DNA–protein interactions: **IHF – the master bender**

Andrew Travers

The crystal structure of *Escherichia coli* integration host factor complexed with DNA reveals how the sequence-specificity of DNA binding can be determined almost entirely by the structural features of the DNA itself and not by direct readout of the base sequence. There are lessons to be drawn for other DNA-binding motifs.

Address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

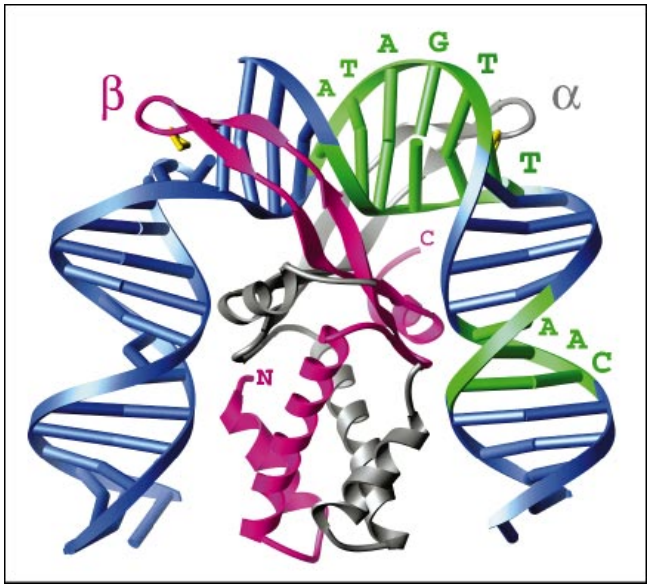
Electronic identifier: 0960-9822-007-R0252

Current Biology 1997, 7:R252–R254

The bacterial type II DNA-binding proteins, which include the abundant HU and the phage-encoded TF1 [1,2], are believed to condense their cognate genomes by binding to them cooperatively and inducing coherent bends in the DNA. These proteins bind to DNA with little, if any, sequence specificity. However, another member of the family, the integration host factor or IHF (for review, see [3]), which is required for site-specific recombination, DNA replication and transcription, binds at specific sites characterized by a limited consensus sequence [4]. It has long been thought that IHF works by inducing a large bend at its binding site [5,6], but with the recent solution of the crystal structure of *Escherichia coli* IHF complexed to a sequence from the phage lambda H′ site, the true magnitude of the distortion is now apparent [7]. Within two-and-a-half turns of the double helix, the DNA executes a U-turn with an overall bend angle of at least 160° and possibly in excess of 180° (Fig. 1). IHF thus heads an exclusive list of big benders including, to date, lymphoid enhancer-binding factor 1 (LEF-1; 120°), high-mobility group protein D (HMG-D; >90°), catabolite gene activator protein (CAP; ~90°) and TATA-binding protein (TBP; 80°) [8–11].

The crystallization of an IHF–DNA complex required that one or both strands of the bound DNA be discontinuous, a device that was also successful in producing crystals of the CAP–DNA complex [9]. In the IHF–DNA complex, the two ~10 kDa subunits of the IHF heterodimer are intertwined to form a compact core from which two long β ribbon arms extend, as in the structure of homologous HU [12,13]. As predicted both from the structure of free HU and from genetic studies, the arms track along the minor groove from the inside to the outside of the wrapped DNA, where they terminate at the two substantial kinks. In addition to these interactions *via* the β arms, IHF also clamps the hairpin by minor-groove contacts to the core of the dimer. All the contacts to the DNA are either in the minor groove or are part of an extensive network of electrostatic interactions with the phosphate backbones.

Figure 1

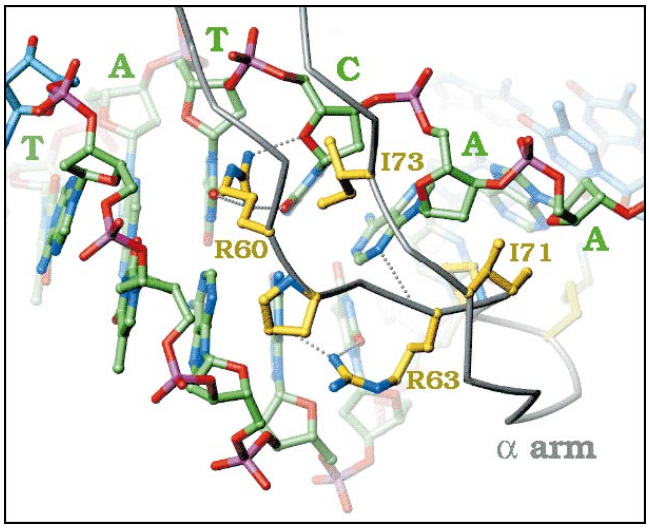


Complex of IHF with site H′. The α subunit is shown in silver and the β subunit in pink. The consensus sequence is highlighted in green and interacts mainly with the arm of α and the body of β. (Reproduced from [7] with the kind permission of P.A. Rice.)

The IHF–DNA structure confirms and extends previous insights into the mechanism by which proteins introduce substantial bends into DNA. All the big benders concentrate the bend at kinks where a single base-step is unstacked and opened towards the minor groove, usually with a positive roll angle of 40–50°. Of these proteins, all but CAP interact principally with the minor groove of DNA and induce the kink by the partial intercalation of a hydrophobic residue between adjacent base pairs. In the type II DNA-binding proteins, this residue is an absolutely conserved proline [4] located at the tips of the β arms. Other proteins that bind the minor groove induce kinks by the insertion of phenylalanine (TBP), leucine (purR), isoleucine (SRY) or methionine (LEF-1) [8,10,11,14–16]. The greater structural rigidity of the proline side-chain may fix the flexible β arms of IHF and stabilize the disposition of the DNA in the immediate vicinity of the kink.

Two distinct mechanisms maintain the DNA bend in the IHF–DNA complex. On the outside of the bend, the hydrophobic intercalation stabilizes the opening of the

Figure 2



Distortion of DNA duplex adjacent to the site of intercalation by the IHF α subunit. The protein Cα trace is shown in grey, with side-chains that interact directly with the DNA in yellow. Carbons in the consensus sequence bases are green; others are blue. (Reproduced from [7] with the kind permission of P.A. Rice.)

minor groove. On the inside, charge neutralization counteracts the enhanced repulsion between the phosphates on opposite sides of the narrowed grooves. This combined 'push–pull' action is also used by the HMG-domain proteins, notably by LEF-1, a short basic region of which neutralizes charges across a narrowed major groove induced by the widening of the minor groove on the opposite face of the double helix [8]. By contrast, TBP stabilizes the induced bend entirely by extensive minor-groove interactions on the outer face of the bend [10,11], and the bending induced by CAP [9] or the histone octamer [17] depends exclusively on charge neutralization on the inner face.

The sequence dependency of IHF binding is an example *par excellence* of indirect readout, in which the conformation of DNA, rather than base-specific contacts, determines the binding site. The 'consensus' sequence consists of two short elements separated by approximately half a turn in only one of the two half-sites. Both these elements contain the trinucleotide TTG. In the first of these elements, TATCAA, two arginines reach into the minor groove to contact the conserved bases. By themselves, these interactions are insufficient to explain the selectivity. Hence, in the absence of any strong sequence-specific contacts, the selection for the TTG trinucleotide must reflect the physiochemical properties of the sequence. The TT/AA step is the site of intercalation of the α subunit, and here the close contact and hydrophobic interaction may be favoured by the lack of a polar 2-amino group in the minor groove [18]. However, the selection for the remainder of the sequence is apparently more subtle.

The distortion at the site of intercalation buckles the A–T base-pair in the CA step (Fig. 2). In the structure, this buckle is resolved asymmetrically at the adjacent TC step by a large tilt angle between the two pyrimidine bases. Interestingly, this resolution would be sterically hindered if a purine base replaced the C. In this position, a T would also be energetically disfavoured, because its methyl group would be exposed to the solvent rather than packing against an adjacent base. The clear message from the crystal structure is that IHF selects its binding site largely on the basis of the structural constraints imposed by the DNA, and sequence 'recognition' is thus indirect.

The second conserved sequence element is located where the IHF α subunit forms one side of the DNA clamp. Here, the minor groove is narrow, consistent with the conservation of the AA/TT step. However, the key to the conservation of the TG/CA step may be its flexibility. Analysis of the crystal structures of DNA oligomers shows that this step is, with the possible exception of TA, the most conformationally variable of all steps [19]. In the CAP–DNA complex, the protein kinks its binding site at TG steps [9], but at the clamp site, the ability of TG to adopt a high twist angle may be crucial. Again, it is DNA structure rather than specific contacts that determines recognition. On the other side of the clamp, the minor groove is again narrow. In the IHF–DNA crystal structure, the β subunit contacts a short oligo(dA) tract, a sequence that favours IHF binding. Here again, it is the ability of this sequence to adopt the required conformation that seems to be the important determinant.

The structure of the A-tract in the IHF–DNA complex also illuminates the long-running debate on the structure of oligo(dA) tracts in particular, and on the structure of DNA in solution in general. When helically phased, such tracts confer intrinsic curvature on DNA. One proposed explanation of this phenomenon is that, in solution, these tracts are themselves curved towards the minor groove. Yet this is at variance with numerous crystal structures in which such tracts are invariably straight [20,21]. The structure of the A-tract in the complex is virtually identical to the structures of these straight tracts in free DNA, indicating that the latter are biologically relevant. Indeed, the overall pattern of DNA curvature in the complex fits very well with the view, proposed for DNA free in solution, that the helical axis is deflected in regions where the minor groove is on the outside of the bent DNA — such that roll angles are positive — but is straight with a zero roll angle when the minor groove is on the inside [20].

Eukaryotic equivalents

In eukaryotes, the equivalents of the bacterial type II DNA-binding proteins are the HMG-domain proteins. The abundant, sequence-independent members of this family, the HMG1/2 proteins, are involved in the maintenance of

chromatin structure [22], and the 'sequence-specific' transcription factors containing this DNA-binding domain introduce a sharp bend into the DNA, thereby bringing other DNA-bound proteins into close spatial proximity [23]. Although the structures of type II DNA-binding proteins and HMG-domain proteins are completely different, in certain natural situations they are functionally equivalent. HU can compensate for the loss of the yeast mitochondrial protein ABF2 and *vice versa* [24], and the HMG-domain proteins NHP6A from yeast [25] and HMG-D from *Drosophila* (S.S. Ner, unpublished observations) can phenotypically rescue *E. coli* strains that lack HU. This functional equivalence argues that these proteins affect DNA structure in a comparable manner, and indeed the structural parallels between the binding of IHF and of HMG-domain proteins to DNA are surprisingly strong. Both types of protein widen the minor groove by partial intercalation of a hydrophobic residue on the outside of the bend, and both minimize the electrostatic repulsion across the narrowed grooves on the inside of the bend by charge neutralization.

It is also remarkable that the binding sites for both IHF and the HMG-domain proteins contain the trinucleotide TTG in their most conserved regions [4,26]. In both cases, partial intercalation occurs at the AA/TT step [8,15,16] and, at least in the LEF-1–DNA complex, as in the IHF–DNA complex, there is a rapid reversion to a B-like DNA structure distal to the TT step. However, the structures of the conserved trinucleotide are not wholly equivalent in the two complexes. In the IHF–DNA complex, the compressed major groove is stabilized by the insertion of the methyl group of a thymine in a hydrophobic pocket, whereas the binding of HMG-D to its cognate binding site is enhanced by the removal of thymine methyl groups at the assumed site of intercalation [18]. Nevertheless, the parallels between the HMG and IHF binding motifs are sufficiently strong that it seems plausible that the selection of TTG by HMG proteins may also be largely dependent on structural considerations.

An unanswered question is the biological rationale for the heterodimeric nature of IHF. Because the DNA in the immediate vicinity of the kink induced by the α chain is 'relaxed' by the single strand nick, it is unclear whether there are detailed differences between the naturally induced deformations by the α and β chains in this region, although the structure suggests that the approaches made by the β chain may be less close. Again there is a possible analogy with the HMG1/2 proteins of vertebrates. These proteins contain two tandem HMG domains, which differ in the structural selectivity of DNA binding [27] and might thus be regarded as fused heterodimers. It will be very interesting to see whether the structure of the HMG1–DNA complex mirrors the pseudosymmetry of the IHF–DNA complex.

References

1. Rouvière-Yaniv J, Gros F: Characterisation of a novel low-molecular weight DNA binding protein from ***Escherichia coli.*** *Proc Natl Acad Sci USA* 1975, 72:3428-3432.
2. Wilson DL, Geiduschek EP: A template selective inhibitor of ***in vitro*** transcription. *Proc Natl Acad Sci USA* 1969, 62:514–520.
3. Nash HA: The HU and IHF proteins: accessory factors for complex protein–DNA assemblies. In *Regulation of gene expression in* Escherichia coli. Edited by Lin ECC, Lynch AS. Austin Texas: R.G. Landes; 1996:149–179.
4. Goodrich JA, Schwartz ML, McClure WR: Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for ***Escherichia coli*** integration host factor (IHF). *Nucleic Acids Res* 1990, 18:4993–5000.
5. Moitioso de Vargas L, Kim S, Landy A: DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* 1989, 244:1457–1461.
6. Goodman SD, Nash HA: Functional replacement of a protein-induced bend in a recombination site. *Nature* 1989, 341:251–254.
7. Rice PA, Yang S-W, Mizuuchi K, Nash HA: Crystal structure of an IHF–DNA complex: a protein-induced U-turn. *Cell* 1996, 87:1295–1306.
8. Love JJ, Xiang L, Case DA, Giese K, Grosschedl R, Wright PE: Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 1995, 376:791–795.
9. Schultz SC, Shields GC, Steitz TA: Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science* 1991, 253:1001–1007.
10 Kim Y, Geiger JH, Hahn S, Sigler PB: Crystal structure of a yeast TBP/TATA-box complex. *Nature* 1993, 365:512–520.
11. Kim JL, Nikolov DB, Burley SK: Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 1993, 365:520–527.
12. Tanaka I, Appelt K, Dijk J, White SW, Wilson KS: 3 Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* 1984, 310:376–381.
13. Vis H, Mariani M, Vorgias CE, Wilson KS, Kaptein R, Boelens R: Solution structure of the HU protein from ***Bacillus stearothermophilus.*** *J Mol Biol* 1995, 254:692–703.
14. Schumacher MA, Choi KY, Zalkin H, Brennan RG: Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science* 1994, 266:763–770.
15. King CY, Weiss MA: The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove — a topological mechanism of sequence specificity. *Proc Natl Acad Sci USA* 1993, 90:11990–11994.
16. Werner MH, Huth JR, Gronenborn AM, Clore GM: Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY–DNA complex. *Cell* 1995 81:705–714.
17. Mirzabekov AD, Rich A: Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. *Proc Natl Acad Sci USA* 1979, 76:1181–1121.
18. Bailly C, Payet D, Travers AA, Waring MJ: PCR-based development of DNA substrates containing modified bases: an efficient system for investigating the role of the exocyclic groups in chemical and structural recognition by minor groove binding drugs and proteins. *Proc Natl Acad Sci USA* 1996, 93:13623–13628.
19. El Hassan MA, Calladine CR: Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* 1996, 259:95–103.
20. Nelson HCM, Finch JT, Luisi BF, Klug A: The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature* 1987, 330:221–226.
21. DiGabriele AD, Sanderson MR, Steitz TA: Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc Natl Acad Sci USA* 1989, 86:1816–1820.
22. Ner SS, Travers AA: HMG-D, the ***Drosophila melanogaster*** homologue of HMG1 protein is associated with early embryonic chromatin in the absence of histone H1. *EMBO J* 1994, 13:1817–1822.
23. Grosschedl R, Giese K, Pagel J: HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet* 1994, 10:94–100.
24. Megraw TL, Chae C-B: Functional complementarity between HMG1-like yeast mitochondrial histone protein HM and the bacterial histone-like protein HU. *J Biol Chem* 1993, 268:12758–12763.
25. Paull TT, Johnson RC: DNA looping by ***Saccharomyces cerevisiae*** high mobility group proteins NHP6A/B. *J Biol Chem* 1995, 270:8744–8754.
26. Churchill MEA, Jones DNM, Glaser T, Hefner H, Searles MA, Travers AA: HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J* 1995 14:1264–1275.
27. Teo S-H, Grasser KD, Thomas JO: Differences in the DNA-binding properties of the HMG-box domains of HMG1 and the sex-determining factor SRY. *Eur J Biochem* 1995 230:943–950.