DNA–PROTEIN INTERACTIONS

Applying a genetic cantilever

Marco E. Bianchi and David M. J. Lilley

INTERACTIONS between DNA and proteins lie at the very heart of development. In the latest issue of *Cell*[1] Marius Clore and his colleagues zoom in near the atomic level to the complex cascade of switches that control the expression of key genes during eukaryotic development. Using nuclear magnetic resonance, they have solved the solution structure of a complex between a DNA octamer and the DNA-binding domain of SRY, the protein factor that is responsible for determining male gender in mammals[2].

SRY is a member of the HMG class of DNA-binding proteins that are characterized by their effect on DNA structure[3]. A striking feature of the SRY–DNA complex is that the DNA is bent and twisted, in marked contrast to DNA in complex with most repressor proteins, in which the protein recognizes a target binding sequence but leaves the DNA largely unperturbed. This effect of SRY illustrates the inherent deformability of DNA, and substantiates proposals that HMG-box proteins function as architectural components to ply and mould DNA[4]. The local structural deformation of the double helix induced by SRY might mediate effects at a distance through the mechanical displacement of DNA segments (and associated factors) at either side of the point of flexure.

The HMG domain was first identified as duplicated 80-amino-acid regions in the abundant non-histone nuclear protein HMG1. Sequences with significant similarity were then found in a variety of transcription factors, including SRY, LEF-1 and UBF and in more than 100 other DNA-binding proteins (reviewed in ref. 5). These proteins appear to be subdivided into classes depending on their sequence selectivity, from the essentially nonspecific HMG1 and 2, through UBF which is partially specific, to SRY, LEF-1 and TCF-1, for which proper consensus DNA-binding sites can be defined. What all these proteins have in common is their marked effect on DNA structure, and an affinity for distorted DNA structures such as four-way junctions[6] and *cis*-platinum adducts[7]. HMG proteins bend their target DNA, retarding electrophoretic mobility and promoting cyclization rates[8], and can even participate as auxiliary factors in site-specific recombination reactions[8,9]. This has led to the idea that the HMG box is an all-purpose DNA bender/wrapper/looper which can be recruited for a variety of DNA transactions, including transcription, repair and recombination — a kind of eukaryotic version of the IHF and HU proteins of *Escherichia coli*.

Three earlier NMR structures of HMG boxes of the non-sequence-selective class[10–12] exhibited the same general fold with minor differences, in which three α-helical segments formed an L-shaped structure stabilized by a hydrophobic core. The longer part of the 'L' comprised the helical C-terminal region and an extended N-terminal section. In the SRY–DNA complex[1] the overall fold is closely similar, but the angle of the 'L' is more obtuse. The concave surface of the protein is perfectly located in the minor groove of the DNA where it buries a large surface area. To achieve this, the DNA is distorted by a succession of positive roll angles and bent back by 70 to 80 degrees. The protein is locked down on to the DNA by a series of interactions with the backbones of both helices. The minor groove is extensively widened to accommodate the protein, accomplished principally by a wedge of five residues that prise apart the central region of the DNA.

A factor in the bending of the DNA is the intercalation of an isoleucine residue at the centre of the structure[1,13]. The extent of bending can vary between different HMG proteins: for example, LEF-1 induces a considerably greater bending than SRY[9]. The protein structure itself might influence the degree of bending, with target DNA being bent more by HMG boxes having a more acute-angled 'L', or the protein may adjust this angle on binding to the DNA; in this regard it would be interesting to know the structure of free SRY in solution. What is clear is that very different binding surfaces in an expanded minor groove can bend and unwind DNA in a similar manner, because the overall shape of the SRY–DNA complex is strongly reminiscent of that formed by the TATA-box-binding protein with its target DNA sequence[14,15].

It is not apparent why four-way junctions should be a universal target for HMG-box proteins. No convex surfaces are presented in the folded junction, but the real target may be the unfolded junction found at low salt concentrations, or some other conformation induced by protein binding. Another outstanding issue is the origin of the sequence specificity of SRY and the LEF-1/TCF-1 families and of the sequence indifference of the HMG1 family. A domain-swap experiment showed that a chimaeric protein consisting of the long arm of TCF-1 fused to the short arm of HMG1 retains the binding and bending ability of TCF-1 (ref. 16). Clore and colleagues[1] have identified a number of contacts to DNA bases distributed along the entire binding surface that would be expected to mediate specificity. Many of these positions are altered in the non-sequence-selective HMG-box proteins.

The structure of the SRY–DNA complex offers a rare opportunity for molecular interpretation of mutations that cause clinical effects. Natural mutants of human SRY are associated with abnormal sexual development and all involve the HMG box. A few mutations can be inherited, which means that sometimes they can be recovered from normal males, and so must be less disruptive than mutations that cause malformation of the gonads and cannot be transmitted. Interestingly, all inherited point mutations map to sites that do not interact with DNA; DNA binding overall is not grossly altered[17,18], suggesting by default that the stability of the protein could be affected. Some *de novo* mutations map to sites that would severely disrupt protein packing and four involve residues that contact DNA and would be expected to reduce the DNA affinity of SRY.

One mutation, involving an isoleucine substitution for a methionine at residue 64, is remarkable in that it does not appreciably reduce the affinity or the specificity for the DNA target but affects the extent of DNA deformation, as detected by the alteration of the electrophoretic mobility of the complex[18]. The methionine involved is part of the wedge that kinks DNA by unstacking the central base pairs, so this conservative substitution might reduce the surface of contact between protein and DNA. Thus, a modest alteration of the SRY-induced DNA structure can translate into a severe disruption of developmental patterns and a serious clinical condition. □

Marco E. Bianchi is in the Department of Genetics and Microbiology, University of Milano, via Celoria 26, 20133 Milano, Italy. David M. J. Lilley is in the CRC Nucleic Acids Structure Research Group, Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK.

1. Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. *Cell* **81**, 705–714 (1995).
2. Sinclair, A. *et al. Nature* **346**, 240–244 (1990).
3. Lilley, D. M. J. *Nature* **357**, 282–283 (1992).
4. Grosschedl, R., Giese, K. & Pagel, J. *Trends Genet.* **13**, 94–100 (1994).
5. Bianchi, M. E. in *DNA–Protein: Structural Interactions* (ed. Lilley, D. M. J.) (IRL, Oxford, in the press).
6. Bianchi, M. E., Beltrame, M. & Paonessa, G. *Science* **243**, 1056–1059 (1989).
7. Pil, P. M. & Lippard, S. J. *Science* **256**, 234–237 (1992).
8. Paull, T. T., Haykinson, M. J. & Johnson, R. C. *Genes Dev.* **7**, 1521–1534 (1993).
9. Giese, K. *et al. Cell* **69**, 185–195 (1992).
10. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C. & Norman, D. G. *Nucleic Acids Res.* **21**, 3427–3436 (1993).
11. Weir, H. M. *et al. EMBO J.* **12**, 1311–1319 (1993).
12. Jones, D. N. M. *et al. Structure* **2**, 609–627 (1994).
13. King, C.-Y. & Weiss, M. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11990–11994 (1993).
14. Kim, Y. C., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–520 (1993).
15. Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **365**, 520–527 (1993).
16. Read, C. M. *et al. EMBO J.* **13**, 5639–5646 (1994).
17. Harley, V. R. *et al. Science* **225**, 453–456 (1992).
18. Pontiggia, A. *et al. EMBO J.* **13**, 6115–6124 (1994).