New and Notable

Mechanisms of Protein Binding to DNA: Statistical Interactions are Important

Anatoly B. Kolomeisky*
Department of Chemistry, Rice University, Houston, Texas

Protein molecules binding to DNA initiate sequences of biochemical transitions that control and regulate all major processes in living cells (1). It has long been understood that the efficient transfer of genetic information is possible when genome-regulating protein molecules are found to be strongly bound to special sequences on DNA, known as specific binding sites. It was later realized that finding these target sequences might be complicated by the existence of a large number ($\simeq 10^6$–10^9 per DNA) of other nonspecific binding sites and because of low concentrations of relevant protein molecules (2–4). Although significant experimental and theoretical efforts to investigate mechanisms of protein binding to DNA have been made, the role and functioning of nonspecific interactions has not been fully understood.

Contemporary theoretical views of protein-DNA interactions indicate that there are two main components of binding forces. One of them is purely electrostatic attraction between oppositely charged DNA and protein molecules that are mostly sequence-independent (2–4). It has been suggested that another contribution comes from particular DNA sequence motifs that strengthen the attraction of protein molecules (2–4). The microscopic origins of this increased affinity are probably due to a combination of van der Waals, hydrogen, covalent, and steric interactions, as well as electrostatic charge patterns recognition (5,6). However, this theoretical picture cannot explain recent high-resolution experimental measurements of genome-wide distribution and binding preferences for protein molecules from transcription preinitiation complexes (7) and for nucleosomes (8). The known protein-DNA binding motifs do not correlate with the obtained distributions of association positions. This suggests that there are additional factors that control and regulate the spatial organization and affinity of protein attachments to DNA. An article by Afek and Lukatsky (9) in this issue of *Biophysical Journal* presents a new possible mechanism of protein-DNA bindings.

Afek and Lukatsky suggest that there is an additional statistical interaction potential between protein and DNA molecules (10). The source of this interaction is due to the specific structure and symmetry of DNA sequences to which the protein molecule binds. Specifically, they have shown that DNA sequences with repeated homogeneous segments (consisting of only dA:dT or dC:dG units) have a stronger affinity for association to DNA-binding proteins, while more heterogeneous sequences are attracted less to the same proteins. The origin of this phenomenon is purely entropic. For DNA sequences with homogeneous tracts there is, on average, a higher probability to have the segment that will more strongly attract the protein molecule. For the more heterogeneous sequences, binding energies are typically close to some average value which is always higher than the most attractive one, leading to weaker protein-DNA associations.

Afek and Lukatsky (9) have quantified the action of this statistical potential by calculating a free-energy landscape for the whole genome and comparing it with experimentally measured binding preferences for several transcription factors (7). An excellent agreement between theoretically predicted and experimentally measured distributions of protein binding preferences has been found. The result of this simple theoretical model is impressive because the only experimental input parameter is the actual sequence of the yeast genome. The strength of the statistical potential has been estimated to be close to 2–3 kcal/mol per each protein, which provides an additional attraction for protein molecules to attach to DNA. One of the most striking results from this theoretical method is the ability to discriminate between different binding preference distributions for different genes. It is important to note that the statistical interaction is a global effect which still has a strong effect locally by modulating the binding affinity at given position.

Although the theoretical picture of protein-DNA binding mechanisms presented by Afek and Lukatsky (9) provides a significant advancement in our understanding of fundamental processes that control and regulate biological systems, it still has many issues and many questions remain unanswered. This theoretical method is built on a thermodynamic quasiequilibrium description of protein binding to DNA, while real cellular systems are highly dynamic and nonequilibrium. It also ignores the complex structure and multiple conformations of DNA and protein molecules that should significantly modify the size of regions accessible for protein-DNA bindings as well as interaction potentials. In addition, this method still does not explain the fundamental issue of how proteins find specific target sites so fast and efficiently (11). It is known that the facilitated target search involves a combination of three-dimensional bulk diffusion and one-dimensional hopping along the DNA (3); however, the role of nonspecific interactions here remains controversial and is not well explained (11). Furthermore, different cellular concentrations of DNA-binding proteins are

Submitted December 26, 2012, and accepted for publication January 15, 2013.

*Correspondence: tolya@rice.edu

Editor: Reka Albert.

0006-3495/13/03/0966/2 $2.00

http://dx.doi.org/10.1016/j.bpj.2013.01.039

not taken into account. Despite these shortcomings, the work by Afek and Lukatsky is a large step forward in our understanding of fundamental issues in protein-DNA interactions, which also provides a quantitative guidance for future experimental studies. In addition, it is an excellent example of how the theoretical method must be utilized for analyzing complex biological phenomena.

REFERENCES

1. Lodish, H., A. Berk, …, J. Darnell. 2000. Molecular Cell Biology, 4th Ed. W.H. Freeman, New York.
2. von Hippel, P. H., A. Revzin, …, A. C. Wang. 1974. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc. Natl. Acad. Sci. USA.* 71:4808–4812.
3. Berg, O. G., R. B. Winter, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry.* 20:6929–6948.
4. von Hippel, P. H., and O. G. Berg. 1986. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA.* 83:1608–1612.
5. Cherstvy, A. G., A. B. Kolomeisky, and A. A. Kornyshev. 2008. Protein-DNA interactions: reaching and recognizing the targets. *J. Phys. Chem. B.* 112:4741–4750.
6. Cherstvy, A. G. 2011. Electrostatic interactions in biological DNA-related systems. *Phys. Chem. Chem. Phys.* 13:9942–9968.
7. Rhee, H. S., and B. F. Pugh. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature.* 483: 295–301.
8. Yuan, G. -C., Y. -J. Liu, …, O. J. Rando. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science.* 309:626–630.
9. Afek, A., and D. B. Lukatsky. 2013. Genome-wide Organization of Eukaryotic Pre-initiation Complex is Influenced by Nonconsensus Protein-DNA Binding. *Biophys. J.* 104:1107–1115.
10. Sela, I., and D. B. Lukatsky. 2011. DNA sequence correlations shape nonspecific transcription factor-DNA binding affinity. *Biophys. J.* 101:160–166.
11. Kolomeisky, A. B. 2011. Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* 13:2088–2095.