J|A|C|S
COMMUNICATIONS

Published on Web 01/03/2007

Fly-Casting in Protein−DNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition

Yaakov Levy,*,†,‡ José N. Onuchic,† and Peter G. Wolynes†

Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093, and Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received August 14, 2006; E-mail: koby.levy@weizmann.ac.il

The remarkable efficiency and specificity of protein−DNA recognition presents many theoretical puzzles given the size of the genome and the large number of molecular species in vivo. Another aspect of protein−DNA interactions, not less perplexing than the high specificity of recognition, is the rapidity with which the DNA target sequence is recognized. Since the pioneering work of von Hippel,[1] it has become well accepted that a protein's search for its target sequence is comprised of both one-dimensional search (sliding) and three-dimensional search (hopping) as well as transfer between DNA segments separated in sequence. The combination of these mechanisms is key to the high efficiency of the search.[1,2] The molecular and physical principles that govern the direct and indirect readout of DNA sequences and their interplay with nonspecific interactions, however, remain unclear.

Many studies of protein−DNA complexes acknowledge the intrinsic flexibility of proteins and nucleic acids because both molecules often undergo conformational changes prior to or upon their recognition. DNA conformational deformations, for example, have been thought to contribute to binding specificity, selectivity, and affinity.[3,4] Protein conformational changes can discriminate between specific and nonspecific binding.[5] Protein disorder can, in principle, facilitate the diffusive search through the "fly-casting" mechanism[6] where a flexible region of the protein partially and nonspecifically binds to a DNA sequence. Indeed, several DNA binding proteins are known to have partially unstructured structures in the unbound state, and the unstructured regions fold upon binding to the target.[7,8] An insightful analysis suggests induced folding by DNA binding, as reflected by the sequential binding of a dimeric transcription factor as unfolded monomers that fold and dimerize on the DNA,[9] is vital in providing both rapid and specific assembly.[10]

In this study, we investigate how electrostatic steering and protein flexibility in DNA recognition synergistically couple in DNA binding. The electrostatic field guides the protein to the DNA (as the interface of protein−DNA complexes is highly charged and complementary) and thus accelerates binding rate. Protein flexibility also facilitates binding via the fly-casting mechanism.[6,11] The electrostatic field introduced by nucleic acids or other polyanions was previously reported to stimulate the folding of Arc-repressor.[12] We show also that the "tidal force" coming from the electrostatics actually facilitates formation of an intermediate state due to partial unfolding that is needed to carry out fly-casting. As a case study, we explore the assembly mechanism of the complex formed between the Ets domain of SAP-1 and its specific DNA sequence (see Figure 1; the Ets domain is stable and structured even in the absence of DNA).[13]

We performed molecular dynamics simulations on the complex between the Ets protein and its specific DNA sequence using a coarse-grained model that corresponds to a funneled energy landscape. To encode the specificity and the minimal frustration of most of the protein's interactions, the energy function is based on the native topology but is supplemented by Coulomb interactions, which are not specific. The protein can diffuse and undergo folding and unfolding processes in a sphere of radius 40 Å centered at the center of mass of the DNA that is kept frozen. Previously, native topology-based models have successfully been applied to describe the physics of protein folding and association.[11] The present model does not include DNA conformational fluctuations. For the Ets−

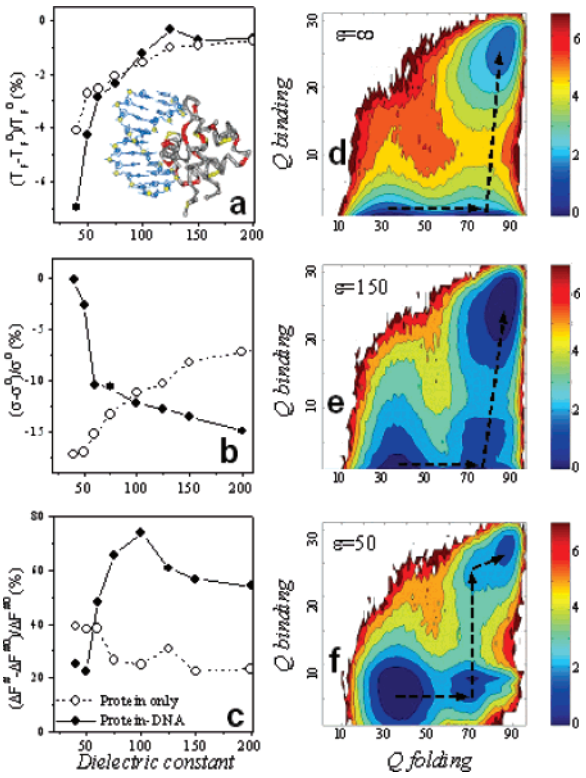

Figure 1. DNA effects on the folding thermodynamics and kinetics. The Ets protein is destabilized by the DNA as the electrostatic forces become stronger (a). The folding reaction becomes less cooperative (b) and its barrier decreases (c). T_F^0, σ^0, and $\Delta F^{\#0}$ are the folding values of the isolated Ets domain based on the native structure-based model without electrostatics. The inset in (a) shows the crystal structure of the complex between the Ets protein and its specific DNA sequence (pdb entry 1bc8). In the coarse-grained model, each residue is represented by a single bead which is positively charged for lysine and arginine (red spheres), negatively charged for glutamic and aspartic acids (yellow spheres), or neutral (gray spheres). Each DNA nucleotide is modeled by three beads: neutral beads at the center of the base and the sugar groups, and a negatively charged bead for the phosphate group. (d−f) Free energy surfaces for protein−DNA assembly. The free energy is shown as a function of $Q_{folding}$ and $Q_{binding}$ (the total number of native contacts in the protein and the protein−DNA interface) for three values of dielectric constant ($\epsilon = \infty$, 150, 50). Larger coupling between folding and binding is seen for stronger electrostatic forces (smaller dielectric constant). Even when the electrostatic forces are relatively weak, protein flexibility is essential for DNA recognition: the protein is bound to the DNA when it is partially folded and completes its folding when it is bound.

† University of California at San Diego.
‡ Weizmann Institute of Science.

10.1021/ja065531n CCC: $37.00

DNA system, the free energy difference between the specific binding (electrostatic and specific contact interactions) and the nonspecific binding (electrostatic interactions alone) is about $3k_BT$.

The electrostatic field of the DNA affects the thermodynamics and kinetics of the folding of the Ets protein. Figure 1a shows that decreasing the dielectric constant decreases the protein stability. For both the isolated protein and in the vicinity of DNA, a shift of the peak of the specific heat curves to lower temperatures is observed when increasing the electrostatic strength. Increasing the strength of the electrostatic forces destabilizes the isolated protein due to frustration between nonspecific electrostatic interactions and the specific native interactions (for SH3 domain, barnase, and barstar, non-DNA binding proteins, we have observed a milder destabilization effect due to electrostatics). In the presence of DNA, a larger destabilization effect is seen as a result of competition between folding and binding. This latter frustration reflects the tidal forces from electrostatics pulling the molecule apart. Under strong electrostatic forces, the negatively charged DNA pulls the protein, and protein−DNA association can therefore advance protein folding. The electrostatic interactions affect the degree of folding cooperativity, as well, as expected for a frustrated system. Increasing the strength of the electrostatic interactions in the presence of DNA results in a less cooperative folding as is evident from the broadening of the specific heat profile when decreasing the dielectric constant (Figure 1b). The broader specific heat profiles for strong electrostatics are due to a larger steering effect of the protein to the DNA (in the absence of DNA, on the other hand, decreasing the dielectric constant results in a higher cooperative folding presumably because of strengthening the five native salt bridges). Along with the decrease of the folding cooperativity, the kinetic folding barrier decreases as well when protein folding occurs around a DNA molecule (Figure 1c).

The destabilization of the Ets protein and the decrease in its folding barrier with increasing electrostatic strength is a manifestation of the fly-casting mechanism in protein−DNA association. The electrostatic forces between the negatively charged DNA and the positively charged binding site of the protein facilitate nonspecific association even when the protein is not completely folded. The protein, thus, is steered to the DNA and remains attached due to the attractive electrostatic interactions. The coupling between protein folding and its binding to DNA is shown in Figures 1d−f for three different strengths of electrostatic forces. With increasing the electrostatic strength, the protein is found to be more bound to the DNA as a partially folded structure and completes its folding together with binding at the specific binding site. These plots also demonstrate how the folding mechanism for smaller dielectric constants involves more intermediates, which give rise to the broader specific heat curves. When the protein−DNA contacts are modeled as short-range Lennard-Jones interactions (i.e., infinite dielectric constant), a fully folded protein is prerequisite for specific DNA recognition.

The coupling between folding and protein−DNA assembly, even for weak electrostatic forces, changes the capture radius of the target as shown by the free energy plotted as a function of the separation distance between the protein and the DNA (Figure 2). A sharper decrease of the free energy is observed when flexibility and electrostatics are both included in contrast to when electrostatic forces alone guide a rigid protein. When electrostatic forces are turned off, the barrier for binding arises either for rigid or flexible protein. It would be interesting to test this idea by studying the recognition of peptide nucleic acids with appropriate proteins since they lack these strong electrostatic forces.[14]

Protein flexibility acts in concert with electrostatic long-range forces to maximize the efficiency of binding. Modulating the unfoldedness of proteins is important within the cell.[15] The involvement of

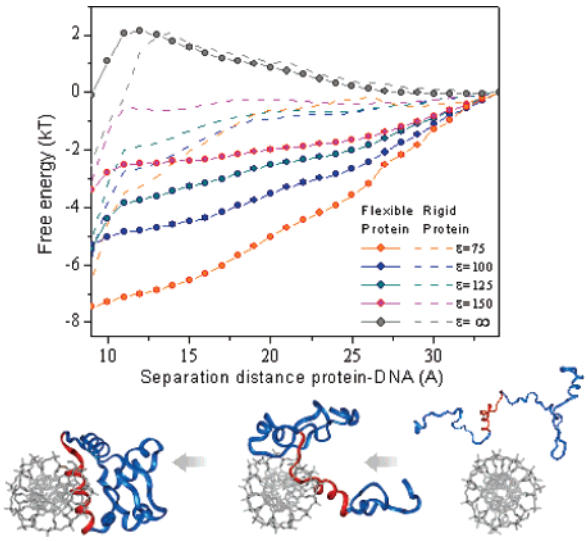

Figure 2. The interplay between fly-casting and electrostatic steering in protein−DNA recognition. The Ets protein is modeled as a flexible polymer that can fold and unfold (circles) and as a rigid folded protein (dashed lines). The free energy of the protein−DNA recognition is shown as a function of the separation distance between the center of mass of the protein and DNA. The DNA recognition by a rigid protein is governed by the electrostatic forces, but when protein flexibility is introduced, fly-casting effects contribute as well. A much more significant decrease of the free energy is observed when protein flexibility is allowed, showing the significance of protein plasticity and unfolding in DNA recognition. Without electrostatics, a barrier appears for both rigid and flexible protein models. The degree of fly-casting increases with the electrostatic strength, suggesting a coupling between electrostatic forces and fly-casting effects in protein−DNA binding.

fly-casting effects in protein−DNA recognition will vary from system to system primarily by the protein folding kinetics, the protein structural plasticity, and the electrostatic screening strength.

Acknowledgment. This work has been funded by the NSF sponsored Center for Theoretical Biological Physics (Grants PHY-0216576 and 0225630) with additional support from MCB-0543906 as well as the National Institutes of Health Grant 5R01GM44557.

Supporting Information Available: Additional computational models and analysis details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Berg, O. G.; Winter, R. B.; von Hippel, P. H. *Biochemistry* **1981**, *20*, 6929−6948.
(2) Halford, S. E.; Marko, J. F. *Nucleic Acids Res.* **2004**, *32*, 3040−3052.
(3) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−11168.
(4) Zhang, Y.; Xi, Z.; Hedge, R. S.; Shakked, Z.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8337−8341.
(5) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. *Science* **2004**, *305*, 386−389.
(6) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−8873.
(7) Spolar, R.; Record, M. *Science* **1994**, *263*, 777−784.
(8) Liu, J. G.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2006**, *45*, 6873−6888.
(9) Kohler, J. J.; Metallo, S. J.; Schneider, T. L.; Schepartz, A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11735−11739.
(10) Slutsky, M.; Mirny, L. A. *Biophys. J.* **2004**, *87*, 4021−4035.
(11) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 511−516.
(12) Rentzeperis, D.; Jonsson, T.; Sauer, R. T. *Nat. Struct. Biol.* **1999**, *6*, 569−573.
(13) Mo, Y.; Vaessen, B.; Johnston, K.; Marmorstein, R. *Mol. Cell* **1998**, *2*, 201−212.
(14) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. *Nature* **1994**, *368*, 561−563.
(15) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.

JA065531N