A Free Energy Pathway for the Interaction of the SRY Protein with Its Binding Site on DNA from Atomistic Simulations

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · AUGUST 2009

Impact Factor: 12.11 · DOI: 10.1021/ja901761a · Source: PubMed

2 AUTHORS, INCLUDING:

Benjamin Bouvier
Institute for the Biology and Chemistry of P…

17 PUBLICATIONS **379** CITATIONS

SEE PROFILE

J|A|C|S COMMUNICATIONS

Published on Web 07/06/2009

A Free Energy Pathway for the Interaction of the SRY Protein with Its Binding Site on DNA from Atomistic Simulations

Benjamin Bouvier* and Richard Lavery

Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS. 7, passage du Vercors, F-69007 Lyon, France

Received March 6, 2009; E-mail: benjamin.bouvier@ibcp.fr

In mammals, the SRY gene on the Y chromosome is a necessary and sufficient condition for the development of the male phenotype.[1] The SRY protein belongs to the SOX (SRY-type HMG box) family of transcriptional regulator proteins. These proteins are involved in numerous biological functions across many species and feature compact DNA binding domains (DBD); they are therefore appealing models for the study of protein−DNA recognition. Moreover, by binding in the minor groove of DNA rather than in the more usual major groove sites, they produce severe DNA deformation and are thus likely to find their target sites partly by so-called indirect recognition,[2] which probes the sequence-dependent mechanics of the DNA double helix. Indeed, SRY and other SOX members are known to bind in vivo to sequences that only partially match their in vitro consensus.[3] Structural transitions during protein−DNA binding are also believed to be at the origin of the passage between a transient nonspecific complex, which facilitates quasi-1D searching for the binding site by protein sliding along DNA,[4] and a tightly bound, specific complex, once the correct target sequence in DNA has been located.[5] However, the nature of these transitions, even whether they principally involve the protein or the DNA, remains largely unknown. The conjunction of combinatorial sequence variations and slow conformational transitions explains that such challenging questions have only recently entered the scope of simulation methods,[6] although some earlier attempts have paved the way.[7]

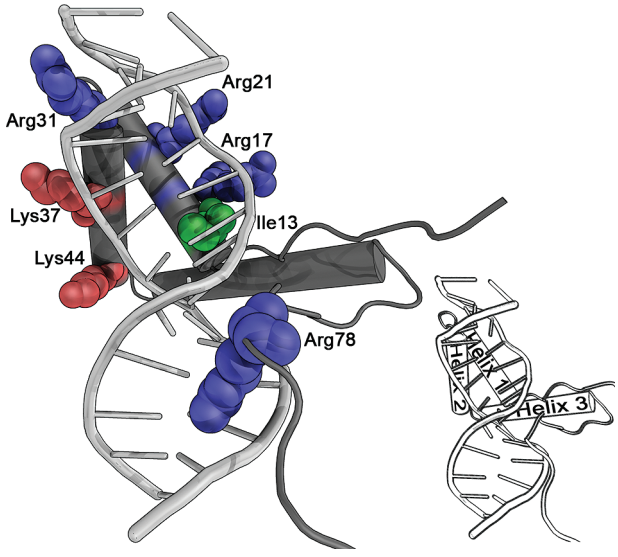

Figure 1. Cartoon representation of the specific SRY/DNA complex, showing color-coded key amino acids (Lys in red, Arg in blue, and Ile in green) and the standard nomenclature of α helices (inset).

As a step toward a better understanding of protein−DNA binding, we have simulated the controlled dissociation of the SRY protein DBD from its cognate DNA sequence. The specific complex, whose conformation has been determined by NMR spectroscopy,[8] involves a 14 base pair DNA oligomer (5′-dCCT**GCACAAAC**ACC-3′) containing the binding site (bold letters). Protein binding leads to minor groove opening, DNA bending by 50° away from the protein and the partial intercalation of an isoleucine side chain (Ile13) at the A8pA9 step (Figure 1). From this initial structure, the pathway of complex dissociation was probed using ∼0.6 μs of all-atom molecular dynamics simulations, in conjunction with a harmonic biasing potential designed to impose a given separation distance between partners with a limited bias on the pathway. The associated free energy profile was also computed (Figure 2).

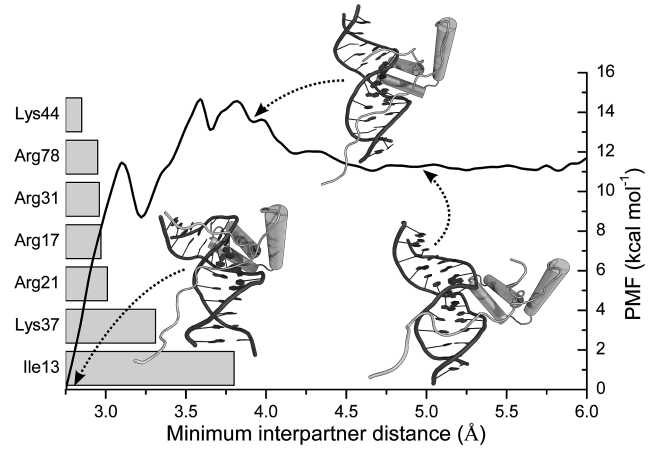

Figure 2. Potential of mean force (PMF, thick line) and persistence of major protein−DNA interactions (gray bars), as a function of the minimum interpartner distance d. Representative structures along the separation pathway are shown. Salt bridges and hydrophobic contacts are considered broken for heavy atom distances greater than 3.5 and 4.5 Å, respectively; residual interactions might however persist.

The overall dissociation mechanism consists of three phases, separated by free energy maxima of 12 kcal mol^{-1} at d = 3.1 Å and of 15 kcal mol^{-1} at d = 3.5−3.8 Å. During the first phase, the protein changes position and releases the DNA toward the ends of the binding site, with little structural change other than in its flexible C- and N-termini, and in side chain rearrangements following loosening of the first salt bridges between cationic arginine and lysine side chains and the anionic DNA phosphate groups. These specifically involve Lys44 (at d = 2.9 Å) which releases the C-terminal end of helix 2, allowing it to move away from DNA, followed by Arg31 at the N-terminus of the same helix which breaks its contact with the 3′-end of DNA strand 1 and then transiently contacts a thymidine on the opposing DNA strand (d = 2.95 Å). Simultaneously, the weakening of a hydrogen bond between Arg78 and an adenine in DNA strand 2 frees the other end of the binding site. At d = 3 Å, salt bridges with Arg17 and Arg21, both on helix

10.1021/ja901761a CCC: $40.75

1 which also carries the intercalating Ile13 residue, are lost, without significantly disturbing Ile13. Loosening of conformational restraints due to strong protein−DNA interactions is reflected in a 3 kcal mol^{-1} stabilization of the dissociating complex at the end of this phase (d = 3.2 Å).

The second phase of the dissociation (3.2 < d < 3.7 Å) is characterized by an extensive reorganization of the DNA, which straightens up and relaxes the base pair deformation around the partial intercalation site as the Ile13 side chain begins to withdraw. At d = 3.3 Å, the salt bridge involving Lys37 on helix 2 breaks, finally freeing this helix and, in turn, weakening the restraints on helix 1. This is accompanied by increasing reorganization of the C- and N-termini of the protein DBD which move away from DNA, breaking salt bridges involving Arg4 and Arg77 at base pair A6 (d = 3.6 Å). Further loss of strong electrostatic interactions and complementary protein−DNA surface interactions during this phase is likely to be compensated by the conformational relaxation of DNA and the stabilizing effect of water, which progressively penetrates between the two partners.

The final phase (d > 3.7 Å) initiates with the extraction of Ile13 from the DNA minor groove (d = 3.8 Å), enabling the partners to rebuild their respective hydration layers. This phase corresponds to an ~3 kcal mol^{-1} gain in free energy as the DNA relaxes, losing the induced axial bend by d = 4 Å and finally adopting a canonical B-like conformation at d = 6 Å. The final dissociation free energy is 11.5 kcal mol^{-1}, for an experimental value of 9.7 kcal mol^{-1}.[9]

Since DNA deformation is expected to play a large role in the specificity of SRY binding, we now look at its conformational rearrangements in more detail. Interestingly, during the first stage of protein release, it actually becomes more bent (65°) than in the initial complex (50°). This change accompanies the loss of the salt bridge with Arg31 at the 3′ end of the binding site, apparently freeing a restraint on its overall shape. Other broken interactions are reflected in more local changes, such as the twist at the T3pG4 step contacted by Arg78, which is strongly unwound in the initial complex (twist 16°) but quickly recovers a normal conformation when this interaction is broken (d = 3.0−3.2 Å). In contrast, the equally underwound partial intercalation site (A8pA9, twist 14°) recovers in stages, reducing its roll from an initial 50° to ~0° at d = 3.5 Å and its rise from 6.1 Å to a canonical 3.4 Å by d = 3.9 Å. The twist at this step is the last feature to adjust, finally adopting a normal value at d = 5 Å. Finally, by d = 6.8 Å the sequence-averaged twist over the central 12 base pairs of the oligomer adopts a canonical value of 33°, at which point the minor groove at the binding site has also contracted to its normal width and allowed the opposing major groove to reopen.

Water, the third partner in the SRY/DNA complex, appears to play a crucial stabilizing role during the dissociation. A few water molecules are present in the specific complex around key residues (Ile13, Arg31). More water enters the growing cleft separating the partners from two locations: starting at d = 3.2 Å, between the DNA backbone and the termini of the protein; at d = 3.6 Å from the other end of the binding site, near Arg31. By d = 3.8 Å, direct protein−DNA contacts are reduced to a patch around Ile13. By d = 4.25 Å, increasing separation destroys the water bridges, allowing independent hydration of the partners to begin.

Inasmuch as we have induced SRY dissociation in small steps (0.05 Å), with extensive sampling to equilibrate each step (typically >6 ns), and using a restraint which minimally biases the dissociation pathway, we can hope that many features of this pathway will also apply to the association of the complex. In this case, we can postulate association to begin with significant DNA deformation when the protein is still far away (d = 5 Å). This deformation, involving helical untwisting and minor groove opening, continues until d = 4 Å and requires only 1−2 kcal mol^{-1}. Considering the free energy range of 4 to 10 kcal mol^{-1} reported by Zacharias[10] for the opening of the DNA minor groove, this hints at an active stabilizing role of the protein. Beyond this point, a further 1−2 kcal mol^{-1} allows for extensive DNA deformation and the start of Ile13 intercalation. Here again, the protein can play an active role, facilitating DNA bending by neutralizing its backbone charges.[11] Completing this process and expelling water to form an extensive protein−DNA interface lead to a metastable intermediate, 2.5 kcal mol^{-1} below the free energy of the separated partners. A barrier of only 3 kcal mol^{-1} then separates the intermediate from the specific complex, characterized by further DNA deformation and stabilization via salt bridges which fully lock helices 1 and 2 into place. This secondary barrier may be linked to unfavorable electrostatic contacts that manifest themselves at such small separation distances,[12] combined with a global loss in entropy despite the release of structural waters.[13] It also confirms the view that DNA binding and bending occur in two separate steps rather than simultaneously, an important feature of indirect recognition reported for the IHF protein.[14]

This view of SRY−DNA association supports the idea that DNA deformation may play a role in determining the specificity, since deformation in the early stages of protein approach is an ideal way of indirectly probing the local base sequence. Whether the nonspecific complex involved in quasi-1D searching along DNA corresponds to the metastable intermediate we have identified at d = 3.2 Å or a more loosely bound state at a larger separation remains to be determined and is the subject of ongoing simulations involving noncognate DNA sequences.

Acknowledgment. The authors acknowledge funding from Agence Nationale pour la Recherche Grant ALADDIN. B.B. is supported by ANR Grant HIPCAL.

Supporting Information Available: Details of methodology and computations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Berta, P.; Hawkins, J. R.; Sinclair, A. H.; Taylor, A.; Griffiths, B. L.; Goodfellow, P. N.; Fellous, M. *Nature* **1990**, *348*, 448–450.
(2) Bewley, C. A.; Gronenborn, A. M.; Clore, G. M. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105–131.
(3) Peters, R.; King, C. Y.; Ukiyama, E.; Falsafi, S.; Donahoe, P. K.; Weiss, M. A. *Biochemistry* **1995**, *34*, 4569–4576.
(4) von Hippel, P. H. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 79–105.
(5) Slutsky, M.; Mirny, L. A. *Biophys. J.* **2004**, *87*, 4021–35. Halford, S. E.; Marko, J. F. *Nucleic Acids Res.* **2004**, *32*, 3040–3052.
(6) Ahmad, S.; Kono, H.; Arauzo-Bravo, M. J.; Sarai, A. *Nucleic Acids Res.* **2006**, *34*, W124–W127. Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J.; Stoddard, B. L.; Baker, D. *Nature* **2006**, *441*, 658–659. Paillard, G.; Lavery, R. *Structure* **2004**, *12*, 113–122.
(7) Tang, Y.; Nilsson, L. *Proteins* **1998**, *31*, 417–433.
(8) Murphy, E. C.; Zhurkin, V. B.; Louis, J. M.; Cornilescu, G.; Clore, G. M. *J. Mol. Biol.* **2001**, *312*, 481–499.
(9) Sanchez-Moreno, I.; Coral-Vazquez, R.; Mendez, P.; Canto, P. *Mol. Hum. Reprod.* **2008**, *14*, 325–330.
(10) Zacharias, M. *Biophys. J.* **2006**, *91*, 882–891.
(11) Kosikov, K. M.; Gorin, A. A.; Lu, X. J.; Olson, W. K.; Manning, G. S. *J. Am. Chem. Soc.* **2002**, *124*, 4838–4847.
(12) Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Das, A.; Beveridge, D. L. *J. Comput. Chem.* **2002**, *23*, 1–14.
(13) Spitzer, G. M.; Fuchs, J. E.; Markt, P.; Kirchmair, J.; Wellenzohn, B.; Langer, T.; Liedl, K. R. *ChemPhysChem* **2008**, *9*, 2766–2771.
(14) Kuznetsov, S. V.; Sugimura, S.; Vivas, P.; Crothers, D. M.; Ansari, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18515–18520.

JA901761A