



**HAL**  
open science

## Protein-DNA Recognition Triggered by a DNA Conformational Switch

Benjamin Bouvier, Krystyna Zakrzewska, Richard Lavery

► **To cite this version:**

Benjamin Bouvier, Krystyna Zakrzewska, Richard Lavery. Protein-DNA Recognition Triggered by a DNA Conformational Switch. *Angewandte Chemie International Edition*, 2011, 50 (29), pp.6516-6518. <10.1002/anie.201101417>. <hal-02395321>

**HAL Id: hal-02395321**

**<https://hal.science/hal-02395321v1>**

Submitted on 9 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Protein-DNA recognition triggered by a DNA conformational switch\*\*

Benjamin Bouvier, Krystyna Zakrzewska and Richard Lavery\*

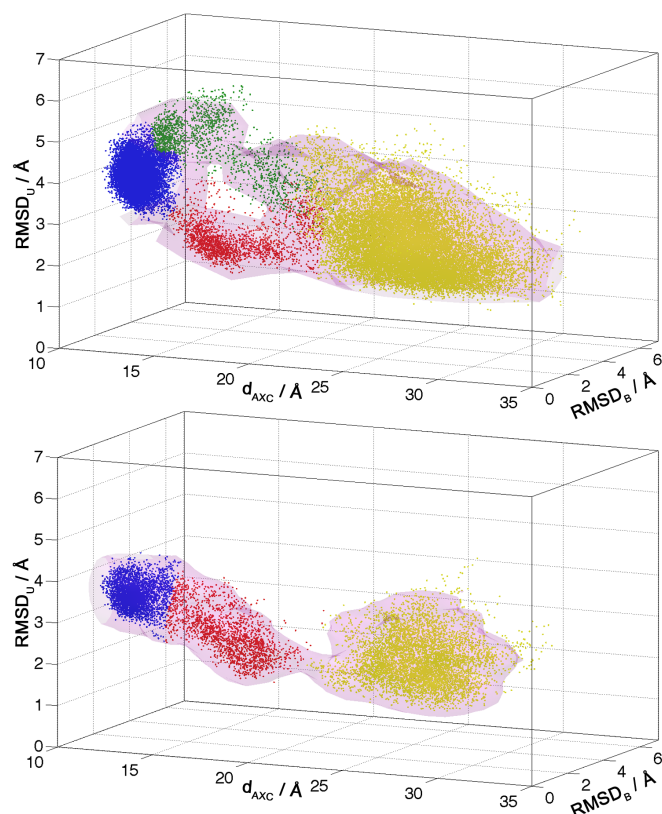
How DNA-binding proteins find their target sites remains a fascinating question. Early work on the Lac repressor showed that proteins specifically bind faster than simple diffusion allows. This led to the idea that recognition could be accelerated by combining diffusion with sliding along DNA and hopping between neighboring strands.<sup>[1,2]</sup> This, in turn, implied that proteins could interact with DNA in a distinct non-specific manner. Experimental work has confirmed that proteins can indeed slide along DNA, although typical sliding distances vary from protein to protein.<sup>[3,4]</sup> Recent work also indicates that sliding follows the helical grooves of DNA.<sup>[5,6]</sup>

Crystallography, using proteins bound to non-cognate sites, NMR spectroscopy and molecular simulation have all provided data on non-specific binding, notably suggesting that proteins maintain similar orientations with respect to DNA in non-specifically and specifically bound states (see, for example.<sup>[7-9]</sup> However, little is known about the transition between these states, although theoretical studies have suggested that a switching mechanism should be involved, possibly involving a protein conformational change.<sup>[10]</sup>

In order to analyze this problem at the atomic level, we carried out a molecular dynamics (MD) study of dissociating a specific protein-DNA complex in water, starting from the experimental bound conformation. We studied the protein SRY, which is involved in gender selection in mammals and linked to a number of gender-related pathologies.<sup>[11]</sup> SRY binds in the minor groove of DNA, optimally at an (A/T)AACAAT sequence,<sup>[12]</sup> opening the minor groove, partially intercalating an isoleucine residue (Ile13) between two adjacent AT base pairs and inducing local unwinding and DNA bending away from the protein.<sup>[13]</sup> Using a specially designed restraint controlling the minimal atomic distance between any pair of non-hydrogen atoms across the protein-DNA interface ( $d_{\text{MIN}}$ ), we were able to control the dissociation of SRY from a 14 base pair (bp) DNA oligomer (5'-CCTGCACAAACACC-3'), without biasing the conformational pathway. Using this approach, roughly 0.6  $\mu\text{s}$  of umbrella sampling led to a free energy profile for the dissociation/association processes.<sup>[14]</sup> This profile showed an 11.5 kcal mol<sup>-1</sup> free energy gain due to SRY binding, which involved passing an energy barrier of roughly 4 kcal mol<sup>-1</sup> at a separation of 4.2-3.5 Å and a secondary 2 kcal mol<sup>-1</sup> barrier at 3.1 Å (Figure S1 Supporting Information). We now investigate the conformational aspects of this pathway to understand the recognition mechanism.

An initial analysis showed that the conformation of SRY

remained remarkably stable during its separation from DNA. Although the N- and C-terminal tails were very flexible, the three  $\alpha$ -helix protein core varied by 2.2 Å RMSD at most (Figure S1 Supporting Information). In contrast, DNA underwent considerable change, linked to the SRY-induced deformations described above. However, our initial analysis of the separation profile also showed that many DNA conformations and protein locations occurred for a single minimal pair distance along the separation pathway. We reduced this problem by using the distance,  $d_{\text{AXC}}$ , from the center of the DNA helical axis to the center of mass of the core region of SRY. This distance varies almost monotonically with the minimal pair distance for  $d_{\text{MIN}} < 6$  Å, but increases more rapidly in the region of the main free energy barrier (Figure S2 Supporting Information). To further characterize DNA conformations we introduced two reference states: the average DNA conformation bound to SRY and the average free B-DNA conformation, following SRY dissociation (see Table S1 Supporting Information).



**Figure 1.** Representation of conformational space for the wild-type (top) and mutated (bottom) SRY/DNA complex. Each dot represents a conformation, color-coded by group (bound: blue, unbound: yellow), path 1: green, path 2: red. The purple envelope is an isodensity surface.

Using the RMSD values with respect to the bound ( $\text{RMSD}_{\text{B}}$ ) and unbound ( $\text{RMSD}_{\text{U}}$ ) DNA reference states and the separation distance  $d_{\text{AXC}}$ , we get a clearer view of what happens to DNA along the free energy profile. Figure 1 plots conformations drawn from the

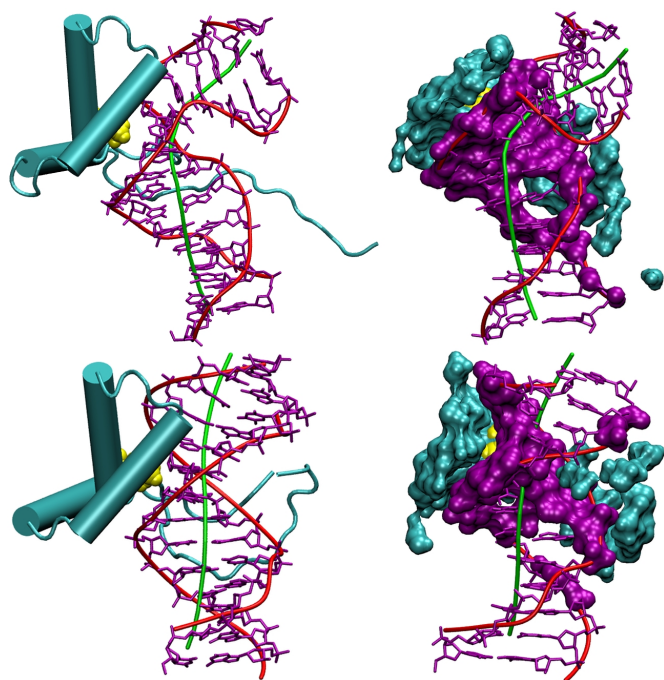
[\*] B. Bouvier, K. Zakrzewska, R. Lavery  
 Université Lyon 1 / CNRS, UMR 5086, Bases Moléculaires  
 et Structurales des Systèmes Infectieux,  
 IBCP, 7 passage du Vercors - F-69367 Lyon, France.  
 E-mail: [richard.lavery@ibcp.fr](mailto:richard.lavery@ibcp.fr)  
 Homepage: <http://www.ibcp.fr>

[\*\*] This work was supported by the ANR Blanc ALADDIN  
 project and by the CNRS. Computational resources were  
 granted by the Centre Informatique National de  
 l'Enseignement Supérieur (CINES).



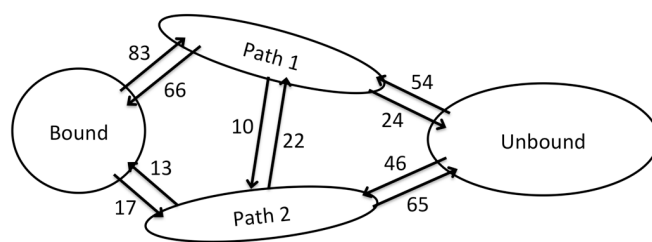
Supporting information for this article is available on the  
 WWW under <http://www.angewandte.org> or from the author.

umbrella sampling as a function of these three variables, clustered (and color-coded) using a neural gas approach (see Supplementary Information). The bound state forms a tight cluster of blue points at  $d_{\text{AXC}} < 13 \text{ \AA}$ , where the bound protein stabilizes the deformed DNA conformation. This state is separated by roughly 4 Å RMSD from the average free conformation, characterized by the loose cluster of yellow points at  $d_{\text{AXC}} > 20 \text{ \AA}$ . The most interesting feature lies between these two regions ( $13 \text{ \AA} > d_{\text{AXC}} > 20 \text{ \AA}$ ), where the DNA conformations clearly split into two paths (green and red). The center of the two-path region occurs around  $d_{\text{AXC}} = 16 \text{ \AA}$ . Here, both paths are  $\approx 3 \text{ \AA}$  RMSD from bound DNA, but while the upper path (path 1) is 5.2 Å RMSD from the free DNA reference, the lower path (path 2) is much less perturbed with an RMSD of only 2.2 Å.



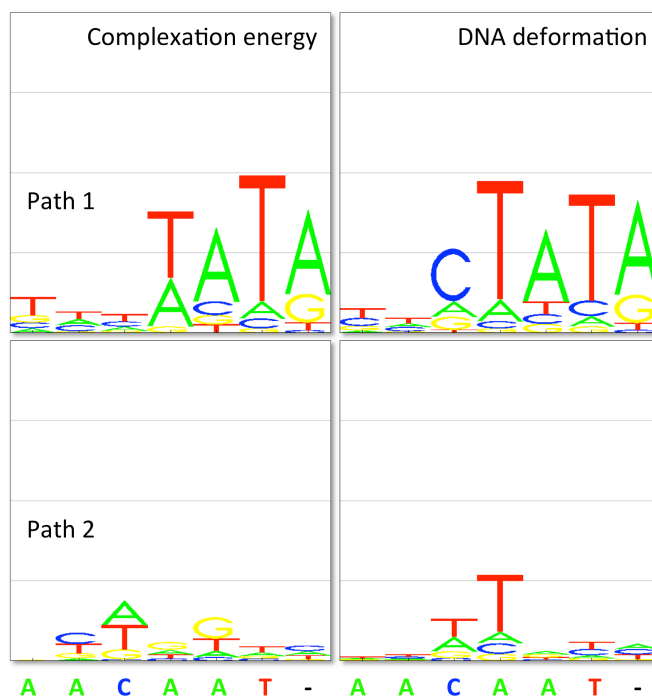
**Figure 2.** Representative conformations of the system (DNA: purple, protein: cyan, Ile13: yellow) along path 1 (top) and path 2 (bottom). Left: cartoon representation highlighting the DNA axis (green tube) and backbone (red tubes); right: surface representation of the SRY/DNA interface.

Figure 2 shows conformations from the center of the two paths. SRY has a similar orientation with respect to DNA in both cases. As SRY approaches DNA it rotates, presumably guided by electrostatic interactions, until its second  $\alpha$ -helix is aligned with the DNA minor groove and its orientation stabilizes (Figure S4 Supporting Information). DNA however reacts differently to the protein along the two pathways (Table S1 Supporting Information). Path 1 shows a sharp kink at the A8pA9 step, where SRY Ile13 will finally intercalate, and also at the following AA step (bp rolls of  $20^\circ$  and  $53^\circ$  respectively) leading to an overall bend of  $59^\circ$ . It also has a strongly reduced twist at A8pA9 ( $12^\circ$ ) and a minor groove much wider than the major groove ( $11.9 \text{ \AA}$  versus  $7.8 \text{ \AA}$ ). In contrast, path 2 shows smaller kinks at A8pA9 and A9pA10 (bp rolls of  $5^\circ$  and  $7^\circ$  respectively) and an overall bend of only  $19^\circ$ . The A8pA9 twist is  $32^\circ$  and the minor groove remains narrower than the major groove ( $10.7 \text{ \AA}$  versus  $12.3 \text{ \AA}$ ). Thus, many features of path 1 are similar to the bound state of DNA, while path 2 is only weakly perturbed from the free state.



**Figure 3.** Percentage of transitions observed from any region of phase space during MD simulations.

These conformational differences cause the SRY-DNA interface to evolve differently as the protein continues to approach. Although the interface looks better packed in path 1 (Figure 2), the total protein-DNA contact area is similar in both paths ( $\approx 100 \text{ \AA}^2$ , Figure 3 Supporting Information). However, at  $d_{\text{AXC}} = 14.5 \text{ \AA}$ , water is abruptly excluded from the path 1 interface, increasing the contact area to  $500 \text{ \AA}^2$ . A similar change occurs along path 2, but only when SRY is much closer ( $d_{\text{AXC}} = 13 \text{ \AA}$ ). Note that  $d_{\text{AXC}} = 14.5 \text{ \AA}$  corresponds to top of the main free energy barrier opposing the formation of the complex ( $d_{\text{MIN}} = 3.75 \text{ \AA}$ ).<sup>[14]</sup> The energetic penalty of dehydrating the binding interface thus occurs more easily along path 1.



**Figure 4.** DNA sequence specificity of the SRY protein represented as logos. Top: path1, bottom: path2; left: using total complexation energy, right: contribution of DNA deformation. Lower line: experimental consensus sequence.

We can now ask what is the most likely route to site-specific binding. Counting the transitions occurring between the regions along the free energy pathway (Figure 3) shows that although paths 1 and 2 are similarly populated (implying a free energy difference of  $< 0.1 \text{ kcal mol}^{-1}$ ), path 1 has a much higher chance of reaching the bound conformation. Path 2 is most likely to return to the unbound state, although it can also transit to path 1.

The emerging picture is that electrostatic interactions attract the protein and orient it with respect to DNA (we cannot easily umbrella sample long-range interactions, but a Poisson-Boltzmann calculation shows a favorable protein-DNA interaction of  $\approx 5$  kcal mol<sup>-1</sup> at  $d_{\text{AXC}}=30$  Å). As the protein reaches the energy barrier linked with interface dehydration, it triggers DNA to make a conformational change, preparing the passage to the bound conformation. Analyzing the sequence-dependence of the complexation energy with the ADAPT approach<sup>[15,16]</sup> at  $d_{\text{AXC}}=16$  Å along path 1 shows that this conformational change allows the protein to partially recognize its target sequence (mostly via DNA deformation). No such recognition occurs along path 2. Following DNA deformation, the protein can continue along path 1 to the bound state.

What happens when SRY approaches the wrong DNA sequence? We tested this possibility by mutating A8pA9 to G8pG9 in the bound conformation and then recalculated and analyzed the corresponding dissociation pathway. Figure S4 (Supporting Information) shows a similar free energy profile with the mutant sequence. Because WHAM free energies include an undefined constant, and because we cannot sufficiently sample conformational space for the dissociated complexes, we have used non-linear Poisson-Boltzmann electrostatic energy calculations at  $d_{\text{MIN}} = 3.5$  Å to position the curves with respect to one another. This implies that the mutant binds SRY roughly 3 kcal mol<sup>-1</sup> less well than the wild-type (in line with experimental data showing that A → G substitutions at positions 8 or 9 lead to weaker binding).<sup>[17]</sup> Figures S5 and S6 (Supporting Information) show that SRY occupies a similar position with respect to DNA at short separations, but that Ile13 intercalation causes more perturbation of DNA structure in the case of the mutant. The consequence for DNA conformation (see Figure 1) is that the bound conformation is lost as soon SRY moves away from the mutant sequence. Path 1 no longer exists, and thus, with an incorrect target sequence, the protein is unable to trigger the DNA conformational switch and no recognition occurs.

This study supports the idea of an energy barrier (at  $d_{\text{MIN}} \approx 4$  Å) linked here to interface dehydration,<sup>[18]</sup> separating the non-specifically and specifically bound states. A sequence-specific DNA conformational switch (rather than a protein switch)<sup>[10]</sup> controls the passage through this barrier.

Received: ((will be filled in by the editorial staff))  
Published online on ((will be filled in by the editorial staff))

**Keywords:** protein-DNA recognition · molecular dynamics · conformational transitions · base sequence specificity · DNA plasticity

- 
- [1] R.B. Winter, O.G. Berg, P.H. von Hippel *Biochemistry* **1981**, *20*, 6961.
  - [2] P.H. von Hippel, O.G. Berg *J Biol Chem* **1989**, *264*, 675.
  - [3] I. Bonnet, A. Biebricher, P.L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, P. Desbiolles *Nucleic Acids Res* **2008**, *36*, 4118.
  - [4] J. Gorman, E.C. Greene *Nat Struct Mol Biol* **2008**, *15*, 768.
  - [5] B. Bagchi, P.C. Blainey, X.S. Xie *J Phys Chem B* **2008**, *112*, 6282.
  - [6] P.C. Blainey, G. Luo, S.C. Kou, W.F. Mangel, G.L. Verdine, B. Bagchi, X.S. Xie *Nat Struct Mol Biol* **2009**, *16*, 1224.
  - [7] H. Viadiu, A.K. Aggarwal *Mol Cell* **2000**, *5*, 889.
  - [8] J. Iwahara, G.M. Clore *Nature* **2006**, *440*, 1227.
  - [9] O. Givaty, Y. Levy *J Mol Biol* **2009**, *385*, 1087.
  - [10] M. Slutsky, L.A. Mirny *Biophys J* **2004**, *87*, 4021.
  - [11] C.Y. King, M.A. Weiss *Proc Natl Acad Sci U S A* **1993**, *90*, 11990.
  - [12] V.R. Harley, R. Lovell-Badge, P.N. Goodfellow *Nucleic acids research* **1994**, *22*, 1500.
  - [13] E.C. Murphy, V.B. Zhurkin, J.M. Louis, G. Cornilescu, G.M. Clore *J Mol Biol* **2001**, *312*, 481.
  - [14] B. Bouvier, R. Lavery *J Am Chem Soc* **2009**, *131*, 9864.
  - [15] C. Deremble, R. Lavery, K. Zakrzewska *Comput Phys Commun* **2008**, *179*, 112.
  - [16] K. Zakrzewska, B. Bouvier, A. Michon, C. Blanchet, R. Lavery *Phys Chem Chem Phys* **2009**, *11*, 10712.
  - [17] V.R. Harley, D.I. Jackson, P.J. Hextall, J.R. Hawkins, G.D. Berkovitz, S. Sockanathan, R. Lovell-Badge, P.N. Goodfellow *Science* **1992**, *255*, 453.
  - [18] V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, J.M. Victor *Phys Rev Lett* **2009**, *102*, 228101.
-

**Entry for the Table of Contents** (Please choose one layout)

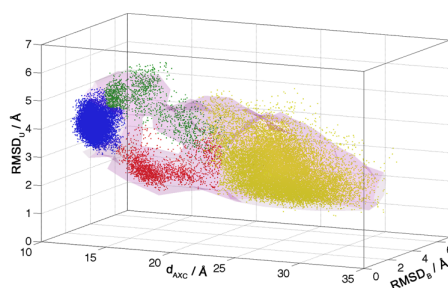
Layout 1:

**Protein-DNA recognition**

Benjamin Bouvier, Krystyna Zakrzewska, Richard Lavery\*

Page – Page

Protein-DNA recognition triggered by a DNA conformational switch



An analysis of molecular dynamics simulations of the SRY/DNA complex shows that as the protein approaches the correct DNA target sequence, it triggers a DNA-conformational switch and allows the passage from a non-specifically bound to a specifically bound state. No switch occurs when SRY binds to a non-consensus DNA sequence.