Published in final edited form as:

J Mol Biol. 2010 October 15; 403(1): 88–102. doi:10.1016/j.jmb.2010.08.031.

Structural Analysis of HMGB-DNA Complexes Reveal Influence of Intercalation on Sequence Selectivity and DNA Bending

Mair E.A. Churchill^{1,2,*}, Janet Klass¹, and David L. Zoetewey²

¹ Department of Pharmacology, The University of Colorado Denver, School of Medicine, Aurora, CO 80045, USA

² Molecular Biology Program, The University of Colorado Denver, School of Medicine, Aurora, CO 80045, USA

Abstract

The ubiquitous eukaryotic High-Mobility-Group-Box (HMGB) chromosomal proteins promote many chromatin-mediated cellular activities through their non-sequence-specific binding and bending of DNA. Minor groove DNA binding by the HMG box results in substantial DNA bending toward the major groove owing to electrostatic interactions, shape complementarity and DNA intercalation that occurs at two sites. Here, the structures of the complexes formed with DNA by a partially DNA intercalation-deficient mutant of *Drosophila melanogaster* HMGB have been determined by X-ray crystallography at a resolution of 2.85 Å. The six proteins and fifty base pairs of DNA in the crystal structure revealed a variety of bound conformations. All of the proteins bound in the minor groove, bridging DNA molecules, presumably because these DNA regions are easily deformed. The loss of the primary site of DNA intercalation decreased overall DNA bending and shape complementarity. However, DNA bending at the secondary site of intercalation was retained and most protein-DNA contacts were preserved. The mode of binding resembles the HMGB1-boxA-cisplatin-DNA complex, which also lacks a primary intercalating residue. This study provides new insights into the binding mechanisms used by HMG boxes to recognize varied DNA structures and sequences as well as modulate DNA structure and DNA bending.

Introduction

HMGB proteins are ubiquitous and essential proteins that have an interesting duality of function, where they act either as DNA-binding proteins in the nucleus (reviewed in ¹) or as extracellular mediators of inflammatory responses (reviewed in ^{2; 3}). In their nuclear role, the most abundant HMGB proteins are the non-sequence-specific chromosomal proteins, but numerous transcription factors, such as the ubiquitous Sox family, also use the HMG box (which is the historical name of the DNA binding domain) to recognize DNA site-specifically (reviewed in ^{1; 4}).

*Corresponding author. mair.churchill@ucdenver.edu.

ACCESSION NUMBERS: Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 3NM9.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The non-sequence-specific chromosomal HMGB proteins induce alterations in DNA structure and interact with numerous other proteins to facilitate a variety of cellular processes. They participate in the function of chromatin remodeling by interacting with nucleosomes^{5; 6; 7; 8; 9}, altering the interaction of nucleosomes with DNA^{10; 11; 12}, and aiding chromatin remodeling machinery^{10; 13; 14}. HMGB1/2 facilitates V(D)J recombination by interacting with recombination activating genes (RAG) proteins and enhancing DNA cleavage rates^{15; 16; 17; 18}. The HMGB proteins regulate transcription through interactions with many types of transcription factors^{19; 20; 21; 22; 23; 24} and proteins necessary for efficient transcript elongation²⁵. HMGB family members also have a role in DNA repair^{26; 27; 28}. In many instances these activities require the DNA binding and bending activity of the HMG box and certain amino acid substitutions, that alter DNA bending, unwinding or looping, have detrimental functional consequences *in vivo*^{24; 29; 30; 31; 32; 33; 34; 35}.

The mechanism of DNA bending is a unique feature of the sequence-specific and non-sequence-specific HMGB proteins, which typically bend DNA approximately 60 ° to 110 ° per bound HMG box (summarized in³⁶). The structures of HMGB-DNA complexes, including the sequence-specific as well as non-sequence specific proteins HMGD and NHP6a, which are *Drosophila melanogaster* or *Saccharomyces cerevisiae* homologues of the metazoan chromosomal protein HMGB1, respectively, revealed that the 75 amino-acid long HMG box has a 3-helix L-shaped structure that binds in the flattened, widened, and bent minor groove of DNA^{37; 38; 39; 40; 41}. HMGD is among the most effective at distorting DNA of the HMG boxes studied to date, as it distorts DNA with a bend angle of 111.1° and average twist angle of 27° in structural and solution experiments^{39; 42}.

In addition to shape complementarity and electrostatic interactions, intercalation into the DNA at two sites in the DNA separated by two base pairs is important for wildtype DNA bending and binding of the non-sequence-specific HMGB proteins, as predicted and observed in the structures of DNA bound to HMGD or NHP6A (Figure 1)^{39; 41; 43}. The primary (1°) intercalation site or wedge is composed of an aliphatic residue in helix 1, Met13 in HMGD, which is further stabilized by two buttressing residues, Leu9, and Tyr12 (HMGD numbering; reviewed in⁴⁴) (Figure 1A). The largest bend angle of the DNA binding site (45°) occurs at an individual base-step (T5-A6), which is the site of Met13 intercalation (Figure 1B). Interestingly, the 1° intercalation site is the only intercalation observed for the sequence-specific HMGB proteins^{37; 38}, and the residues in the 2° wedge instead make sequence-specific hydrogen bonds with the DNA, which also distorts DNA in this region of the binding site. The 2° intercalating wedge of the non-sequence specific HMGB proteins at the N-terminus of helix $\alpha 2$, is composed of a single or two adjacent aliphatic or aromatic amino acid residues^{39; 41; 43}. The 2° intercalation site is equally important for linear DNA binding and architecture-specific modes of DNA binding^{24; 29; 30; 31; 45; 46}(Figure 1B and C).

In contrast to the HMG boxes of other non-sequence-specific HMGB proteins, the alanine at the 1° intercalation site of box A of HMGB1 lacks the ability to intercalate DNA. It is not known how this domain may bind and bend linear DNA, as there are no structures of this domain bound to unmodified DNA, although the structure of it bound to cisplatin-modified DNA revealed the importance of the 2° intercalation site in recognition of the kinked DNA site⁴⁷. Interestingly, systematic examination of DNA bending of a series of alanine amino acid substitutions of HMGD also revealed that substitution of the 1° intercalating residue to alanine (M13A) decreased the binding affinity to linear DNA or pre-bent (disulfide crosslinked DNA) by 6-fold and 9-fold respectively. However, unexpectedly only a slight decrease in DNA bending was observed in a ligase mediated circularization study³⁶.

Despite the importance of intercalating residues, the structural basis of their influence on HMGB function has been limited because only a few NMR or X-ray crystallography structures of DNA-bound complexes of variants of these proteins have been determined. Therefore, to better understand the role of 1° intercalation, we determined the co-crystal structure of the HMGD^{M13A} bound to a DNA decamer. The structure, which provides six different views of the protein-DNA complex, reveals how DNA binding and bending are accomplished in a DNA-bending defective HMGB protein and new insights into the nature of non-sequence-specific DNA recognition by HMGB proteins.

Results and Discussion

Crystallization, structure determination, and overall structure

In order to understand how deficiencies in intercalation might alter HMGB-DNA recognition, we solved the structure of the HMGD^{M13A} protein co-crystallized with DNA. The substitution of the 1° intercalating methionine with alanine in HMGD was previously shown to alter its DNA bending and binding properties³⁶. One crystal that was suitable for x-ray diffraction studies of HMGD^{M13A} with the duplex DNA fragment (dGCGATATCGC) formed in a triclinic unit cell after many co-crystallization attempts with a variety of different DNA oligonucleotides. A crystal of this size and quality could not be reproduced, despite seeding and other approaches, which precluded determination of experimental phases for the structure. The structure was solved by molecular replacement using the Molrep program in the CCP4 suite of programs with the wild type (WT) HMGD protein as a model^{39; 48}. After locating one protein molecule in the unit cell, use of a locked rotation/translation function was essential for finding the additional proteins. Molecular averaging was required for density improvement so that DNA could be visualized in the electron density maps^{49; 50}. The final model contains six protein molecules (chains A, D, G, J, M, and P), 5 DNA decamers (chains B/C, E/F, H/I, K/L, and N/O), and four water molecules. Because of apparent static disorder in the crystal, one of the DNA strands (strand L) was modeled with a guanosine added at position one making strand L 11 bp, since modeling the strand in alternate conformations did not improve the model. The final model was refined to a resolution of 2.85 Å, with good geometry and stereochemistry and acceptable R-values (29.5/24.5; Table 1).

The five DNA decamers stacked end-to-end throughout the crystal, with curved as well as relatively straight regions of DNA (Figure 2A). The electron density defined the protein and DNA very well (Figure 2B), except for a few regions of DNA that lacked bound protein. For these regions, the DNA was quite disordered as indicated by high B-factors, but the electron density was still interpretable.

The six HMGD^{M13A} proteins formed a variety of interactions with the DNA and each other in the crystal (Figure 3). An interesting feature of the protein-DNA packing was the interaction between pairs of HMG boxes that occurred along the DNA and across adjacent duplexes (Figure 2A). Two sets of HMG boxes, A/M, and D/J, each have a head-to-head arrangement on the DNA (Figure 2A and in detail in Figure 4A and B), whereas molecules G and P bind to the DNA in a head-to-tail fashion (Figure 2A). The length of the DNA in each binding site is 8 base pairs, except for the binding sites of domains G and P that are nine base pairs, which happen to be those that bind DNA in a head-to-tail configuration. There are no direct interactions between any two proteins bound to adjacent sites along the DNA. However, extensive protein-protein packing occurs between alternate proteins, and this may facilitate the formation of this complex structure. For two sets of protein pairs, D/P and G/J, the amino acid residues 44–65 of helix $\alpha 3$ interact with the same region of the other protein, which is in an antiparallel orientation, giving rise to a homotypic contact. Additionally, the contacts between proteins in the different layers in the crystal show two

types of heterotypic interactions. For A, D, and G and their symmetry mates (A-A_{symm}, D-D_{symm}, and G-G_{symm}), residues 41–44 of helix $\alpha 2$ contact the “top end” of the HMG box at residues 2, 73, and 74. The other heterotypic interaction occurs between molecules A-J, D-M and G-P, where the region containing helix $\alpha 1$ and the helix $\alpha 1$ - $\alpha 2$ loop (aa 25–29) contact residues of the N-terminal strand (aa 6–14) and helix $\alpha 3$ 55–58 of the other protein in the pair. This packing interaction is similar to that previously observed for the WT HMGB protein-DNA complex³⁹, and suggests that particular protein-protein interactions may be relevant to the formation of higher-order protein-DNA complexes. Unusual effects of HMGB proteins on DNA structure have been observed in single molecule stretching studies, where low HMGB concentrations fit to a flexible hinge mode, but relatively high protein concentrations of HMGB fit a cooperative filament model⁵¹. We speculate that protein-protein interactions such as the ones observed in this structure may contribute to this unusual mode of DNA binding.

HMGB^{M13A} binds DNA across DNA junctions

The HMG boxes bind to the DNA minor groove in a wide variety of configurations and surprisingly all were seen to bind across duplex DNA junctions that were formed by adjacent DNA molecules in the unit cell or adjacent unit cells. The DNA ladder diagram for each of the HMGB proteins, as well as the top and side views of the protein-DNA complexes in Figure 3, illustrate the positions of the DNA within each of the protein-DNA binding sites. For molecules A and G, the junction is near the middle of the protein binding site that we have arbitrarily defined as the position of Leu9, but for molecule J, the junction is near the N-terminus of the HMG box, and for M and P it is positioned near the opposite end of the HMG box, whereas molecule D was the most centered on the DNA of all the complexes. This unique arrangement of the HMGB^{M13A} on the DNA differed dramatically from the central location of the wildtype HMGB on the same DNA sequence³⁹, but is consistent with HMGB intercalation at a junction between duplex and non-duplex DNA⁵². A central arrangement of HMG boxes on their binding sites has been observed in other structures of HMGB-DNA complexes^{37; 39; 40; 41; 46}. Likely explanations for the unusual positioning observed across DNA junctions are that numerous protein-protein interactions are favorable for crystal packing and the more weakly binding and bending HMGB^{M13A} mutant may interact near the DNA ends because these regions are more easily deformed than the center of the DNA. Other similar HMG boxes, such as the A box of HMGB, also preferentially bind easily distorted and/or pre-bent DNA, and indeed there is a preference of the HMGB-box A for these types of DNA sites relative to box B, which generally prefers unmodified linear DNA⁵³.

Conservation of DNA contacts in the HMGB^{M13A} mutant-DNA complexes

A subset of the protein-DNA interactions are conserved despite the unique DNA binding environment and differences in the details of the protein-DNA interactions observed in the structure of HMGB^{M13A} bound to DNA (Figure 3). In every single complex, Arg7, Leu9, Ser10, Tyr12, Arg20, Val32, and Thr33, and in most complexes, Asn17, Trp43, Arg44, and Ala36 interact with the DNA via van der Waals interactions, short-range ion pairs, or through hydrogen bonding. Ser10, Asn17, Tyr12, and Trp43, all make direct hydrogen bonds with the DNA, but in the case of Ser10 these are water mediated (Figure 4C and D).

The details of the protein-DNA interactions are quite different between the six models, but several patterns emerged. Lys6, Arg7, Arg20, Lys37, Arg44 and Lys60 are basic residues that either directly interact with the DNA or would be expected to contribute to electrostatic interactions with the DNA. These interactions occur primarily through the phosphodiester backbone, and thus may steer the protein to its DNA binding site. In addition, Arg7 binds within the minor groove, in some cases directly to DNA bases, and likely contributes to the

substantial narrowing of the minor groove that is observed in several of the complexes, including the A/M and D/J proteins that have a head-to-head arrangement on the DNA (Figure 4A and B and Table II). This head-to-head configuration of proteins bound to DNA and the extensive contacts of Arg7 are remarkably like those observed for the designed hybrid (SRY-HMGB1boxB) di-domain-DNA complex⁴⁶. In contrast to the four bp spacing between the equivalent arginine contacts in the hybrid di-domain structure (Arg7 C α -Arg 91 C α of 28 Å), the A/M and D/J pairs of HMG boxes are separated by only two and zero base pairs (Arg7 C α -C α of 18 and 16 Å), respectively. Although there is no linker connecting the A/M or D/J HMG boxes, the distance spanning the C-terminus of A (or D) to the N-terminus of M (or J) is short enough to be bridged easily by any of the linker lengths observed in the dual-domain HMGB proteins^{1; 44; 54}. Together these data support an important role for the arginine at position seven in binding within a substantially narrowed minor groove potentially stabilizing the head-to head configuration of HMG boxes on adjacent binding sites.

The HMGD^{M13A} mutant maintains the ability to intercalate DNA at the 2° site

Close examination of the 2° intercalation wedge formed by Val32 and Thr33 (Figure 3) showed direct interactions of these residues with the DNA. For DNA intercalation to occur, the “intercalating” amino acid residue(s) must protrude into the base stack, which requires a distortion of the DNA, such as increased inter-base pair roll. As expected, Ala13 did not intercalate the DNA at all. However, true intercalation by the 2° intercalation-site residues Val32 and Thr33 occurred in a majority of the complexes (Figure 4E and Table II). The A, D, G, and J complexes, but not the M and P complexes, showed intercalation presumably because the position of Val32 and Thr33 occurs at the last base step of the helix in molecules M, and P, very near to the DNA junction (Figure 3 and Suppl. Table I). The intercalation sites have a characteristically widened minor groove of between 9.6 to 11.2 Å (Table II) and positive base pair roll angles of between 10.2 ° and 45.3 ° (Table II and Suppl. Table I). These values compared well to the values measured at the 2° intercalation site in the wildtype HMGD-DNA complex for the roll angle (32 °) and minor groove width (10.1 Å)³⁹. Therefore, the substitution of the methionine with alanine at the 1° intercalation site had little apparent influence on the magnitude of deformation in DNA structure at the 2° intercalation site, which is only two base steps away.

The intercalation mutant alters the contacts within the HMGD-DNA complex

The most notable differences in protein-DNA contacts between the wildtype HMGD and the HMGD^{M13A} mutant occurred in the vicinity of the substituted methionine residue at the 1° intercalation position. Analysis of the contacting residues using CNS found that the conserved residue interactions, as described earlier, generally occur in both the wildtype and mutant complexes with some exceptions. As expected, the Ala13 residue had fewer van der Waals interactions with the DNA than those seen with the minor groove intercalation observed for Met13. However, Leu16, which makes tight van der Waals contacts with the DNA in the wildtype complex, contacted the DNA only in the G mutant complex. Overall, the region surrounding the intercalation residue had fewer van der Waals contacts with the DNA in the HMGD^{M13A} mutant than the wildtype HMGD, which suggests that the intercalation itself serves to bring the HMG box in closer contact with the DNA minor groove.

To determine whether these differences in contact residues altered the packing of the protein DNA complexes, the interfacial surface areas were analyzed. The buried surface areas calculated using CNS for the HMGD^{M13A} complexes are between 659 and 934 Å² (Table II). With the exception of molecules D and G, which had the largest buried surface areas, the other complexes had substantially less buried surface area than the wildtype HMGD (821 Å²

in Table II). The D and G proteins have the most extensive intercalation by the 2° intercalation wedge (Figure 3 and 4E). This measurement of buried surface area corresponds to the solvent excluded surface and does not give as clear an indication of the tightness of the protein-DNA complex as does the calculation of packing density. Therefore, the program Fast Atomic Density Evaluation (FADE) was used to analyze the interfacial packing density for each complex (Table II). In this case both the area of the interfacial surface and the magnitude of the packing density score were significantly greater for the wildtype HMGD-DNA complex than for any of the complexes formed with the HMGD^{M13A} mutant. Interestingly, the packing density we calculated for HMGB-box A bound to cisplatin modified DNA⁴⁷ was also quite low and similar in magnitude to the HMGD^{M13A} mutant. Taken together, the DNA contact, buried surface area, and packing density calculations clearly demonstrate that the loss of the 1° intercalating residue has a large impact on the ability of the HMG box to form a tight complementary surface interaction with the linear DNA.

The HMGD^{M13A} mutant is defective in DNA bending

Since the M13A mutant had previously been shown to be slightly deficient at bending DNA in previous solution studies³⁶, we examined the DNA structure and effects on DNA bending in the structure. Because the binding sites occupy more than one DNA duplex, it was not possible to obtain direct bending angles for each HMGD^{M13A} binding site. However, the global axis of DNA curvature for each complex was measured using the program CURVES⁵⁵. A qualitative view of the bend in the DNA that was produced by HMGD^{M13A} compared to the wildtype HMGD (Figure 5A) reveals that two of the models (D and J) show a distinct bend in the helix axis, but the other models either did not show significant bends in the helix axis or were not amenable to the calculation, because of the position of the junction between the duplex DNA fragments. However, the position and magnitude of bending of the helix axis for D and J was considerably less than that observed for the wildtype HMGD. The differences in DNA bending that were induced by the HMGD^{M13A} proteins might be related to the different sequence and junction environments. Interestingly, HMG boxes do have the ability to adopt a fairly wide range of bend angles of 78 ° with a standard deviation of 23 ° when binding to mixed sequence DNA^{42; 56; 57}.

The local DNA structure induced by the binding of the proteins showed conserved features. The program CURVES was used to measure DNA structural parameters as well as local bending angles for each complex. The average helical twist, base pair roll and propeller twist values were indicative of quite highly distorted B-DNA structures (Table III). The overall local and global curvature measurements also indicate that the mutant bends the DNA less than the wildtype HMGD. The largest global curvature noted was for molecule D of 43.2 °, which is half of the overall bending observed for HMGD (99.4 °) (Table III). As noted earlier, the local bending parameters and the bending angles at the 2° site were similar to the wildtype HMGD, which indicates that the 1° site is needed for the majority of the DNA distortion or bent-DNA stabilization that leads to overall DNA bending by the HMG box.

The HMGD^{M13A} mutant protein structure is altered

The HMG boxes in the HMGD^{M13A} structure are generally similar to previously published HMGD and numerous HMGB-type proteins. However, the variability between the multiple models gives the opportunity to examine the most structurally malleable regions of the domain in complex with DNA. The superposition of the proteins in the structures gave an average r.m.s.d. for backbone atoms of 0.7 Å (Table IV). The most closely related proteins are A, D, and G, which are close to the expected coordinate error for this structure 0.6 Å. Compared to the free WT HMGD (PDB ID 1hma), the average r.m.s.d. was 1.68 Å⁵⁸,

whereas the r.m.s.d compared to the WT HMGD protein bound to DNA (PDB ID 1qrv) averages 1.02 Å. Compared to the HMGB1-box A bound to cisplatin DNA, the r.m.s.d. average was much closer (1.04 Å) than the wildtype HMGD protein, which was 2.1 Å^{39; 47}. Therefore, the HMG box in the mutant structure adopts a slightly different conformation than the WT HMGD bound to DNA in a different context and this conformation is closer to box A of HMGB1, which also lacks an intercalating residue at the 1° intercalation site⁴⁷.

The HMG boxes of the HMGD and HMGD^{M13A} differ in a region localized to 1° intercalation site involved in DNA bending and binding. In Figure 5, the models of all of the six HMGD^{M13A} mutant structures were superimposed with the WT HMGD protein and the complexes with the DNA were analyzed. The N-terminal strand and helix α 1 were shifted relative to the rest of the protein (Figure 5B), which brings all of the residues in the 1° intercalating wedge into slightly different positions relative to their positions in the WT protein. These structural differences are a consequence either of the mutation or DNA binding because the WT HMGD was used as the molecular replacement model. Specifically, Leu16 exists in a different rotamer than in the WT protein, Asn17 is oriented differently, Leu9 moved away from alpha helix 3 toward the DNA, and the phi/psi angles of position 13 have changed slightly. In wildtype HMGD, Met13 has van der Waals interactions with Tyr12 and Leu9 and similar interactions have been observed in the free or bound proteins of several species of HMGB proteins^{39; 41; 59; 60; 61}. The M13A mutant is less thermally stable than the wildtype protein, presumably because it lacks some of these stabilizing interactions³⁶.

The M13A mutant also results in differences in the disposition of the HMG box in the minor groove. Figures 5C and D show that the N-terminal strand of the HMG box lies closer to the DNA in the M13A mutant than the wildtype HMGD, even though in the wildtype HMGD the 1° intercalation site itself is closer to the minor groove bases. This corresponds with a much narrower and deeper minor groove in this region of the HMGD^{M13A} models than was seen for the wildtype HMGD complex (see R7 groove widths and depths in Table II). Furthermore, additional contacts with the DNA by Lys4 (Figure 3 and 4A and 4B) occur in several of the HMGD^{M13A}-DNA complexes (Figure 3 and 4A and 4B). The slightly narrowed minor groove in the region exists in the structure of box A of HMGB1 bound to the cisplatin modified DNA⁴⁷ and suggests that the 1° intercalation does more than just intercalate the DNA, but also permits the domain to settle more deeply into the minor groove and generate a more sustained bend along a longer stretch of the DNA in the binding site. The similarities between the HMGD^{M13A} complex with unmodified DNA and box A of HMGB1 bound to the cisplatin-modified DNA⁴⁷ in the shape of the minor groove and dominance of the 2° interacting residue are quite striking. These results are consistent with the idea that the type of residue at 1° intercalating position defines whether the HMG domain adopts structure-specific (preferential binding to pre-bent or deformed DNA) binding or DNA-bending (binds to and bends linear DNA) mode of interaction with DNA^{53; 62; 63}.

Non-sequence-specific DNA recognition

The characteristic ability of the non-sequence-specific HMGB proteins to bind to different DNA sequences was visualized for the first time in this structure. Ser10, which is the position of a specificity determinant in the sequence-specific HMGB proteins interacts with the DNA similarly in the wildtype and mutant structures. Ser10 forms hydrogen bonds to a GuaN3 position in the DNA at E9 via a water molecule in molecule D and in molecule P, a water mediates its interaction with the Ade4 N3 position of strand K (Figure 4C and D). In the other complexes Ser10 is also adjacent to purines, but ordered water molecules were not detected in the electron density. In all cases Ser10 is near a purine, except for molecule J, where it interacts with the sugar-phosphate backbone of the DNA (Figure 3).

The interaction maps for the protein-DNA contacts (Figure 3) illustrate that there was no particular sequence register for any of the HMGD^{M13A} proteins on the DNA other than the selection of a pyrimidine-purine base step at the site of the 2° intercalation. Molecules A and G or D and J intercalate the DNA at a CG or TA base steps, respectively. In the wildtype HMGD-DNA crystal structure the 1° intercalation occurred at a TA base step and the 2° occurred at the adjacent purine-purine base step. The NMR structure of NHP6A bound to DNA has the 1° intercalation at a TG base step and the 2° intercalation at a TT base step 2 base steps away⁴¹. Whether these differences are due to the intrinsic intercalation preferences of these two proteins with the particular DNA sequences used in the studies or differences due to the structure determination methods is not known. In the absence of the 1° intercalating residue the 2° intercalation dominates and there appears to be a preference for this intercalation at a pyrimidine-purine base step. Therefore, although HMGD^{M13A} has the ability to recognize different DNA sequences and structures, it has sequence preferences for intercalation. Although the HMGB proteins are non-sequence-specific, HMGD does exhibit some sequence selectivity for AT-rich sequences and sequences containing pyrimidine-purine base steps such as TA or TG⁶⁴.

Conclusions and implications for HMGB protein-DNA recognition

Many studies have suggested that HMG boxes induce bends into DNA with quite nonuniform sequence-dependent bend angles^{31; 42; 56; 57}, and this structure provides mechanistic insight into this phenomenon as well as how HMG boxes may bind across duplex junctions and recognize different DNA sequences. Despite the differences in the details structure of the mutant HMGD-DNA complexes, the HMG box maintained the majority of the direct amino acid residues-DNA contacts. However, loss of the 1° intercalation site in the HMGD^{M13A} mutant uncovered a pyrimidine-purine sequence selectivity of the 2° intercalation that was not observed for the wildtype HMGD, which suggest that binding stabilization due to the 1° interacting residue is dominant over the sequence-selectivity of the 2° intercalation site. The structures also highlighted arginine at position 7 in the binding in a narrow DNA minor groove, perhaps contributing to the head-to-head configuration of HMGB1/2 with HMG box binding sites having different interdomain spacing⁴⁶. Finally, we suggest that the 1° intercalation residue is important for close packing of the HMG box with the DNA and a specific conformation of the protein and DNA minor groove. Indeed, widened minor grooves and close DNA contacts in the 1° site region were seen in all of the structures of the sequence-specific HMGB protein-DNA complexes^{37; 40}, as well as NHP6A and the designed di-domain (SRY-HMGB1 hybrid) structures that have 1° intercalation^{39; 41; 46}. In the absence of the 1° intercalation, the HMGD^{M13A} domain bound to DNA resembles the HMGB1 box A-bound to cisplatin modified DNA⁴⁷ in the vicinity of the non-intercalating alanine residue and has a similar looser fit in the DNA binding site. Thus, the HMG box has a remarkable ability adapt to a wide variety of DNA sequences and structures that are encountered in chromatin and the 1° and 2° intercalation residues play a crucial role in being able to recognize and bind to them for correct function *in vivo*.

Materials and methods

Production and purification of protein and DNA

The DNA encoding the gene for HMGD^{M13A} was truncated to the size of the minimal HMG box of 74 residues. Standard PCR and subcloning protocols used the pET13a plasmid containing the longer HMGD (100 aa) Met13 substituted to alanine gene inserted between *Bam*HI and *Nde* I sites as the template³⁶ and synthetic oligonucleotide primers (Qiagen) 5'-GAGATATACATATGTCTGATAAGCCAAAACGCCCACTCTCCGCCTACAGTCTGTGGCTC-3' and 5'-AAGGATCCCTATTACTTCTTGCTCTTCTT-3', with the *Nde*I and

BamHI sites, respectively (underlined). The pET-D74-M13A sequence was verified by DNA sequencing.

The HMGD^{M13A} expression and purification protocol was similar to HMGD protocols that have been described previously^{36; 65}. Briefly, the pET-D74-M13A plasmid was transformed into BL-21(DE3) cells, and the protein was overexpressed by induction with IPTG followed by expression for four hours at 37°C. The purification protocol used high salt lysis, isocratic DEAE Sephacel chromatography (Pharmacia) and SP-Sephacel cation exchange chromatography (Pharmacia) followed by isocratic reversed phase-HPLC and SP-Sephacel cation exchange chromatography to remove protein modifications and contaminants⁶⁵. The protein was concentrated and exchanged into water using a stirred cell concentrator (Amicon). The protein mass was confirmed by electrospray or MALDI mass spectrometry, and UV spectroscopy was used to determine the protein concentration using an extinction coefficient of 19,100 (M cm)⁻¹. DNA oligonucleotides (Qiagen) of the sequence GCGATATCGC were purified using size-exclusion chromatography and concentrated in water, as described previously⁶⁶.

Co-crystallization and data collection

Co-crystallization was conducted using freshly prepared protein and DNA that were combined into a drop of final concentration 1.2 mM HMGD^{M13A} protein, 1.18 mM duplex DNA fragment (GCGATATCGC), 5 mM MES-Na pH 5.25, 10 mM NaCl, and 7.6 % PEG 3350 equilibrated against 0.5 mL 32% PEG 3350 in hanging drop vapor diffusion trays. After 17 h the drops were streak seeded from smaller crystals from nearly identical conditions, and allowed to sit undisturbed in a warm (~24 °C) isolated area for three weeks. Needle-shaped crystals formed sporadically, and only one reached the size used here, 200 µm × 30 µm × 30 µm. The crystal was flash frozen prior to data collection at the UC Denver Biomolecular X-ray Crystallography Center using the Rigaku-MSR R-AxisIV++ area detector with confocal optics and an XTREAM cryocooling system. Images of 0.5° were collected for 40 min each over a range of 360°, and the data were processed using DENZO and SCALEPACK from the HKL suite of programs⁶⁷.

Structure solution

The structure was solved using molecular replacement methods. The Molrep program from the CCP4 program suite⁴⁸ was used with an HMGD-74 monomer (PDB ID 1qrv) as a search model to find an initial rotational solution. The position of the molecule with the best rotation solution molecule was fixed at the origin of the triclinic cell, and successive protein molecules were placed relative to the original protein in an iterative process of the locked rotation/translation function in Molrep until six proteins were located in the unit cell. Crystallography and NMR system (CNS)⁶⁸ rigid body refinement with flexible non-crystallographic symmetry, solvent flattening and anisotropic/fixed isotropic restrictions was used with medium resolution data, 12–3.5 Å, and energy minimization with the same restrictions was used to refine the initial positions of the proteins. Density modification, using a mask around the proteins that was created in CNS and refined using MAMA⁵⁰, was required to observe electron density for the DNA. As initial DNA was built more electron density appeared and building of DNA and rebuilding of previously truncated portions of the proteins were added iteratively to the model. Calculation of model phases, followed by Hendrickson-Lattman coefficient blurring and use of Solomon in the solvent flipping mode clarified even more density, and eventually the final model contained 6 proteins and 5 DNA decamers⁶⁸. Refinement improvements were made using REFMAC with restrained molecular averaging of the proteins, prior to final rounds of refinement without non-crystallographic symmetry restraints⁴⁸. There is one large solvent region in the model that

appears to be large enough to contain at least one more protein, but the electron density was not interpretable in this region (Figure 2).

Structural analysis

The structure was analyzed for stereochemical and geometrical quality using PROCHECK⁶⁹. Structural comparisons among HMG boxes were conducted using LSQMAN⁴⁹. DNA structure analyses were conducted using CURVES⁵⁵, and 3DNA⁷⁰ as indicated in the figures and tables. The contacts were identified using CNS⁶⁸, and interfacial packing densities were calculated using FADE⁷¹. Figures were generated using MOLSCRIPT⁷², PyMol⁷³, Raster3D⁷⁴, VMD⁷⁵ and Photoshop (Adobe).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We appreciate the helpful suggestions and contributions of Ryan Fletcher, Frank Murphy, Sarah Roemer, Rui Zhao, and other members of the Churchill lab. We acknowledge the University of Colorado Biomolecular X-ray crystallography Center, which is supported by the HHMI and University of Colorado Comprehensive Cancer Center. This work was supported by NIH grants R01GM59456 and R01GM079154 to MEAC.

References

1. Stros M. HMGB proteins: interactions with DNA and chromatin. *Biochim Biophys Acta*. 2010; 1799:101–13. [PubMed: 20123072]
2. Lotze MT, Tracey KJ. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. *Nat Rev Immunol*. 2005; 5:331–42. [PubMed: 15803152]
3. Bianchi ME, Manfredi AA. High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol Rev*. 2007; 220:35–46. [PubMed: 17979838]
4. Thomas JO, Travers AA. HMGI and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci*. 2001; 26:167–74. [PubMed: 11246022]
5. Stros M, Kolibalova A. Interaction of non-histone proteins HMGI and HMGI2 with core histones in nucleosomes and core particles revealed by chemical cross-linking. *Eur J Biochem*. 1987; 162:111–8. [PubMed: 3816775]
6. Jackson JB, Pollock JM Jr, Rill RL. Chromatin fractionation procedure that yields nucleosomes containing near-stoichiometric amounts of high mobility group nonhistone chromosomal proteins. *Biochemistry*. 1979; 18:3739–48. [PubMed: 476083]
7. Goodwin GH, Mathew CG, Wright CA, Venkov CD, Johns EW. Analysis of the high mobility group proteins associated with salt-soluble nucleosomes. *Nucleic Acids Res*. 1979; 7:1815–35. [PubMed: 537911]
8. Ura K, Nightingale K, Wolffe A. Differential association of HMGI and linker histones B4 and H1 with dinucleosomal DNA: structural transitions and transcriptional repression. *The EMBO Journal*. 1996; 15(18):4959–4969. [PubMed: 8890169]
9. Churchill MEA, Changela A, Dow LK, Krieg AJ. Interactions of high mobility group box proteins with DNA and chromatin. *Methods Enzymol*. 1999; 304:99–133. [PubMed: 10372358]
10. Bonaldi T, Langst G, Strohner R, Becker PB, Bianchi ME. The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. *Embo J*. 2002; 21:6865–73. [PubMed: 12486007]
11. Nightingale K, Dimitrov S, Reeves R, Wolffe A. Evidence for a shared structural role for HMGI and linker histones B4 and H1 in organizing chromatin. *EMBO J*. 1996; 15:548–561. [PubMed: 8599938]
12. Bustin M, Reeves R. High-Mobility-Group chromosomal proteins: Architectural components that facilitate chromatin function. *Prog Nuc Acid Res Mol Biol*. 1996; 54:35–100.

13. Hartlepp KF, Fernandez-Tornero C, Eberharter A, Grune T, Muller CW, Becker PB. The histone fold subunits of *Drosophila* CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol Cell Biol.* 2005; 25:9886–96. [PubMed: 16260604]
14. Formosa T, Eriksson P, Wittmeyer J, Ginn J, Yu Y, Stillman DJ. Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *Embo J.* 2001; 20:3506–17. [PubMed: 11432837]
15. Dai Y, Wong B, Yen YM, Oettinger MA, Kwon J, Johnson RC. Determinants of HMGB proteins required to promote RAG1/2-recombination signal sequence complex assembly and catalysis during V(D)J recombination. *Mol Cell Biol.* 2005; 25:4413–25. [PubMed: 15899848]
16. Kriatchko AN, Bergeron S, Swanson PC. HMG-box domain stimulation of RAG1/2 cleavage activity is metal ion dependent. *BMC Mol Biol.* 2008; 9:32. [PubMed: 18380906]
17. Aidinis V, Bonaldi T, Beltrame M, Santagata S, Bianchi ME, Soanopolou 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. [PubMed: 10490593]
18. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* 1997; 16:1665–2670.
19. Zwilling S, Konig H, Wirth T. High-mobility-group protein-2 functionally interacts with the POU domains of Octamer transcription factors. *EMBO J.* 1995; 14:1198–1208. [PubMed: 7720710]
20. Boonyaratankornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME, Taraseviciene L, Nordeen SK, Allegretto EA, Edwards DP. High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. *Mol Cell Biol.* 1998; 18:4471–4487. [PubMed: 9671457]
21. Ge H, Roeder RG. The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J Biol Chem.* 1994; 268:17136–17140. [PubMed: 8006019]
22. Jayaraman L, Moorthy NC, Murthy KGK, Manley JL, Bustin M, Prives C. High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev.* 1998; 12:462–472. [PubMed: 9472015]
23. Zappavigna V, Falciola L, Citterich M, Mavilio F, Bianchi M. HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. *EMBO J.* 1996; 15:4981–4991. [PubMed: 8890171]
24. Roemer SC, Adelman J, Churchill ME, Edwards DP. Mechanism of high-mobility group protein B enhancement of progesterone receptor sequence-specific DNA binding. *Nucleic Acids Res.* 2008; 36:3655–66. [PubMed: 18474528]
25. Guermah M, Palhan VB, Tackett AJ, Chait BT, Roeder RG. Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. *Cell.* 2006; 125:275–86. [PubMed: 16630816]
26. Yuan F, Gu L, Guo S, Wang C, Li GM. Evidence for involvement of HMGB1 protein in human DNA mismatch repair. *J Biol Chem.* 2004; 279:20935–40. Epub 2004 Mar 09. [PubMed: 15014079]
27. Giavara S, Kosmidou E, Hande MP, Bianchi ME, Morgan A, d'Adda di Fagagna F, Jackson SP. Yeast Nhp6A/B and mammalian Hmgb1 facilitate the maintenance of genome stability. *Curr Biol.* 2005; 15:68–72. [PubMed: 15649368]
28. Lange SS, Mitchell DL, Vasquez KM. High mobility group protein B1 enhances DNA repair and chromatin modification after DNA damage. *Proc Natl Acad Sci U S A.* 2008; 105:10320–5. [PubMed: 18650382]
29. Pontiggia A, Rimini R, Harley VR, Goodfellow PN, Lovell-Badge R, Bianchi ME. Sex-reversing mutations affect the architecture-specificity of SRY-DNA complexes. *EMBO J.* 1994; 13:6115–6124. [PubMed: 7813448]
30. Agresti A, Scaffidi P, Riva A, Caiolfa VR, Bianchi ME. GR and HMGB1 interact only within chromatin and influence each other's residence time. *Mol Cell.* 2005; 18:109–21. [PubMed: 15808513]

31. Sebastian NT, Bystry EM, Becker NA, Maher LJ 3rd. Enhancement of DNA flexibility in vitro and in vivo by HMGB box A proteins carrying box B residues. *Biochemistry*. 2009; 48:2125–34. [PubMed: 19236006]
32. Haykinson MJ, Johnson RC. DNA looping and the helical repeat in vitro and in vivo: Effect of HU protein and enhancer location on Hin invertasome assembly. *EMBO J*. 1993; 12:2503–2512. [PubMed: 8508775]
33. 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. [PubMed: 9468494]
34. Ross ED, Hardwidge PR, Maher LJ 3rd. HMG proteins and DNA flexibility in transcription activation. *Mol Cell Biol*. 2001; 21:6598–605. [PubMed: 11533247]
35. Becker NA, Kahn JD, Maher LJ 3rd. Eukaryotic HMGB proteins as replacements for HU in *E. coli* repression loop formation. *Nucleic Acids Res*. 2008; 36:4009–21. [PubMed: 18515834]
36. Klass J, Murphy FVI, Fouts S, Serenil M, Changela A, Siple J, Churchill MEA. The Role of Intercalating Residues in Chromosomal High-Mobility-Group Protein DNA Binding, Bending and Specificity. *Nucleic Acids Research*. 2003; 31:2852–2864. [PubMed: 12771212]
37. Love JJ, Li X, 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. [PubMed: 7651541]
38. 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. [PubMed: 7774012]
39. 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. [PubMed: 10581235]
40. Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM. Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol*. 2001; 312:481–499. [PubMed: 11563911]
41. Masse JE, Wong B, Yen YM, Allain FH, Johnson RC, Feigon J. The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. *J Mol Biol*. 2002; 323:263–84. [PubMed: 12381320]
42. Dragan A, Churchill MEA, Klass J, Crane-Robinson C, Read C, Privalov P. DNA Binding of a Non-Sequence Specific HMG-D Protein is Entropy Driven with a Substantial Non-Electrostatic Contribution. *J Mol Biol*. 2003; 331:795–813. [PubMed: 12909011]
43. Balaeff A, Churchill MEA, Schulten K. Structure prediction of a complex between the chromosomal protein HMG-D and DNA. *Proteins*. 1998; 30:113–135. [PubMed: 9489920]
44. Murphy FV, Churchill MEA. Nonsequence-specific DNA recognition: a structural perspective. *Structure, Folding, and Design*. 2000; 8:R83–89.
45. Saito K, Kikuchi T, Yoshida M. The mechanism of sequence non-specific DNA binding of HMG1/2-box B in HMG1 with DNA. *Protein Eng*. 1999; 12:235–42. [PubMed: 10235624]
46. Stott K, Tang GS, Lee KB, Thomas JO. Structure of a complex of tandem HMG boxes and DNA. *J Mol Biol*. 2006; 360:90–104. [PubMed: 16813837]
47. Ohndorf UM, Rould MA, He Q, Pabo CO, Lippard SJ. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature*. 1999; 399:708–712. [PubMed: 10385126]
48. Bailey S. The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallogr*. 1994; D50:760–763.
49. Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr*. 1996; 52:842–57. [PubMed: 15299650]
50. Kleywegt GJ, Jones TA. Software for handling macromolecular envelopes. *Acta Crystallogr D Biol Crystallogr*. 1999; 55:941–4. [PubMed: 10089342]
51. McCauley M, Hardwidge PR, Maher LJ 3rd, Williams MC. Dual binding modes for an HMG domain from human HMGB2 on DNA. *Biophys J*. 2005; 89:353–64. [PubMed: 15833996]
52. Cerdan R, Payet D, Yang JC, Travers AA, Neuhaus D. HMG-D complexed to a bulge DNA: an NMR model. *Protein Sci*. 2001; 10:504–18. [PubMed: 11344319]

53. 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; 294:373–87. [PubMed: 10610765]
54. Gangelhoff TA, Mungalachetty PS, Nix JC, Churchill ME. Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A. *Nucleic Acids Res.* 2009; 37:3153–64. [PubMed: 19304746]
55. Lavery R, Sklenar H. Defining the Structure of Irregular Nucleic Acids: Conventions and Principles. *J Biomol Structure Dynamics.* 1989; 6:655–667.
56. Zhang J, McCauley MJ, Maher LJ 3rd, Williams MC, Israeloff NE. Mechanism of DNA flexibility enhancement by HMGB proteins. *Nucleic Acids Res.* 2009; 37:1107–14. [PubMed: 19129233]
57. Dragan AI, Read CM, Makeyeva EN, Milgotina EI, Churchill ME, Crane-Robinson C, Privalov PL. DNA binding and bending by HMG boxes: energetic determinants of specificity. *J Mol Biol.* 2004; 343:371–93. [PubMed: 15451667]
58. Jones DNM, Searles A, 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. [PubMed: 7922039]
59. Dow LK, Jones DN, Wolfe SA, Verdine GL, Churchill ME. Structural studies of the high mobility group globular domain and basic tail of HMG-D bound to disulfide cross-linked DNA. *Biochemistry.* 2000; 39:9725–36. [PubMed: 10933789]
60. Hardman CH, Broadhurst RW, Raine ARC, Grasser KD, Thomas JO, Laue ED. Structure of the A-domain of HMG1 and its Interactions with DNA as Studied by Heteronuclear Three- and Four-Dimensional NMR Spectroscopy. *Biochemistry.* 1995; 34:16596–16607. [PubMed: 8527432]
61. Weir HM, Kraulis PJ, Hill CS, Raine A, Laue ED, Thomas JO. Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.* 1993; 12:1311–1319. [PubMed: 8467791]
62. He Q, Ohndorf UM, Lippard SJ. Intercalating residues determine the mode of HMG1 domains A and B binding to cisplatin-modified DNA. *Biochemistry.* 2000; 39:14426–35. [PubMed: 11087395]
63. Webb M, Payet D, Lee KB, Travers AA, Thomas JO. Structural requirements for cooperative binding of HMG1 to DNA minicircles. *J Mol Biol.* 2001; 309:79–88. [PubMed: 11491303]
64. 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. [PubMed: 7720717]
65. Dow L, Changela A, Hefner H, Churchill M. Oxidation of a critical methionine modulates DNA binding of the *Drosophila melanogaster* high mobility group protein, HMG-D. *FEBS Letters.* 1997; 414:514–520. [PubMed: 9323026]
66. Murphy FV IV, Sehy JV, Gao Y-G, Dow LK, 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.
67. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.
68. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. *Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination.* *Acta Crystallogr.* 1998; D54:905–921.
69. Laskowski RA. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst.* 1993; 26:283–291.
70. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003; 31:5108–21. [PubMed: 12930962]
71. Mitchell JC, Kerr R, Ten Eyck LF. Rapid atomic density methods for molecular shape characterization. *J Mol Graphics and Modelling.* 2001; 19:325–330.
72. Kraulis PJ. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *Journal of Applied Crystallography.* 1991; 24:946–950.
73. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA: 2008.

74. Merritt EA, Bacon DJ. Raster3D: Photorealistic Molecular Graphics. *Methods in Enzymology*. 1997; 277:505–524. [PubMed: 18488322]
75. Humphrey W, Dalke A, Schulten K. VMD - visual molecular dynamics. *J Molec Graphics*. 1996; 14:33–38.
76. Kleywegt GJ, Jones TA. Detecting folding motifs and similarities in protein structures. *Methods Enzymol*. 1997; 277:525–45. [PubMed: 18488323]

Figure 1.

HMGB sequences and structures. (A) Sequence alignment of HMGB1 boxes A and B, NHP6A, and HMGD. The sequences are aligned and numbered according to the HMGD structure with the alpha helices depicted by rectangles. Residues known to intercalate the DNA are shaded in grey. (B) The ribbon diagram shows the crystal structure of wildtype HMGD bound to DNA (PDB ID 1qrv), with the intercalation sites indicated by arrows and the intercalating residues labeled. (C) The ribbon diagram shows the crystal structure of HMGB1 box A bound to cisplatin-modified DNA (PDB ID 1ckt), with the intercalation site indicated by an arrow and the intercalating residue labeled.

Figure 2.

Structure of the HMGD^{M13A}-DNA complex. (A) Overall crystal packing diagram. The unit cell is shown with additional DNA from adjacent asymmetric units to illustrate the mode of packing within the crystal. The model contains six HMGD^{M13A} proteins and five DNA decamers that are arranged in a continuous DNA helical stack throughout the crystal. (B) Electron density of a protein-DNA contact region. The σ_a -weighted 2Fo-Fc map contoured at 1.2 σ is shown in tan, with the underlying model colored according to the B-factor of the model, from blue (lowest B factors) to red (highest B-factors). Selected residues are labeled with the chain ID, residue number, and residue name for clarity.

Figure 3.

HMGD-M13A Protein-DNA Interactions. The DNA ladder diagrams on the left show the contacts made by each protein with the DNA binding site, and are labeled according to the chains in PDB ID 3NM9. The amino acids that make major contacts with the DNA are shown with lines to indicate inferred intermolecular hydrogen bonding or short-range ion pairs, and amino acid residues that are shown make substantial van der Waals interactions with the DNA. The identity of each DNA position is shown in each sugar ring. The sites of DNA intercalation are shaded in grey. On the right of each ladder diagram are the individual protein-DNA complexes shown in a view from the side and from the top to illustrate the different configurations of DNA to which the proteins bind.

Figure 4.

Detailed views of protein-DNA interactions. Two pairs of proteins form head-to-head interactions in the crystal. The binding sites of molecules A and M (A) and D and J (B) are shown with the protein colored with rainbow coloring from blue to red along each pair of HMG boxes from the N-terminus of domain A or D to the C-terminus of molecule M or J. The DNA is in grey. Amino acid residues Lys4, Pro5, Lys6, Arg7, Pro8, Ala13, Val32, and Thr33 are shown in stick form. (C) Close up views of the water mediated interaction of Ser10 with DNA purine bases are shown for (C) chains D-E/F and chains P-K/L. The stick diagram is colored with carbon in grey, oxygen in red, nitrogen in blue, phosphorus in orange, and with the water molecule shown as a red sphere. The putative hydrogen bonds are shown as black dotted lines with distances indicated. (E) A close up view of the intercalation by Val32 and Thr33 observed in chain G bound to chains H/I. The stick diagram is colored with carbon in grey for protein and in green for DNA, oxygen in red, nitrogen in blue and phosphorus in orange. The intercalation of Val32 and Thr33 occurs at the strand H Cyt2-Gua3 base step in the DNA.

Figure 5.

Structural differences between the wildtype HMGD and HMGD^{M13A} DNA complexes. (A) HMGD^{M13A} bends DNA less than wildtype HMGD. All of the HMGD^{M13A} structures with associated DNA were superimposed onto the wildtype HMGD-DNA complex (PDB ID 1qrv) and calculated helix axes generated by CURVES were compared. The wildtype HMGD complex is shown in red. The calculated helix axes of the DNA bound to models D and J are shown in blue and the axes for the other binding sites are shown in grey. (B) Superposition of HMG boxes showing HMG boxes alone. The HMGD wildtype protein is shown in red and HMGD^{M13A} models are in grey. (C) The N-terminal strand interactions with the DNA minor groove are shown. (D) The fit of alpha helices 1 and 2 in the DNA minor groove.

Table I

Data collection and refinement statistics

P1 space group cell dimensions	44.75 × 71.70 × 89.02 Å
	$\alpha = 92.49^\circ$; $\beta = 91.12^\circ$; $\gamma = 107.10^\circ$
Resolution range (high resolution shell) (Å)	20–2.80 (2.91–2.80)
Unique reflections	23,992
Redundancy	2.6
R_{sym}^a	6.0 (20.8)
Completeness (%)	92.3 (82.1)
Intensity (I/σ)	21.9 (4.6)
Refinement	
Resolution Range (high resolution shell)(Å)	20–2.85 (2.92–2.85)
R_{free}^b (%)	29.5 (44.9)
Rworking (%)	24.5 (42.2)
Model	
Number of protein atoms	3510
Number of DNA atoms	2042
Number of solvent atoms	4
Average B factor (Å ²)	58.1
R.m.s.d. bond lengths (Å)	0.009
R.m.s.d. bond angles (°)	1.34
Ramachandran Analysis	87.7% most favored; 12.3% allowed

$$^a R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$$

$^b R_{\text{free}}$ calculated with an excluded set of 1,118 reflections (5.15%)

Table II

Structural parameters of the protein-DNA complexes.

Protein and bound DNA chains		Site	Local axis curvature (°) ^c	Intercalation site Roll (°) ^c	Minor groove width (Å) ^c	Minor groove depth (Å) ^c	Buried surface area (Å ²) ^d	FADE packing surface (points)	FADE packing density (score)
A	B/C	2°	7.4	11.5	9.6	3.2	665.9	842	-0.098
	N/O	R7			5.6	4.8			
		L9			nd	nd			
D	F/E	2°	13.4	14.73	11.2	1.2			-0.082
		L9			10.2	1.8	800.8	1267	
	K/L	R7			nd	nd			
G	F/E	R7			4.0	5.8			0-101
		L9			8.1	3.7	933.8	1471	
	H/I	2°	24.5	45.2	11.4	-0.01			
J	L/K	2°	10.1	10.2	10.9	2.6			-0.026
	E/F	L9			5.2	5.6	707.8	971	
		R7			nd	nd			
M	O/N	L9			10.2	2.3			-0.037
		R7			3.8	5.3	659.2	902	
	I/H ^a	*		*	nd	nd			
P	L/K	R7			4.3	5.0			-0.095
		L9			nd	nd	733.4	1029	
	C/B	*		*	nd	nd			
HMGD	C/D	1°	29.0	48.50 ^b	11.8 ^b	0.3 ^b			-0.139
							821.2	1665	
	C/D	2°	18.5	32.00 ^b	10.1 ^b	1.6 ^b			

^afrom a symmetry related molecule.

^bfrom 39.

^cfrom CURVES 55

d_p from CNS 68

* no apparent 2° intercalation site

nd due to inability to determine groove width near DNA ends

Table III

Average structural parameters of the DNA.

DNA duplex	Bound proteins	Helical twist (std) ^a	Roll (std) ^a	Propeller twist (std) ^a	Global helix axis curvature ^b
B/C	A, P	34.7° (9.5)	2.2° (4.22)	-9.6° (10.2)	5.5°
E/F	D, G, J	33.3° (4.8)	8.4° (6.6)	-8.7° (6.3)	43.2°
H/I	G, M	38.5° (9.2)	7.2° (17.9)	-6.8° (4.9)	8.9°
K/L	D, J, P	34.7° (11.4)	8.3° (9.3)	-17.9° (8.5)	34.5°
N/O	A, M	35.1° (4.8)	4.1° (5.1)	-9.5° (6.5)	10.3°
1qrv.pdb					99.4°

^aFrom 3DNA 70

^bFrom CURVES 55

Table IV

Structural comparison of the HMGD^{M13A} proteins to wildtype HMGD

HMGD ^{M13A} Protein Chain	r.m.s.d. HMGD	r.m.s.d. Chain A	r.m.s.d. Chain D	r.m.s.d. Chain G	r.m.s.d. Chain J	r.m.s.d. Chain M
A	0.94					
D	0.94	0.54				
G	0.87	0.61	0.53			
J	1.15	0.95	0.63	0.79		
M	1.12	0.62	0.56	0.67	0.90	
P	1.13	0.85	0.61	0.76	0.61	0.78

Calculated using LSQMAN ⁷⁶