized by more homogeneous distribution of electrostatic potential, whereas local inhomogeneities with the most positive and negative areas correspond to promoter sites. It should be noted that individual promoters discussed here vary in the design of their electrostatic profiles but all of them, in contrast to coding regions, are characterized by inhomogeneous complex-shaped patterns. These characteristic variations of electrostatic potential of DNA may be related to RNA polymerase – DNA recognition by specifying promoter sites as electrostatic traps or barriers for the enzyme. In addition, alternating areas of negative and positive potential in promoter sites may enforce charged RNA polymerase molecules to orient properly relative to the transcription start point.

Thus, DNA electrostatic component may be one of the determining factors allowing RNA polymerase to identify promoter sites in genomes.

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (grants: RFBR naukograd-04-04-97275 and RFBR 04-04-49635).

References

Fogolari F., Elcock A.H., Esposito G., Viglino P., Briggs J.M., McCammon J.A. Electrostatic effects in homeodomain-DNA interactions // J. Mol. Biol. 1997. V. 267(2). P. 368–381.

Kamzolova S.G., Sivozhelezov V.S., Sorokin A.A., Dzhelyadin T.R., Ivanova N.N., Polozov R.V. RNA polymerase—promoter recognition. Specific features of electrostatic potential of “early” T4 phage DNA promoters // J. Biomol. Struct. Dyn. 2000. V. 18(3). P. 325–334.

Labeots L.A., Weiss M.A. Electrostatics and hydration at the homeodomain-DNA interface: chemical probes of an interfacial water cavity // J. Mol. Biol. 1997. V. 269(1). P. 113–128.

Misra V.K., Hecht J.L., Sharp K.A., Friedman R.A., Honig B. Salt effects on protein-DNA interactions. The lambda cI repressor and EcoRI endonuclease // J. Mol. Biol. 1994. V. 238(2). P. 264–280.

Sorokin A.A. Functional analysis of E. coli promoter sequences. New promoter determinants. Ph.D. Thesis. Pushchino, Institute of Theoretical and Experimental Biophysics RAS. 2001.