J|A|C|S COMMUNICATIONS

Published on Web 10/05/2009

Disordered Tails of Homeodomains Facilitate DNA Recognition by Providing a Trade-Off between Folding and Specific Binding

Ágnes Tóth-Petróczy,[†,‡] Istvan Simon,[‡] Monika Fuxreiter,[‡] and Yaakov Levy*[,†]

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel, and Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1518 Budapest, Hungary

Received June 29, 2009; E-mail: koby.levy@weizmann.ac.il

Specific DNA binding is a critical element of all processes related to the maintenance of genetic information. Proteins playing role in transcription and DNA repair have been developed to efficiently discriminate between their target sequences and the vast pool of nonspecific ones. Substrate selectivity can also be influenced by protein regions that lack a well-defined structure in solution and fold only upon binding to their specific targets.[1] For example, exchanging the disordered N-terminal segments (N-tails) in Antennapedia (Antp) and thyroid transcription factor 1 (TTF-1) homeodomains abolished binding to both cognate sites.[3] N-tails were also sufficient to discriminate between the specificities of Antp and Sex combs reduced (Scr) homeodomains in vivo.[5] Despite their importance in determining the specificity of homeodomains,[6] the mechanism by which disordered N-tails modulate target selection is largely unknown.

N-tails of homeodomains contribute to complex formation by increasing the number of contacts with DNA. Primarily highly charged N-tails form electrostatic interactions with DNA phosphates, but specific contacts with the cognate bases have also been observed.[2] The long-range electrostatic forces between the N-tails and DNA might also interfere with their folding process, as seen in the case of Ets transcription factor.[7] Disordered N-tails can also modulate the affinity of homeodomains for DNA via the interplay between the cost of folding and the entropy gain due to the released waters.[8] As an alternative scenario, N-tails can influence the selectivity of homeodomains by interacting with other proteins.[5]

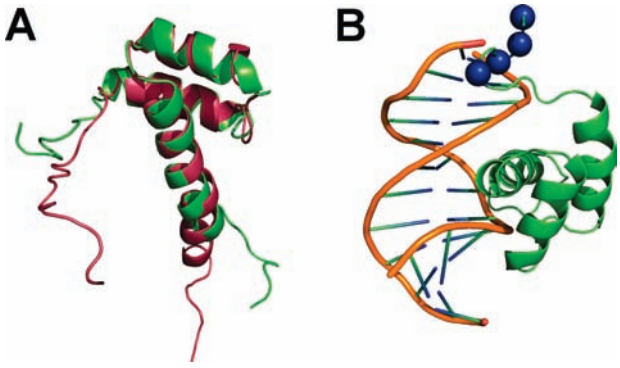

Figure 1. NMR structures of the *Drosophila* Antennapedia and NK-2 homeodomains complexed with DNA. (A) Superposition of the Antp (green, PDB code 1ahd[2]) and NK-2 (purple, PDB code 1nk2[4]) homeodomains ($rmsd_{backbone}$ = 1.28 Å). The N-tails of NK2 and Antp are composed of 7 and 14 residues, respectively. (B) Interactions between the Antp homeodomain and DNA. The recognition helix is in the major groove, and the disordered N-tail is inserted in the minor groove. Positively charged residues in the N-tail are colored blue.

To elucidate how N-tails facilitate target selection of homeodomains, native-topology-based model (Gō model) simulations[9] were performed on the full and N-tail-truncated variants of the Antp[2] and NK-2[4] proteins (Figure 1). These coarse-grained techniques were successfully applied to explain binding-coupled folding of disordered proteins in good agreement with experiment.[10]

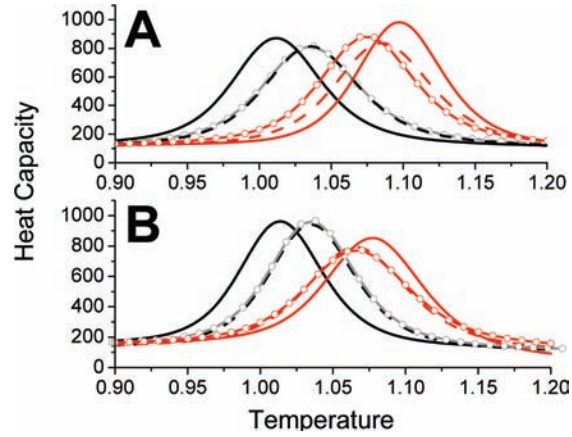

Figure 2. Folding heat capacity curves of the (A) Antp and (B) NK-2 homeodomains. Folding was studied in the free form (black) and in the presence of DNA (red) at a salt concentration of 0.01 M using full-length (solid) and truncated (dashed) forms of the homeodomains as well as full-length proteins with uncharged tails (open circles).

The curves of the specific heat capacity of both full Antp and NK-2 are shifted to lower temperatures relative to truncated (no N-tail) Antp and NK-2 (Figure 2), indicating that the disordered N-tails decrease the stability of the homeodomains to a similar extent. The decrease in the transition temperatures for folding (where the folded and unfolded states are equally populated) by the N-tail is in accord with experiment.[11] The decreased protein stability of Antp and NK-2 is reflected by the larger average size (measured by the radius of gyration) and a smaller number of native contacts at various temperatures (Figure S1 in the Supporting Information). NK2 and Antp with uncharged N-tails show behavior very similar to that of truncated Antp and NK2, indicating that the disordered tails affect the unfolded state by non-native charge−charge interactions. In contrast, in the presence of DNA, the N-tails increase the stability of the homeodomain cores, resulting in a smaller radius of gyration (Figure S1), a larger number of native contacts, and a 4−6% higher transition temperature (Figure 2) in both homeodomains. This suggests that the N-tails do not compete with the native protein−protein contacts in charged form, likely because of their preference for DNA. In the absence of the tail charges, however, there is still competition with the native interactions.

N-tails also increase stability of homeodomains by reducing the coupling between folding of the protein core and DNA binding (Figure S2). As a consequence, the full-length Antp folds before it

[†] Weizmann Institute of Science.
[‡] Hungarian Academy of Sciences.

10.1021/ja9052784 CCC: $40.75

tightly associates with DNA, and the specific interactions are established with the folded protein core. In the absence of the N-tail, however, specific DNA contacts compete with native protein contacts, thereby increasing the frustration of the protein-core folding as a result of the tight coupling between the two processes. Hence, the full-length Antp may be viewed as a two-domain protein: the N-tail anchors the homeodomain to DNA, while the core of the protein folds and subsequently forms specific interactions at the protein−DNA interface. In the absence of the N-tail, there are fewer constraints on the protein, and it folds and binds concomitantly.

The long-range electrostatic interactions between the DNA and the charged N-tail can also serve to modulate the folding of the homeodomain core. This may accelerate DNA binding by increasing the capture radius of the protein, as proposed by the fly-casting model.[7,12,13] In the case of Antp, the presence of the N-tail decreases the time required to form all of the specific contacts between the Antp core and the DNA by ∼25% (Figure 3). It is important to note that no native contacts between the N-tail and the DNA were defined and that only their electrostatic interactions were taken into account. In this way, the simulations were not biased toward specific contacts between the N-tail and the DNA. This observation demonstrates that anchoring the homeodomain to the DNA even by nonspecific electrostatic interactions accelerates the formation of specific interactions all along the interface. In addition to nonspecific electrostatic interactions, various specific contacts between the N-tail and the DNA were found (A11, G12, A13, A19, A20, T21; see Figure S3), in agreement with experiment.[2]

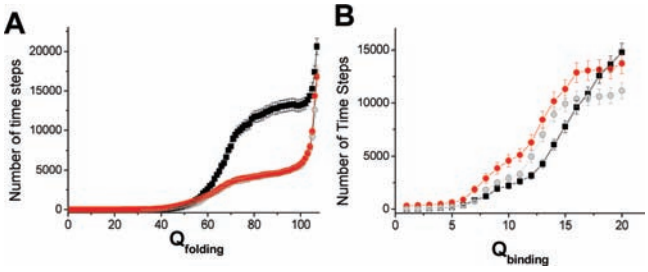

Figure 3. Kinetics of folding and binding of Antp to DNA with and without the N-tail. (A) Time (in simulation steps) required to form a fraction of native contacts in the presence of DNA. (B) Time required to form the interfacial native contacts between Antp and DNA. Full-length Antp (solid, black) is compared to truncated protein (dashed, red) and to uncharged N-tail Antp (gray).

The N-tail of Antp improves the binding affinity (Table 1) as compared with the truncated homeodomain, in accord with experimental data.[11] Because it underestimates the hydrophobic effect, however, the simulation finds the binding to be enthalpy-driven, while the entropy term is unfavorable (as water was only implicitly taken into account). Thermodynamic data derived directly from the simulations were thus corrected with the estimated magnitude of the hydrophobic effect (Table 1), approximated by the number of waters released from the first hydration shell upon formation of protein−DNA interactions[14] and the entropy gain estimated on a per-molecule basis[15] Upon incorporation of the hydrophobic effect, the N-tail provides a significant contribution to the binding entropy, as reflected by the almost 30% increase in $T\Delta S_{\mathrm{bind}}$ relative to the value for truncated Antp. This demonstrates that burying a large surface upon binding of a disordered protein segment to DNA can be a considerable component in the binding free energy that ultimately balances the enthalpic terms provided by specific contacts. The more favorable entropy for the full protein even without the hydrophobic correction indicates that the tail retains substantial flexibility in the bound state.

Table 1. Thermodynamic Data (kcal mol^{-1}) Computed for Antp Binding to Cognate DNA without and with the Hydrophobic Effect (HE) in Comparison with Experimental Results[11]

	without HE			with HE			experiment		
	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$	ΔG	ΔH	$-T\Delta S$
Antp	−1.8	−4.4	2.6	−26.0	−4.4	−21.8	−13.0	−2.9	−10.1
truncated Antp	−1.7	−4.3	2.6	−23.0	−4.3	−18.7	−11.4	−1.7	−9.7

In summary, disordered N-terminal tails influence the substrate selectivity of homeodomains both kinetically and thermodynamically. N-tails primarily serve to anchor the protein to DNA by tidal electrostatic forces, facilitating productive contacts all along the interface and thus increasing the speed of complex formation via the fly-casting mechanism. N-tails are also engaged in specific interactions with the DNA. On the other hand, the balance between the entropy cost of the binding-induced folding and the entropy gain due to the hydrophobic effect may serve to fine-tune the specificity of homeodomains. This might affect substrate selectivity via a larger sequence context and may also account for environment-dependent variations in selectivity. The architecture of DNA-binding proteins with long disordered tails can be advantageous not only for specific recognition but also for searching the DNA using the nonspecific binding mode. The existence of the tail may improve sliding dynamics along the DNA[16] as well as protein hopping between different segments of DNA.[17]

Acknowledgment. This work was funded by grants from the Hungarian Scientific Research Fund (OTKA) (K72569, MRTN-CT-2005-019566 to M.F.) and supported in part by the Kimmelman Center for Macromolecular Assemblies, the Center for Complexity Science, and the Israel Science Foundation (Y.L.). Y.L. is the incumbent of the Lilian and George Lyttle Career Development Chair.

Supporting Information Available: Computational methods and supporting data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777.
(2) Billeter, M.; Qian, Y. Q.; Otting, G.; Muller, M.; Gehring, W.; Wuthrich, K. *J. Mol. Biol.* **1993**, *234*, 1084.
(3) Damante, G.; Di Lauro, R. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5388.
(4) Gruschus, J. M.; Tsao, D. H.; Wang, L. H.; Nirenberg, M.; Ferretti, J. A. *Biochemistry* **1997**, *36*, 5372.
(5) Zeng, W. L.; Andrew, D. J.; Mathies, L. D.; Horner, M. A.; Scott, M. P. *Development* **1993**, *118*, 339.
(6) Joshi, R.; Passner, J. M.; Rohs, R.; Jain, R.; Sosinsky, A.; Crickmore, M. A.; Jacob, V.; Aggarwal, A. K.; Honig, B.; Mann, R. S. *Cell* **2007**, *131*, 530.
(7) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. *J. Am. Chem. Soc.* **2007**, *129*, 738.
(8) Crane-Robinson, C.; Dragan, A. I.; Privalov, P. L. *Trends Biochem. Sci.* **2006**, *31*, 547.
(9) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121.
(10) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comput Biol* **2008**, *4*, e1000060.
(11) Dragan, A. I.; Li, Z.; Makeyeva, E. N.; Milgotina, E. I.; Liu, Y.; Crane-Robinson, C.; Privalov, P. L. *Biochemistry* **2006**, *45*, 141.
(12) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868.
(13) Marcovitz, A.; Levy, Y. *Biophys. J.* **2009**, *96*, 4212.
(14) Fuxreiter, M.; Mezei, M.; Simon, I.; Osman, R. *Biophys. J.* **2005**, *89*, 903.
(15) Jana, B.; Pal, S.; Maiti, P. K.; Lin, S. T.; Hynes, J. T.; Bagchi, B. *J. Phys. Chem. B* **2006**, *110*, 19611.
(16) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087.
(17) Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12855.

JA9052784