Recognition of distorted DNA structures by HMG domains

Andrew Travers

Recent biochemical and structural studies have shown that the preferential recognition of distorted DNA structures, including DNA bulges, four-way junctions and *cis*-platinated DNA, by HMG domains is dependent on residues immediately preceding the second α helix of the L-shaped HMG domain.

Addresses
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK; e-mail: aat@mrc-lmb.cam.ac.uk

Current Opinion in Structural Biology 2000, **10**:102–109

0959-440X/00/$ – see front matter

Abbreviations

FRET	fluorescence resonance energy transfer
HMG	high-mobility group
TBP	TATA-binding protein

Introduction

HMG DNA-binding domains both bend DNA substantially and bind to distorted DNA structures, such as four-way junctions, with high affinity. These domains are subdivided into two major classes — the sequence-specific domains found in transcription factors and the non-sequence-specific domains found in abundant chromosomal proteins. The latter are further subdivided into A and B domains. The differences in selectivity between the sequence-specific and non-sequence-specific domains can be ascribed to a few residues, whereas the determinants of the recognition of distorted DNA structures are located in a short region between helices I and II of the HMG domain.

The high-mobility group (HMG) domain is a DNA-binding motif that occurs both in sequence-specific transcription factors, including LEF-1 and SRY, and in abundant chromosomal proteins typified by the vertebrate proteins HMG1 and HMG2, which bind DNA with little or no sequence specificity. The global fold of this domain is well conserved and consists of three α helices (I–III) arranged in an L shape [1–7]. This fold is maintained by one major and two minor hydrophobic cores, the former consisting of three aromatic residues stacked edge to face. The relative disposition of these three residues is maintained in all structures of free and complexed HMG domains, with the exception of SRY, for which the reported arrangement is different [7]. The concave surface of the HMG domain contacts the minor groove of linear B-form DNA, inducing a substantial bend [5,7]. In addition, both sequence-specific and non-sequence-specific HMG domains exhibit a strong preference for binding to distorted DNA structures, including four-way junctions, *cis*-platinated DNA, DNA microcircles and supercoiled DNA [8–13].

Vertebrate proteins HMG1, HMG2 and HMG2a contain two HMG domains, A and B, connected by a basic linker to a long, highly acidic tail (Figure 1). The invertebrate counterparts, such as *Drosophila* HMG-D and HMG-Z, contain only a single HMG domain followed by an extended basic region and a short acidic tail. By contrast, the *Saccharomyces* counterparts NHP6A and NHP6B have a basic N-terminal extension and no acidic tail. Plant HMG domain proteins also contain a basic N-terminal extension, but have acidic tails of variable length. The sequence-specific domains are structurally more homologous to the B domain of HMG1 than to the A domain, but differ by containing a shorter helix III. They also contain a short extended region beyond the HMG fold, with additional DNA-binding determinants. The HMG domains of NHP6A and HMG-D are structurally highly homologous to the B-domain of HMG1, except that in NHP6A and also possibly in HMG-D, helix III is bent rather than straight. By contrast, the A domain of HMG1 differs significantly from the B domain with respect to the relative disposition of helix I [4]. In addition, in the B domain helix I is sharply bent, whereas in the A domain this helix is straight. The trajectory of the helix I–II loop also differs between these two domains, that in the A domain being two amino acids longer.

So far, three structures of HMG domains with their DNA binding sites have been described: solution structures of the SRY domain with a DNA octamer [7] and the LEF-1 domain with a DNA 15-mer [5] and, more recently, the crystal structure of the A domain of HMG1 with a 20-mer containing a *cis*-platinum adduct [14••]. In addition, crystals of an HMG-D–decamer complex diffracting to 2.2 Å resolution have been reported [15], but no structure is yet available. Three modelled structures of HMG domains with different DNA species have also been described: an HMG-D–DNA complex based on molecular dynamics simulations using the coordinates of TATA-binding protein (TBP)-bound DNA as a starting point [16]; a complex of NHP6A with duplex DNA derived from nuclear Overhauser enhancement (NOE) data [17••]; and a complex of HMG-D with dA_2-bulge DNA built from footprinting data [18•]. The last two models used the coordinates of the DNA in the LEF-1–DNA complex as a starting point for modelling the binding of the HMG domain. Representative structures are shown in Figure 2a, whereas Figure 2b summarises the observed and predicted amino acid–DNA contacts.

Structural basis of DNA binding and bending

The general features of the described HMG domain–DNA complexes are highly conserved. In B-type domain complexes, the DNA-binding face of the domain presents a hydrophobic surface that conforms to a wide and shallow minor groove. In the centre of the surface, a hydrophobic

Figure 1

Schematic structures of HMG domain proteins of the HMG1/2 class from different organisms. The structure of the transcription factor LEF-1 is shown for comparison.

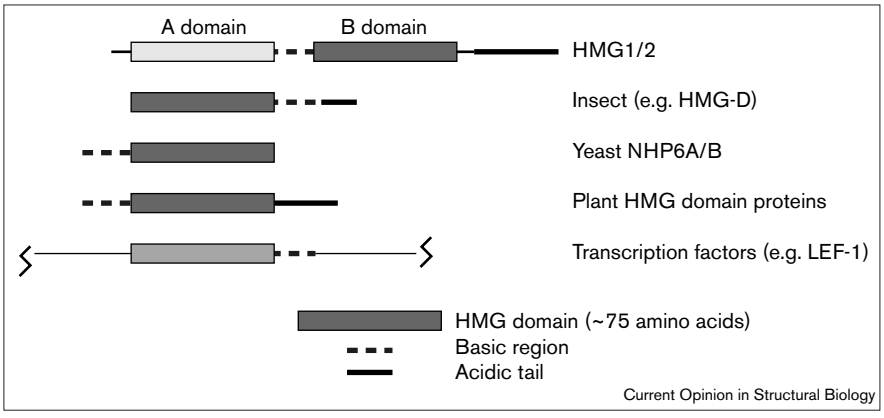
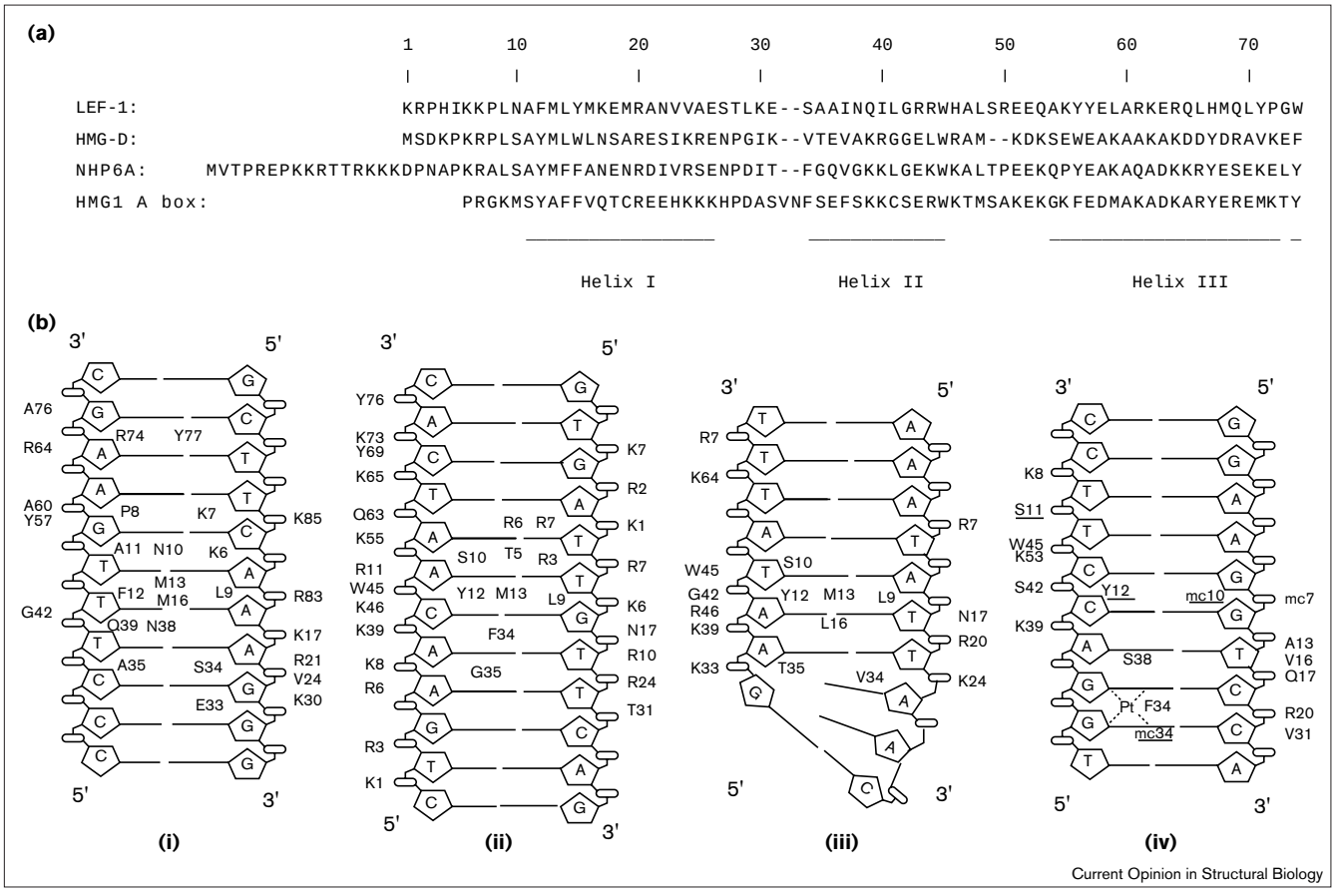


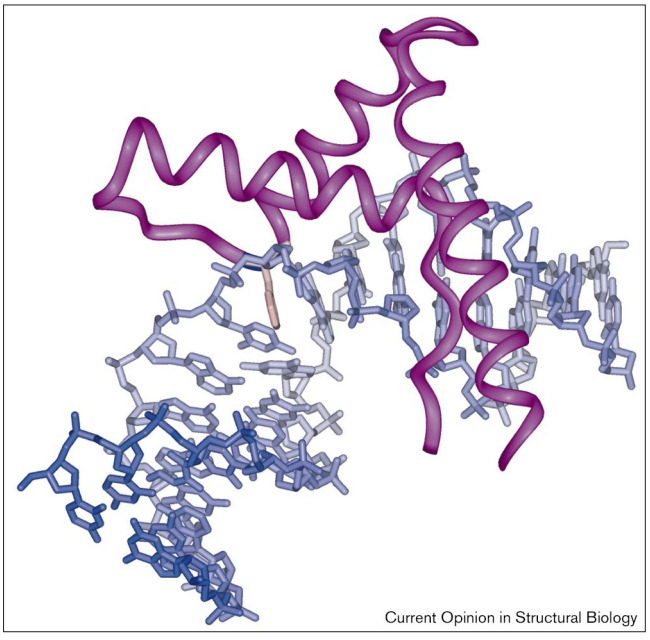
Figure 2



(a) Sequences of different HMG domains showing conserved residues. The numbering of the residues is by homology and corresponds to that used in the text. **(b)** Cylindrical projections of protein–DNA contacts made by **(i)** LEF-1 with its consensus binding site [5], **(ii)** NHP6A with the SRY DNA binding site [17••], **(iii)** HMG-D with bulged DNA [18•] and **(iv)** domain A of HMG1 with *cis*-platinated DNA [14••]. The amino acids are positioned on the basis of proximity. Underlined residues indicate water-mediated contacts and mc indicates a mainchain contact. The contacts shown in (ii) and (iii) are mainly inferred from modelling.

wedge is inserted deep into the minor groove of the DNA. Within this wedge, one residue (methionine in LEF-1 [5], NHP6A [17••] and HMG-D [18•]; isoleucine in SRY [19]) partially intercalates between two base pairs (Figure 2). With the exception of one model [15], in which Leu9 and Met13 partially and alternatively intercalate into successive base steps, the components of this wedge are conserved, comprising residues 9, 12, 13 and 43. The

Figure 3



Crystal structure of domain A of HMG1 (purple) with *cis*-platinated DNA (blue) [14••]; note the insertion of phenylalanine (pink) into the widened minor groove.

hydrophobic surface is flanked by a set of conserved basic residues that bind to the phosphodiester backbones and stabilise the complex. The interactions of the A domain of HMG1, as exemplified by its complex with *cis*-platinated DNA, are similar but, in this case, the domain lacks the intercalating residue at position 13 and consequently does not fit as deeply into the minor groove [14••] (Figure 3).

What is the basis of the distinction between sequence-specific and non-sequence-specific binding? At position 10, asparagine is wholly conserved among the known sequence-specific HMG domains, whereas in the HMG domains of HMG1 and its counterparts, this residue is normally serine (with one exception, dSSRP1, in which it is threonine) [20]. The striking correlation between the nature of this residue and the sequence specificity of the domain has been previously noted [21]. In the LEF-1–DNA and SRY–DNA complexes, Asn10 makes tripartite polar contacts with the C–G and A–T base pairs preceding the site of intercalation of Met13 and Ile13, respectively [5,7]. The sidechain of Ser10 is shorter and, as modelled [17••,18••], could, in principle, make direct contacts to a pyrimidine base or indirect contacts to a purine through water-mediated hydrogen bonds. Such contacts, in contrast to those made by Asn10, which should specify the CpA step in the SRY-binding site, would be expected to be essentially sequence-independent and thus account for the lack of sequence specificity of this class of HMG domain. Nevertheless, HMG-D has been reported to show a preference for binding to sequences that contain a TpG/CpA base step [22]. Interestingly, when binding to the same DNA sequence as one tested for SRY-specific binding, NHP6A is modelled to intercalate via Met13 at a TpG/CpA step, rather than at the adjacent ApA/TpT step (into which Ile13 of SRY intercalates) and also binds in the opposite orientation to SRY [7,17••]. Although NHP6A and HMG-D are completely homologous over the relevant region of the domain (residues 9–14: LSAYML), the basis for this moderate selectivity is uncertain. One possibility is that, for NHP6A and HMG-D, complex formation is initiated at a prebent YpR step, which the methionine sidechain could easily penetrate [17••].

A second region in which the sequence-specific and the non-sequence-specific domains differ is the helix I–II loop immediately before helix II. In the sequence-specific domains, residue 34 is hydrophilic. In the LEF-1–DNA complex, Ser34, together with Asn38, hydrogen bonds to both base pairs of an ApA base step, thus contributing to the specificity of binding. In the SRY–DNA complex, the DNA is too short to determine whether the HMG domain makes similar interactions in this region. By contrast, in non-sequence-specific domains, residue 34 is invariably hydrophobic and is postulated to intercalate two base steps from the site of intercalation of Met13 [16]. This prediction has been verified for Phe34 of the A domain of HMG1 binding to *cis*-platinated DNA [14••] and for Val34 of HMG-D binding to bulged DNA in solution [18•]. In the latter case, there is also evidence that Thr35 intercalates at the same step, although the corresponding residue in the A domain (Ser35) does not contact DNA [14••,18•]. A different mode of interaction has been proposed for Phe34 of NHP6A. In free NHP6A, Phe34 extends the hydrophobic patch comprising Leu9, Tyr12, Met13 and Ala16 [16] on the concave surface of the protein. Allain *et al.* [17••] propose that this orientation is maintained in the complex, so that Phe34 would be inserted in the minor groove almost perpendicular to the edge of two base pairs. This insertion, which strictly is not an intercalation, would force apart the base pairs and induce a bend at the base step adjacent to that intercalated by Met13, but should not seriously compromise minor groove binding.

The sequence selectivity of the transcription factor HMG domains may also be dependent on a minor groove interaction of a tyrosine residue in an extended region distal to helix III. In LEF-1, this residue (Tyr77) penetrates deep into the minor groove and contacts guanine and thymine bases in a CT base step. In the non-sequence-specific HMG domains, helix III is usually longer and there is no corresponding tyrosine residue.

These studies reveal that differences in the specificity and bending properties of the HMG domains depend on changes in only a few residues. Functional versatility is accomplished by amino acid variation, but with little accompanying variation in the overall configuration of the fold.

DNA binding and bending

Estimates of the extent of bending induced by the binding of a single HMG domain vary substantially according

to the method used. For sequence-specific domains, bend angles of 117 ± 10° and 70–80° were determined for the solution structures of LEF-1–DNA and SRY–DNA complexes, respectively [5,7]. These values compare with 80° and 75° for similar complexes, as determined by polyacrylamide gel electrophoresis [23,24]. In contrast, the extent of bending induced by the HMG domains of the abundant chromosomal HMG proteins is uncertain and its determination has been complicated by the lack of sequence-specific binding. Recently, however, several studies have determined the deformation induced by these proteins or their isolated HMG domains by measuring the induced decrease in the end-to-end distance of a short DNA fragment upon protein binding using fluorescence resonance energy transfer (FRET). Lorenz *et al.* [25•] showed that the binding of yeast NHP6A, *Drosophila* HMG-D and HMG-Z, and *Chironomus* cHMG1a to an 18 base pair DNA fragment with one dA_2 bulge all resulted in a conformational change in the DNA structure consistent with a total bend angle of 95 ± 5°. The removal of the C-terminal acidic stretch from the last three proteins did not change the bend angle. In good agreement with this value, using the same technique with a slightly different assumed R_0 value (45 Å compared with 50 Å), Jamieson *et al.* [26•] determined a bend angle of approximately 90° for the complex of the B domain of HMG1 with *cis*-platinated DNA. In contrast to these measured and calculated bending angles, however, Heyduk *et al.* [27] reported that a single molecule of cHMG1a induces a bend of about 150°. Using a linear 30 base pair DNA sequence, they determined a decrease in the dye-to-dye distance upon binding cHMG1a from 100 Å (free, linear DNA) to 50.5 Å (bound DNA) using FRET. This corresponds to an induced bend angle of 150°, assuming that a single HMG domain untwists the bound DNA by 90°. Similar values were also obtained for several plant HMG domain proteins [28]. A bend angle of this magnitude within the span of the HMG-domain-binding site would almost certainly result in steric clashes between the DNA strands on the inside of the apex of the bend. A possible explanation for this result and the discrepancy with other reports is that the 30 base pair fragment used can accommodate two bound HMG molecules. On a linear 30 base pair DNA molecule, an HMG domain could occupy a number of different positions, resulting in a spread of 10–15 Å in possible end-to-end distances, whereas if two HMG molecules are bound, the sites available to each individual protein molecule would be restricted. A bend angle of 150° induced by two cHMG1a molecules binding to the DNA fragment, rather than the assumed single molecule, would also be more consistent with the bonding angle of 190 ± 10° for a complex of two HMG-D molecules with a 27 base pair DNA fragment containing two dA_2 bulges [25•]. Taken together, these data indicate that the bend angles induced by the B-type HMG domains, including the sequence-specific domains, are all very similar (falling in the narrow range of 90 to 120°).

The value of approximately 95° determined by FRET for the binding a single HMG domain to prebent DNA compares with values of 120–135° and at least 70° calculated from the extent of circularisation of short duplex DNA fragments by NHP6A [13] and HMG-D [29], respectively. Although a high degree of confidence can be placed in FRET distance measurements, it is possible that any instability of the HMG–DNA complexes used in these measurements could result in a slight underestimate of the actual bend angle.

In contrast to the binding of the B domain of HMG1 to *cis*-platinated DNA, the same ligand DNA is bent by only 61° when crystallised in a complex with the A domain [14••]. This is consistent with previous observations showing that the A domain is less effective at circularising an 88 base pair DNA fragment than the B domain [30]. The smaller bend angle in the HMG1 domain A–*cis*-platinum DNA complex relative to the HMG1 domain B–*cis*-platinum DNA complex [14••] can be ascribed to the lack of a second intercalating residue corresponding to Met13 in HMG-D (the corresponding residue in the A domain is alanine).

Bending determinants

The structural data have identified two major sites of bending induced by each of the B-type HMG domains. For the non-sequence-specific class, bending is induced principally by the partial intercalation of residue 13 in helix I and also by residues in the distal part of the helix I–helix II loop (34/35 in HMG-D and 34 in NHP6A). The sequence-specific class of HMG domain retains partial intercalation by residue 13 and, even though it lacks a hydrophobic residue at position 34, the LEF-1 domain still induces a bend at the corresponding base step [5], implying that structural complementarity between this part of the domain and the minor groove may in itself contribute to bending. In addition, however, the interaction of the DNA with the tyrosine residue distal to helix III is important for maximal DNA bending [5]. There is already substantial evidence that the mutation of Ile13 in SRY results in diminished bending [31] and Yen *et al.* [13] have also demonstrated that Met13 of NHP6A is essential for 75 base pair, but not for 98 base pair, circle formation. The 98 base pair circles are still smaller than those formed without the protein, however, suggesting the presence of another bending locus. Indeed, mutation of Phe34 to alanine in NHP6A reduces its affinity for linear DNA by only twofold, but substantially impairs DNA circularisation [17••]. By contrast, corresponding mutations of either Val34 or Thr35 of HMG-D reduce affinity for DNA by 5–10-fold [18•]; however, only the T35A mutant protein is substantially impaired in circularisation. More surprisingly, the mutation of Val34 to phenylalanine, while having little effect on affinity, causes defective circularisation to a similar extent to the Thr35Ala mutation. These data indicate that Thr35, but not Val34, is a critical residue for DNA bending, although both are important for binding. The differential effects of

the Val34Phe and Val34Ala mutations on DNA bending and binding also show that these two functions are at least partially separable.

Taken together, these observations emphasise the importance of the region in the immediate vicinity of the N terminus of helix II for DNA bending by HMG domains structurally homologous to the B domain of HMG1 and accord well with the results of structural studies and modelling, and, in particular, with the deep penetration of HMG-D Val34 and Thr35 into the minor groove [18•]. The striking functional differences between proteins with valine and phenylalanine at position 34 are consistent with the suggestion that Val34 may intercalate in a base step together with Thr35, whereas Phe34 abuts against the edges of two base pairs. This different mode of interaction with DNA could explain why the Val34Phe mutant of HMG-D is defective in circularisation, yet binds to linear DNA fragments with comparable affinity to the wild-type protein.

The structural basis of the introduction of a substantial DNA bend by an HMG domain is, in principle, similar to that observed for the TBP [32–34]. In both cases, the minor groove is widened by an extensive hydrophobic protein surface binding within the groove. This widening itself induces a positive roll angle, but the major contribution to the resultant bend is a consequence of kinks introduced by the partial intercalation of a hydrophobic residue at one or two base steps. The bacterial IHF heterodimer, which is a homologue of HU (the prokaryotic functional equivalent of the eukaryotic HMG1/2 class of proteins), also bends DNA by 180°, again by kinking the bound DNA by partial intercalation at two sites separated by eight base pairs.

Role of basic extensions

In the LEF-1–DNA complex, the basic C-terminal extension of the HMG domain is located in the compressed major groove opposite the widened minor groove and was postulated to stabilise the induced bend by charge neutralisation. This suggestion is consistent with many subsequent *in vitro* observations [12,23]. A similar location for the basic N-terminal extension of NHP6A was also observed [17••]. In this structure, a bend in helix III of the HMG domain serves to direct the basic region proximal to the N-terminal strand of the domain into the major groove. An analogous function has been suggested for the bend in helix III of HMG-D [18•].

Structural basis for the recognition of distorted DNA

The high affinity of HMG domains for distorted DNA structures is exhibited not only by the sequence-specific domains, but also by domains homologous to both the A and B domains of HMG1. Early models for the binding of the HMG domain to such DNA structures, including *cis*-platinated DNA [24,35], four-way junctions [36] and DNA containing a disulfide cross-link between the 5′ adenines of a 5′-AATT-3′ region in complementary strands of DNA [37], were based on the assumption that the apex of the DNA bend fits into the angle of the L-shaped surface of the protein. That is, the residue corresponding to Met13 in HMG-D intercalates either at the centre of the pre-existing bend or in the centre of the cross-over in a four-way junction. More recently, however, Pohler *et al.* [38] demonstrated that both sequence-specific and non-sequence-specific HMG domains bind to the extended, rather than the stacked X, form of four-way junctions and suggested, on the basis of molecular modelling, that the loop between helices I and II was located on the minor groove surface of the junction. This general location has been confirmed by a footprinting study by Webb and Thomas [39••], although the detailed interactions proposed by the latter authors differ from those of Pohler *et al.* [38]. Similarly, a footprinting study by Payet *et al.* [18•] on the binding of HMG-D to a dA_2-bulged DNA fragment demonstrated that the main protection from hydroxyl radical cleavage lies on one side of the bulge. These authors concluded that the HMG domain of HMG-D binds asymmetrically with respect to the bulge. All these later studies [18•,38,39••] implicate the residues immediately preceding the N terminus of helix II as the region interacting with the DNA distortion.

These conclusions are fully consistent with two recent definitive studies of the structures of complexes of HMG domains with *cis*-platinated DNA [14••,26•]. The crystal structure of a complex between the A domain of HMG1 and *cis*-platinated DNA [14••] shows that the A domain binds to one side of the *cis*-platinum lesion in an analogous manner to that deduced for the HMG domain of HMG-D binding to a DNA bulge. Also, Phe34, corresponding to Val34 in HMG-D, is shown to partially intercalate between the platinum cross-linked GpG site, thus increasing the roll angle of this step. Using footprinting data, Webb and Thomas [39••] have modelled the binding of HMG1 and the A domain alone to four-way junctions. The A domain confers on HMG1 its high selectivity for four-way junctions and, in the model, is inserted into the central hole on the minor groove side of the junction itself, whereas the B domain binds to one of the extended arms. The high selectivity of the A domain for four-way junctions in particular is postulated to be dependent both on the stacking of Phe34 against an exposed base pair at the end of one arm and on the greater positive charge in this region of the A domain relative to other domains. This latter feature could facilitate the insertion of the domain into the distorted structure and stabilise the resulting complex. This characteristic could explain the higher affinity of the A domain relative to B domain for four-way junctions.

Mutational studies reinforce the conclusion that the determinants of structural DNA recognition may reside in a different part of the domain from the primary intercalating residues and that the region immediately

preceding helix II is important for interaction with distorted DNA. In the SRY domain, the replacement of Ile13 with valine abolished binding to linear DNA, but had no effect on binding to four-way junctions. The replacement of the isoleucine by alanine actually enhanced its affinity for junctions [31]. Similarly, the mutation of a crucial tryptophan residue in the principal hydrophobic core of the A domain of HMG1 does not significantly affect binding to four-way junctions [24], although the domain itself is largely unfolded in solution [40]. Two other mutations, however, Arg9Gly and Pro28Ala, reduce the affinity of the isolated A domain for four-way junctions [24]. In the model of Webb and Thomas [39••], Arg9 makes an electrostatic contact with a phosphate in one of the junction arms, whereas the Pro28Ala mutation would be expected to alter the relative orientation of helices I and II, which protrude into the open centre of the junction. Similarly, mutations at positions 34 and 35 of HMG-D decrease the affinity of the protein for linear and bulged DNA, but affect the former to a greater extent [18•]. This supports the conclusion that these residues interact in close proximity to the bulge.

The feature of distorted DNA targeted by HMG proteins is a 'hole' in the DNA structure introduced by the distortion, as exemplified by the cross-over itself in four-way junctions [38,39••,41] and by the wide and deep minor groove created by the bulged DNA bases. HU, the bacterial analogue of HMG1/2, also binds selectively to four-way junctions [42]. The proposed mechanisms of recognition for both classes of protein are very similar, involving the interaction of an exposed hydrophobic residue, phenylalanine in the helix I–II loop of the A domain of HMG1 and proline in the turn joining the two DNA-binding β strands of each HU monomer.

Biological roles for HMG domain proteins

HMG DNA-binding domains both bend DNA substantially and bind with high affinity to distorted DNA structures. What is the biological relevance of these properties? The pleiotropic phenotype of an HMG1 knockout mouse [43•] provides little insight into this question. One postulated role for DNA bending is to promote contact between proteins bound at relatively distant sites on the DNA, thus facilitating the assembly of complex nucleoprotein structures [44]. There is substantial evidence for such a role for the sequence-specific HMG domain proteins. Recently, evidence has been presented for the facilitation of the RAG1/2-mediated V-D-J recombination reaction by mammalian HMG1 [45–47]. HMG1 is required for the formation of the small DNA loop between the sites of recombination and is recruited by the homeodomain of RAG1. This function has an analogy in bacteria, in which HU is required for invertasome formation between the bound Hin invertase and FIS bound to an inversion enhancer 100 base pairs distant.

HMG domains have a higher affinity for distorted DNA structures, such as four-way junctions, than for most other DNA ligands (except microcircles) [12]. What are the potential natural distorted ligands? The HMG domain binds preferentially to the open, rather than the biologically active stacked X, form of four-way junctions and neither DNA bulges nor *cis*-platinated DNA are of frequent occurrence in natural DNA. The gapped DNA structures arising from the action of DNA glycosylases on damaged DNA are very similar to DNA bulges, however, and could be expected to bind HMG domains with high affinity. If so, one potential role for the chromosomal HMG protein could be to protect the lesions introduced by the glycosylases and to perhaps provide a suitable DNA binding surface for the nucleotide excision repair chaperone protein XRCC1 [48].

Conclusions

During the past year, two major questions regarding the interaction of HMG domains with DNA have been largely answered — by how much do the HMG domains of the abundant HMG1 class of chromosomal proteins bend DNA and what are the determinants of the preferential recognition of distorted DNA structures? FRET measurements indicate that the DNA bend angle in complexes of prebent DNA (*cis*-platinated and bulged DNA) with domains structurally homologous to the B domain of HMG1 is close to 100° and is relatively independent of the particular domain studied. By contrast, the bend angle in complexes of *cis*-platinated DNA with the A domain is 61°. The difference in the bend angle induced by these two types of domain correlates with the structural observation that B domains partially intercalate into the DNA at two base steps, whereas the A domain, at least in the published structure, partially intercalates at only a single step. This latter partial intercalation is mediated by residue 34, which has also been invoked as a determinant of the preferential recognition of DNA bulges and may play a role in the recognition of four-way junctions. It is suggested that this residue and its immediate structural environment are a major determinant of the binding of HMG domains to distorted DNA structures. One of the main determinants of sequence-dependent DNA recognition by the HMG-domain transcription factors and the lack of recognition of specific DNA sequences by the HMG domains of the HMG1 class is believed to be residue 10, which is asparagine in the former and serine in the latter. Although the correlation with recognition properties is persuasive, the molecular basis for the discrimination needs to be definitively established (but see Note added in proof).

Note added in proof

Since the submission of this review, the crystal structure, at 2.2 Å resolution, of a complex of the HMG domain of HMG-D with an 11 base pair linear DNA fragment has been published [49••].

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Weir HM, Kraulis PJ, Hill CS, Raine ARC, Laue ED, Thomas JO: **Structure of the HMG box motif in the B-domain of HMG 1.** *EMBO J* 1993, **12**:1311-1319.

2. Read CM, Cary PD, Crane-Robinson C, Driscoll PC, Norman DG: **Solution structure of a DNA-binding domain from HMG1.** *Nucleic Acids Res* 1993, **21**:3427-3436.

3. Jones DNM, Searles MA, Shaw GL, Churchill MEA, Ner SS, Keeler J, Travers AA, Neuhaus D: **The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*.** *Structure* 1994, **2**:609-627.

4. Hardman CH, Broadhurst RH, Raine ARC, Grasser KD, Thomas JO, Laue ED: **Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three and four dimensional NMR spectroscopy.** *Biochemistry* 1995, **34**:16596-16607.

5. 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.

6. Van Houte LPA, Chuprina VP, Van der Wetering M, Boelens R, Kaptein R, Clevers H: **Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4.** *J Biol Chem* 1995, **270**:30516-30524.

7. 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.

8. Bianchi ME, Beltrame M, Paonessa G: **Specific recognition of cruciform DNA by nuclear protein HMG1.** *Science* 1989, **243**:1056-1059.

9. Sheflin LG, Spaulding SW: **High mobility group protein 1 preferentially conserves torsion in negatively supercoiled DNA.** *Biochemistry* 1989, **28**:5658-5664.

10. Pil PM, Lippard SJ: **Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin.** *Science* 1992, **256**:234-237.

11. Sheflin LG, Fucile NW, Spaulding SW: **The specific interactions of HMG1 and 2 with negatively supercoiled DNA are modulated by their acidic C-terminal domains and involve cysteine residues in their HMG1/2 boxes.** *Biochemistry* 1993, **32**:3238-3248.

12. Payet D, Travers A: **The acidic tail of the high mobility group protein HMG-D modulates the structural specificity of DNA binding.** *J Mol Biol* 1997, **266**:66-75.

13. Yen YM, Wong B, Johnson RC: **Determinants of DNA binding and bending by the *Saccharomyces cerevisiae* high mobility group protein NHP6A that are important for its biological activities. Role of the unique N terminus and putative intercalating methionine.** *J Biol Chem* 1998, **273**:4424-4435.

14. •• Ohndorf U-M, Rould MA, He Q, Pabo C, Lippard SJ: **Basis of cis-platin modified DNA by high-mobility-group proteins.** *Nature* 1999, **399**:708-712.
The authors describe the crystal structure of the A domain of HMG1 complexed with *cis*-platinated DNA; this definitive structure resolves some controversial issues.

15. Murphy FV IV, Sehy JV, Dow LK, Gao YG, Churchill MEA: **Co-crystallization and preliminary crystallographic analysis of the high mobility group domain of HMG-D bound to DNA.** *Acta Crystallogr* 1999, **D55**:1594-1597.

16. Balaeff A, Churchill ME, Schulten K: **Structure prediction of a complex between the chromosomal protein HMG-D and DNA.** *Proteins* 1998, **30**:113-135.

17. •• Allain FH, Yen YM, Masse JE, Schultze P, Dieckmann T, Johnson RC, Feigon J: **Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding.** *EMBO J* 1999 **18**:2563-2579.
This NMR study of the NHP6A–DNA complex identified specific intramolecular contacts as a basis for non-sequence-specific binding by the abundant chromosomal HMG domain proteins.

18. • Payet D, Hillisch A, Lowe N, Diekmann S, Travers A: **The recognition of distorted DNA structures by HMG-D – a footprinting and molecular modelling study.** *J Mol Biol* 1999, **294**:79-91.
The authors demonstrated asymmetric binding of HMG-D to a DNA bulge and identified Val34 and Thr35 in the helix I–helix II loop as important determinants of DNA binding and bending.

19. King C-Y, Weiss MA: **The SRY high-mobility-group box recognises DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity.** *Proc Natl Acad Sci USA* 1993, **90**:11990-11994.

20. Baxevanis AD, Landsman D: **The HMG-1 box protein family: classification and functional relationships.** *Nucleic Acids Res* 1995, **23**:1604-1613.

21. Travers AA: **Reading the minor groove.** *Nat Struct Biol* 1995, **2**:615-618.

22. 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.

23. Lnenicek-Allen M, Read CM, Crane-Robinson C: **The DNA bend angle and binding affinity of an HMG box increased by the presence of short terminal arms.** *Nucleic Acids Res* 1996, **24**:1047-1051.

24. Falciola L, Murchie AI, Lilley DMJ, Bianchi M: **Mutational analysis of the DNA binding domain A of chromosomal protein HMG1.** *Nucleic Acids Res* 1994, **22**:285-292.

25. • Lorenz M, Hillisch A, Payet D, Buttinelli M, Travers A, Diekmann S: **DNA bending induced by high mobility group proteins studied by fluorescence resonance energy transfer (FRET).** *Biochemistry* 1999, **38**:12150-12158.
This study demonstrates that the DNA bending angle induced by a variety of HMG B-type domains upon binding bulged DNA is essentially the same. The reported values sharply disagree with the conclusions of Heyduk *et al.* [27].

26. • Jamieson ER, Jacobson MP, Barnes CM, Chow CS, Lippard SJ: **Structural and kinetic studies of a cisplatin-modified DNA icosamer binding to HMG1 domain B.** *J Biol Chem* 1999, **274**:12346-12354.
The authors contrast the bending induced by the B domain of HMG1 upon binding to *cis*-platinated DNA with that revealed by the crystal structure of the HMG1 A domain bound to a similar ligand.

27. Heyduk E, Heyduk T, Claus P, Wisniewski JR: **Conformational changes of DNA induced by binding of *Chironomus* high mobility group protein 1a (cHMG1a). Regions flanking an HMG1 box domain do not influence the bend angle of the DNA.** *J Biol Chem* 1997, **272**:19763-19770.

28. Wisniewski JR, Krohn NM, Heyduk E, Grasser KD, Heyduk T: **HMG1 proteins from evolutionary distant organisms distort B-DNA conformation in similar way.** *Biochim Biophys Acta* 1999, **1447**:25-34.

29. Churchill MEA, Changela A, Dow LK, Krieg, AJ: **Interactions of high mobility group proteins with DNA and chromatin.** *Methods Enzymol* 1999, **304**:99-133.

30. 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.

31. Pontiggia A, Rimini R, Harley VR, Goodfellow PN, Lovell-Badge R, Bianchi ME: **Sex-reversing mutations affect the architecture of SRY-DNA complexes.** *EMBO J* 1994, **13**:6115-6124.

32. Kim Y, Geiger JH, Hahn S, Sigler PB: **Crystal structure of a yeast TBP/TATA-box complex.** *Nature* 1993, **365**:512-520.

33. Kim JL, Nikolov DB, Burley SK: **Co-crystal structure of TBP recognizing the minor groove of a TATA element.** *Nature* 1993, **365**:520-527.

34. Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, Dickerson RE: **How proteins recognize the TATA box.** *J Mol Biol* 1996, **261**:239-254.

35. Farid RS, Bianchi ME, Falciola L, Engelsberg BN, Billings PC: **Differential binding of HMG1, HMG2 and a single HMG box to cisplatin-damaged DNA.** *Toxicol Appl Pharmacol* 1996, **141**:532-539.

36. Peters R, King CY, Ukiyama E, Falsafi S, Donahue PK, Weiss MA: **An SRY mutation causing human sex reversal resolves a general mechanism of structure-specific DNA recognition: application to the four-way junction.** *Biochemistry* 1995, **34**:4569-4576.

37. Wolfe SA, Ferentz AE, Grantchavora V, Churchill MEA, Verdine GL: **Modifying the helical structure of DNA by design: recruitment of an architecture-specific protein to an enforced DNA bend.** *Chem Biol* 1995, **2**:213-221.

38. Pohler JRG, Norman DG, Bramham J, Bianchi ME, Lilley DMJ: **HMG box proteins bind to four-way DNA junctions in their open conformation.** *EMBO J* 1998, **17**:817-826.

39. •• Webb M, Thomas JO: **Structure-specific binding of the two tandem HMG boxes of HMG1 to four-way junction DNA is mediated by the A domain.** *J Mol Biol* 1999, **294**:373-387.
This footprinting study of the binding of intact HMG1 and its isolated A and B domains to four-way junction DNA provides novel insights into the preference of the A domain for this structure.

40. Teo SH, Grasser KD, Hardman CH, Broadhurst RW, Laue ED, Thomas JO: **Two mutations in the HMG-box with very different structural consequences provide insights into the nature of binding to 4-way junction DNA.** *EMBO J* 1995, **14**:3844-3853.

41. Hill DA, Pedulla ML, Reeves R: **Directional binding of HMG-I(Y) on four-way junction DNA and the molecular basis for competitive binding with HMG-1 and histone H1.** *Nucleic Acids Res* 1999, **27**:2135-2144.

42. Kamashev D, Balandina A, Rouviere-Yaniv J: **The binding motif recognized by HU on both nicked and cruciform DNA.** *EMBO J* 1999, **18**:5434-5444.

43. • Calogero S, Grassi F, Aguzzi A, Voigtlander T, Ferrier P, Ferrari S, Bianchi ME: **The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice.** *Nat Genet* 1999, **22**:276-280.
The first report of the physiological effects of deleting HMG1 from the genome.

44. Travers AA, Ner SS, Churchill ME: **DNA chaperones: a solution to a persistence problem?** *Cell* 1994, **77**:167-169.

45. van Gent DC, Hiom K, Paull TT, Gellert M: **Stimulation of V(D)J cleavage by high mobility group proteins.** *EMBO J* 1997, **16**:2665-2670.

46. West RB, Lieber MR: **The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step.** *Mol Cell Biol* 1998, **18**:6408-6415.

47. Aidinis V, Bonaldi T, Beltrame M, Santagata S, Bianchi ME, Spanopoulou E: **The RAG1 homeodomain recruits HMG1 and HMG2 to facilitate recombination signal sequence binding and to enhance the intrinsic DNA-bending activity of RAG1-RAG2.** *Mol Cell Biol* 1999, **19**:6532-6542.

48. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP: **Solution structure of the single-strand break repair protein XRCC1 N-terminal domain.** *Nat Struct Biol* 1999, **6**:884-893.

49. •• Murphy FV IV, Sweet RM, Churchill MEA: **The structure of a chromosomal high-mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition.** *EMBO J* 1999, **18**:6610-6618.
The important structure described demonstrates unambiguously that this chromosomal HMG domain partially intercalates into its DNA ligand at three positions, two major and one minor. The major partial intercalations are separated by an intervening base step and are effected by Met13 at one step and by Val34 and Thr35 at the other. In addition, Ala38 partially intercalates into the intervening step. The DNA in the complex is bent by 111° and has an average twist angle of 27°. Three residues, Ser10, Arg20 and Lys51, form water-mediated hydrogen bonds to DNA bases. The authors point out that this sequence-neutral interaction of Ser10, in concert with the sequence-neutral intercalation of Val34 and Thr35, contrasts with sequence-specific interactions in the LEF-1–DNA complex and could thus contribute substantially to the lack of sequence-specific DNA recognition by the HMG domain of HMG-D. The details of the structure are also in general agreement with the effects of mutations in the HMG domain (e.g. [18•]).