Why are genome regulatory regions AT-enriched?

Alexander A. Osypov

Laboratory of Cell Genome Functioning, Institute of Cell Biophysics of RAS, 142290, Pushchino MR, Russia, aosypov@gmail.com

Genome DNA physical properties define its shape in the functional space and influence its interactions with proteins, esp. for transcription regulation. DNA is highly charged and electrostatics contributes greatly to the subject. DEPPDB was developed to provide all available information on these properties of genome DNA combined with its sequence and annotation of biological and structural properties of genome elements and whole genomes, organized on a taxonomical basis.

DEPPDB and its tools [1,2] (deppdb.psn.ru or electrodna.psn.ru) were used to carry out the analysis.

Electrostatic potential is distributed non-uniformly along DNA molecule and correlates, though not corresponds exactly, with GC content, strongly depending on the sequence arrangement and its context (flanking regions). Binding frequency of RNA polymerase to DNA along the genome, measured in direct experiment, correlates to the calculated electrostatic potential.

Transcription regulation areas have electrostatic potential and other physical properties peculiarities. Binding sites of transcription factors of different protein families in different taxa are located in long areas of high electrostatic potential. Electrostatics distribution on transcription factors protein molecule surface reflects the DNA electrostatics of its binding sites. Promoters in average have high value and heterogeneity of electrostatic potential profile. The transcription starting sites of prokaryotic genomes are characterized by extensive (hundreds of bp) zone of high electrostatic potential and some peculiarities directly around TSS. This is associated with protein binding and formation of physical properties due to transcription machinery. Specific details of the TSS electrostatic potential architecture are similar in related taxa. Promoters up-element demonstrates electrostatic nature. High AT content is ofthen associated with up-element, however no text consensus is discovered and functionality of this region is defined by its physical properties. Strong T4, early T7-like, phage Lambda and E.coli ribosomal promoters with pronounced up-element have high levels of the electrostatic potential within it, and electrostatics is responsible for its functioning during the global transcription switch under the T4 bacteriophage infection accompanying reversion of the polymerase alpha subunit local charge.

Electrostatic effects on genome functioning interact with other physical properties of DNA, in particular - bending, thermal stability, supercoiling. They may interact with electrostatics in both, formation, and transcription regulation. The distribution of curved DNA in promoter regions is evolutionarily preserved, and is mainly determined by temperature of habitat. Mesophilic genomes may have different intensity in curvature, while thermophiles and hyperthermophiles lack it overall because of the life under temperature above the curvature-relaxing point that renders this property useless in transcription regulation. Magnitude of DNA curvature is related to AT content, it was shown that strongly curved DNA fragments

must possess high A+T content (reverse is not true). There is no decrease in size and prominence of electrostatic deep in extremophyles, proving importance of electrostatics and its differential role vs curvature.
Mycobacterium leprae is intracellular parasitic bacterium that causes leprosy in human in tropical countries. It has undergone massive genome reduction over time as a result of its parasitic nature, discarding more than half its genes with many important metabolic activities and their regulatory circuits. Twenty seven percent of the M. leprae genome consists of pseudogenes that have functional counterparts in M. tuberculosis. Electrostatic properties of Mycobacterium leprae genes starts reflect massive pseudogenization and strictly intracellular parasitic life with reduced transcription regulation. It has smoothed electrosatic profile compared to its close relatives and far less pronounced increase of electrostatics over upstream region where extended electrostatic deep is commonly found. Moreover, the genes remained functional overall possess even less electrostatic typical features, reflecting the deminishing need for extensive transcriptional regulation. Obviously the lost genes needed more diverse regulation with more prominent role of electrostatics.
Electrostatics plays important and universal role in transcription regulation in prokaryotes, affecting proteins binding probability and positioning accuracy. It may influence horizontal gene transfer, transcription regulation systems evolution and contribute to genome regulatory regions high AT content.
Physical properties formation principles affect such fundamental problems as Chargaff's II rule, redundancy of the genetic code, neutrality of synonymous substitutions; and justify the fundamental idea of DNA phenotype, defining the new principle of biophysical bioinformatics.
The authors are grateful to IMPB RAS for hosting the DB.

1. A. A. Osypov, G.G. Krutinin, S. G. Kamzolova (2010) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA, *J Bioinform Comput Biol,* **8(3):** 413-425.
2. A. A. Osypov, G.G. Krutinin, E.A. Krutinina, S. G. Kamzolova (2012) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA elements, *J Bioinform Comput Biol,* **10(2):** 1241004.
3. Kamzolova, A.A. Sorokin, T.D. Dzhelyadin P.M., Beskaravainy, A.A. Osypov (2005) Electrostatic potentials of E. coli genome DNA, *J Biomol Struct Dyn,* **23(3):** 341-346.

4. A. A. Osypov, G. Krutinin, E. Krutinina, S. Kamzolova (2014) Electrostatic properties of natural DNA palindromes – transcription factors binding sites and terminators, *FEBS Journal,* **281(s1):** 681.

5. E. Krutinina, G. Krutinin, S. Kamzolova, A. A. Osypov (2014) Electrostatics of prokaryotic transcription factors match that of their binding sites, *FEBS Journal,* **281(s1):**. 682.

6. A. A. Osypov, G. Krutinin, E. Krutinina, S. Kamzolova (2014) New in DEPPDB – a portal for electrostatic and other physical properties of genome DNA and its elements, *FEBS Journal,* **281(s1):** 631.
7. A. A. Osypov (2014) DNA electrostatics: its role in transcription regulation in prokaryotes, a new natural selection factor, and a base for DNA phenotype and biophysical bioinformatics, *FEBS Journal,* **281(s1):** 680.
8. E. A. Krutinina, G. G. Krutinin, S. G. Kamzolova, A. A. Osypov (2011) New Insights into Protein-DNA Electrostatic interactions: Beyond Promoters to Transcription Factors Binding Sites, *J Biomol Struct Dyn,* **28(6):** 1137-1138.