The A·T-DNA-binding Domain of Mammalian High Mobility Group I Chromosomal Proteins

A NOVEL PEPTIDE MOTIF FOR RECOGNIZING DNA STRUCTURE*

Raymond Reeves‡ and Mark S. Nissen‡

From the ‡Department of Biochemistry/Biophysics and §Program in Genetics and Cell Biology, Washington State University, Pullman, Washington 99164

We have determined the domains of the mammalian high mobility group (HMG)I chromosomal proteins necessary and sufficient for binding to the narrow minor groove of stretches of A·T-rich DNA. Three highly conserved regions within each of the known HMG-I proteins is closely related to the consensus sequence T-P-K-R-P-R-G-R-P-K-K. A synthetic oligopeptide corresponding to this consensus “binding domain” (BD) sequence specifically binds to substrate DNA in a manner similar to the intact HMG-I proteins. Molecular Corey-Pauling-Koltun model building and computer simulations employing energy minimization programs to predict structure suggest that the consensus BD peptide has a secondary structure similar to the antitumor and antiviral drugs netropsin and distamycin and to the dye Hoechst 33258. In vitro these ligands, which also preferentially bind to A·T-rich DNA, have been demonstrated to effectively compete with both the BD peptide and the HMG-I proteins for DNA binding. The BD peptide also contains novel structural features such as a predicted Asx bend or “hook” at its amino-terminal end and laterally projecting cationic Arg/Lys side chains or “bristles” which may contribute to the binding properties of the HMG-I proteins. The predicted BD peptide structure, which we refer to as the “A·T-hook,” represents a previously undescribed DNA-binding motif capable of binding to the minor groove of stretches of A·T base pairs.

High mobility group (HMG)I protein HMG-I is the most thoroughly studied member of a group of isoform mammalian nonhistone chromosomal proteins that appear to be preferentially expressed in undifferentiated, neoplastically transformed, and rapidly proliferating cells (1-8). HMG-Y is a second member of this family found in abundance in human and mouse cells but differs from HMG-I by possessing an internal deletion of 11 amino acids (6, 8, 9). The functional mRNAs coding for the HMG-I and HMG-Y proteins seem to be derived by alternative splicing from a common precursor mRNA and may represent only two of the most abundant forms of several alternatively processed HMG-I-like mRNA species in human and murine cells (6, 8). Collectively the isoform members of this family are referred to as the HMG-I proteins (8). The number of members and the degree of variation within the mammalian HMG-I family are currently unknown. Nevertheless, based on either their biochemical or biological characteristics other potential members have been reported including the c satellite protein of monkey cells (10), the HMG-I' protein of virus-transformed rat fibroblasts (3), and the HMG-like protein of rat hepatoma cells (11). Definitive assignment of these proteins to the HMG-I family must, however, await determination of their primary sequences.

The HMG-I proteins have been shown to specifically bind in vitro to the minor groove of A·T-rich regions of double-stranded DNA (10, 12, 13). This DNA-binding characteristic has led to several postulated functions for these proteins including nucleosome phasing (10), the 3'-end processing of mRNA transcripts (14), and possible involvement in amplification of autonomously replicating sequences in mouse cells (15). The HMG-I proteins may also be involved in transcriptional regulation of genes containing, or in close proximity to, A·T-rich regions of DNA such as those found in the 3'-untranslated regions (3'-UTRs) of many lymphokine and growth factor genes (13) and in the enhancer sequence found in the spacer regions of rat rRNA genes (11).

Recently the human HMG-I proteins have been shown to specifically bind to the human immunoglobulin light chain promoter octamer (OTF-1 and OTF-2) DNA sequence, suggesting that these proteins might also contribute to the regulation of genes containing this regulatory element (16). In these instances the HMG-I proteins might affect gene regulation either by acting as transcriptional activators (11, 16) and/or by changing the conformation of DNA or chromatin and hence altering the accessibility of A·T-rich regions of DNA to binding by other transcription factors (8). It is known from their primary sequences that the HMG-I proteins (6, 8) have the overall structure of typical Ptashne type transcriptional activator proteins possessing both a DNA-binding domain and a highly acidic COOH terminus (17). Furthermore, in vitro binding of HMG-I has been demonstrated to alter both the conformation and thermal stability of A·T-rich regions of DNA (9).

Immunocytochemical studies have localized members of the HMG-I family in vivo to the A·T-rich G/Q and C bands of mammalian metaphase chromosomes (18). These and other results strongly suggest that, in addition to their possible involvement in gene regulation (11-16) and in the establishment or maintenance of the undifferentiated state of chromatin in non-metaphase cells (5-7), the HMG-I proteins may also play important roles in the chromosomal structural changes occurring during mitosis and cell division (1, 8, 18).
Thus, members of the HMG-I protein family appear to have multifaceted involvements in several aspects of cellular metabolism. Most of which are probably mediated by the unusual ability of these proteins to selectively bind to A·T-rich regions of duplex DNA. In this regard it is of interest that the A·T-binding antiviral and antitumor drugs distamycin A and netropsin, as well as the dye Hoechst 33258, effectively compete in vitro with HMG-I proteins for binding to DNA (8, 9), suggesting that all of these molecules share some structural similarities (8, 9, 12, 26).

Here we report the determination of the regions of the HMG-I proteins involved in specific binding to the minor groove of A·T-rich B-form DNA. An 11-residue long synthetic peptide, T.P.K.R-P.R-P.K-P.K, corresponding to a conserved consensus binding domain found repeated three times in all known HMG-I proteins, was demonstrated to have the same specificity as whole HMG-I proteins in DNA-binding and footprinting assays. Molecular modeling studies indicate that the consensus peptide structure represents a new DNA-binding unit different from the previously well characterized α-helix, β-sheet, and zinc finger motifs. This newly defined DNA-binding domain, which we call the “A·T·hook” motif, has certain structural similarities to distamycin A, netropsin, and Hoechst 33258 but also contains novel features that help to clarify many of the previously unexplained aspects of HMG-I binding specificity.

**METHODS AND MATERIALS**

**Purification of Murine HMG-I and HMG-Y**—Crude HMG protein samples were obtained by extracting murine MRL ascites cells with 5% perchloric acid (9). Proteins were precipitated from the pooled extracts by adding an equal volume of 30% trichloroacetic acid. The precipitated HMG proteins were collected by centrifugation, washed with acetone, dried, and resuspended in water. HMG-I and HMG-Y proteins were purified from the crude extract by reverse-phase high-performance liquid chromatography as previously described (4). Following chromatography the purity of the protein samples was determined by acid-urea polyacrylamide gel electrophoresis (4). HMG protein concentrations were determined spectrophotometrically using the relationships ε = 74,000 l/mmol cm (HMG-I) and ε = 68,000 l/mmol cm (HMG-Y). These relationships were obtained from amino acid composition analysis of solutions of purified proteins of known absorbance at 270 nm.

**DNA Purification and Footprint Analysis**—The 3' UTR of the bovine interleukin-2 cDNA was isolated by standard methods (19) as a 300-base pair HindIII/EcoRI restriction fragment from the plasmid pBluescript. This plasmid was constructed by ligating the BcII/EcoRI 3' UTR of pBIL-2 cDNA was performed as previously described (22). Scans of biotinylated 3' UTR DNA was mixed with a 5-fold molar excess of column followed by ethanol precipitation. In a typical experiment, samples were obtained by extracting murine MRL ascites cells with 50 mM NaCl, 1 mM EDTA, 3' UTR DNA, and various concentrations of HMG-I or -Y protein or peptides were titrated with the fluorescent dye Hoechst 33258. DNA concentration was fixed in all experiments at 100 nM as phosphate. Final concentrations of Hoechst 33258 ranged from 0 to 50 nM. Hoechst 33258 was obtained from Sigma. Fluorescence of the dye was excited at 354 nm and observed at 450 nm using a Shimadzu RF-540 fluorescence spectrophotometer. The change in fluorescence (ΔF) of the sample due to binding of the dye by DNA was expressed as the difference of the fluorescence of the test solution and a “blank” containing only buffer and dye.

As previously demonstrated (37), binding of Hoechst 33258 to DNA in the presence or absence of competitors may be analyzed using the Michaelis-Menten equation. Curve fitting of the experimental data (ΔF versus [dye]) was performed using the nonlinear regression analysis program Enzfitter (24). This permitted direct determination of the dissociation constant of the kapp in the presence of competing ligand. Using the relationship kapp = Kd(1 + [ligand]/Ks) the dissociation constant of the competing ligand was determined from a plot of kapp versus [ligand].

**Molecular Modeling**—Scale molecular models of the 11-amino acid DNA-binding domain peptide and of double-stranded A·T-rich B-form DNA sequences were constructed using CPK space-filling atomic model components (Harvard Apparatus, Inc., S. Natick, MA). A computer graphics study was conducted at the VADAMS Laboratory, Washington State University, using a graphics display terminal (Tektronix 4701) controlled by a VAX3 computer. The MacroModel (version 2.5) interactive molecular modeling software system (25) was used for computer model simulations. Molecular geometry optimizations were determined using the Powell and Dunitz (32) and Ponder TNGC algorithms to refine crude peptide structures to minimum energy employing the AMBER and OPLSA force field parameters.

**RESULTS**

**HMG-I DNA-binding Domains Determined by Protease Digestion**—Fig. 1 compares the amino acid sequences of the human and mouse HMG-I and HMG-Y proteins as determined by either direct peptide sequencing or deduced from isolated cDNAs. As previously noted (6, 8, 9), the human and murine HMG-I proteins differ from their HMG-Y isoforms by containing an additional 11 internal amino acids (residues 35-45). Also, as shown in bold type in Fig. 1, each of these HMG-I proteins contain two highly conserved palindromic peptide motifs, P-R-G-R-P (residues 57-61 and 83-87), and a third partially degenerate motif, R-G-R-P (residues 26-29), that are in each case flanked by basic arginine or lysine residues. It has been suggested that these conserved P-R-G-R-P palindromic sequences are likely to be the DNA-binding domains of the HMG-I proteins (27). Synthetic peptides corresponding to this conserved pentameric sequence alone failed, however, to specifically bind to A·T-rich DNAs known

---

2. Reeves and M. S. Nissen, unpublished data.
also found to be repeated with a surprisingly high degree of protected peptides and the known sequences of human and contained the conserved versions of the palindromic amino DNA binding. An II-amino acid consensus "protected" pep-
ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

ties. As will be demonstrated below, this synthetic oligopep-

Similar results were obtained with the murine HMG-Y pro-
teins except that the fragment corresponding to peptide-l was
to amino acid residues 31-55 and peptide- to residues 56-71.

Specific regions of the protein from protease digestion. Sub-
test DNA (the 3'-UTR of the bovine interleukin-2 cDNA)
specific A.T-binding. Trypsin had been end-labeled with a biotinylated nucleotide "tag" that had been end-labeled with a biotinylated nucleotide "tag" was mixed with pure murine HMG I protein under conditions known to be appropriate for specific A.T-binding. Trypsin was added to the mixtures and, following limited digestions, the resistant DNA-peptide complexes isolated by sequential addition of excess streptavidin and biotin-cellulose. The specific retained HMG-I peptides were eluted from the cellulose support and sequenced.

Results obtained from three separate experiments involving murine HMG-I indicated that two contiguous peptides were specifically and reproducibly protected by DNA binding from limited trypsin digestion. Protected peptide-1 corresponded to amino acid residues 31-55 and peptide-2 to residues 56-71. Similar results were obtained with the murine HMG-Y protein except that the fragment corresponding to peptide-1 was not recovered. Since trypsin cleaves peptides past the carbonyl groups of either lysine or arginine residues and since the experiments involved limited digestions it was possible that the protected peptides observed were larger than the actual DNA-binding domains of the HMG-I proteins. As shown in Fig. 2, comparison of the amino acid sequence of the DNA protected peptides and the known sequences of human and mouse HMG-I proteins revealed that a limited part of the DNA-protected fragments of the murine HMG-I protein was also found to be repeated with a surprisingly high degree of similarity in three separate regions of all of the known HMG-I proteins.

Furthermore, as shown by the bold letters in Fig. 2, all of the similar peptide regions within the HMG-I proteins also contained the conserved versions of the palindromic amino acid motifs previously postulated to be involved in specific DNA binding. An 11-amino acid consensus “protected” peptide, Thr-Pro-Lys-Arg-Pro-Arg-Gly-Arg-Pro-Lys-Lys, was therefore synthesized and tested for its DNA-binding capacities. As will be demonstrated below, this synthetic oligopeptide, hereafter referred to as the HMG-I "binding domain" (BD) peptide, was found to specifically bind to the minor groove of A.T-rich DNA in a manner indistinguishable from the intact HMG-I proteins although the affinity of binding was less than that of the intact proteins.

Competition of HMG-I Proteins and Peptides with Hoechst 33258 for DNA Binding—Fig. 3A demonstrates that the extent of specific binding of Hoechst 33258 to substrate DNA, as measured by fluorescence of the DNA-Hoechst complexes, is competitively inhibited by addition of increasing amounts of pure HMG-I protein to the reaction mixture. In this experiment increasing concentrations of murine HMG-I protein (0-10 nM) were added to a fixed amount of 3'-UTR DNA (100 nM) in the presence of increasing amounts of Hoechst dye (0-50 nM) and the fluorescence intensity of the resulting dye-DNA complexes measured at 450 nm. In all of the competition experiments described the concentrations of Hoechst 33258 employed were below the level of primary site saturation (data not shown) to avoid a known nonspecific mode of dye binding to DNA (37).

By using the nonlinear regression analysis program Enzfitter (24), the $K_o$ for the dye was determined directly from plots of changes in fluorescence and dye concentrations and the $K_o$ for HMG-I binding was determined from plots of $K_{app}$ versus [HMG I], where $K_{app}$ is the apparent dissociation constant for dye binding in the presence of a competitive inhibitor (Fig. 3, A and B). Such analyses gave a $K_o$ value for Hoechst 33258 of approximately 9.62 nM and a $K_o$ for the HMG-I protein of approximately 1.02 nM. Competition experiments were also performed using pure HMG-Y (Fig. 3C) and a $K_o$ of approximately 1.38 nM was obtained for this protein. Since the free energy, $\Delta G$, is calculated as $RT\ln K$ at 25 °C, the $\Delta G$ of Hoechst dye, HMG-I and HMG-Y were determined to be 10.9 kcal/mol, respectively. The $K_o$ and $\Delta G$ values obtained for Hoechst 33258 (9.62 nM; 10.9 kcal/mol) are similar to values (4.27-5.62 kcal/mol) obtained by others for this dye (52). Similarly, the $\Delta G$ values obtained for the HMG-I proteins (12.1-12.3 kcal/mol) are close to values reported by others (12.2-12.7 kcal/mol) for specific binding of netropsin to DNA (34).

Fig. 4. A and B, shows the results of dye competition experiments using the synthetic 11-amino acid HMG-I binding domain peptide. In these experiments the concentrations of the 3'-UTR substrate DNA and the Hoechst dye were the same as those used above for the whole HMG-I proteins but the concentration range of the competing BD peptide varied from 0 to 10 nM. The average $K_o$ determined for the BD peptide was 9.97 nM with a $\Delta G$ of 6.82 kcal/mol. While this value is several orders of magnitude lower than the $k_e$ values for intact HMG-I proteins, the BD peptide is still capable of

![Fig. 2. Highly conserved regions of the HMG-I proteins and consensus DNA-binding domain (BD) peptide.](https://example.com/fig2.png)
specifically competing with the Hoechst dye for binding (Fig. 4A).

A number of synthetic "control" peptides were also employed for similar types of competition experiments. The sequences of these control peptides were derived from either parts of the BD peptide itself, regions of the HMG-I proteins thought not to be involved in specific DNA binding, or peptides that had overall amino acid compositions similar to (but sequences different from) regions of the IIGM-I proteins. These control peptides included: 1) P-R-G-R-P (the conserved HMG-I palindromic sequence); 2) T-P-R-K (kentsin, similar in sequence to the amino-terminal residues of the BD peptides); 3) V-S-P-G-T-A-L-V-G-S-Q (the 11-amino acid peptide deleted from HMG-I in the isoform HMG-Y); and 4) K-G-R-P-P-K (a scrambled HMG-I protein sequence).

Representative results of these competition experiments, as illustrated by using the palindromic P-R-G-R-P peptide sequence, are shown in Fig. 4C. In contrast to the BD peptide, neither the palindromic peptide (Fig. 4C), nor any of the above synthetic control peptides, specifically competed with the Hoechst dye for binding to 3'-UTR DNA. All of these control peptides did, however, show a small degree of nonspecific interaction with DNA (cf. Fig. 4C), probably as a result of ionic bonding outside of the minor groove of the helix (37). In addition it was noted that one synthetic octameric peptide, K-R-R-P-R-P-K (corresponding to the BD peptide lacking 2 amino-terminal and 1 carboxyl-terminal residues), did seem to give a low, but discernible level, of specific competition with Hoechst dye for binding (data not shown). These results

![Diagram](https://via.placeholder.com/150)

**Fig. 3.** A, binding of Hoechst 33258 to A-T-rich DNA in the presence or absence of competing HMG-I(Y) protein. The solid curve (+) represents the change in fluorescence observed when 3'-UTR DNA was titrated with Hoechst 33258 in the absence of competing ligands. The dotted curves represent titration of DNA in the presence of 1.39 \(nM\) (A), 2.77 \(nM\) (m), 4.14 \(nM\) (v), 5.22 \(nM\) (○), and 6.89 \(nM\) (El) HMG-I protein. Similar results were obtained using HMG-Y protein. In all experiments the DNA concentration was fixed at 100 \(nM\) (expressed as phosphate) and the dye concentration varied from 0 to 50 \(nM\). Kapp of dye binding to DNA in the presence of competing protein was determined by nonlinear regression analysis of the AF versus [dye] curves. B and C, dissociation constant \(K_d\) of the specific DNA-protein interaction was determined from the ratio of intercept to slope of plots of \(K_{app} = [HMG-I] + K_{app} \times [HMG-Y].

![Diagram](https://via.placeholder.com/150)

**Fig. 4.** A, binding of Hoechst 33258 to 3'-UTR DNA in the absence of competing ligand (solid curve, +) or in the presence of 0.803 \(\mu M\) (C), 2.01 \(\mu M\) (A), 4.01 \(\mu M\) (○), 6.00 \(\mu M\) (v), and 9.95 \(\mu M\) (m) BD peptide. The DNA concentration was fixed at 100 \(nM\) and the dye concentration varied from 0 to 50 \(nM\). B, plot of \(K_{app} \times [BD peptide\). The \(K_d\) of the peptide-DNA complex was determined from the ratio of intercept to slope of the plot. C, Hoechst 33258 binding to 3'-UTR DNA in the absence of competitor (solid curve, +) or the presence of 10 \(\mu M\) PRGRP peptide. Conditions as in panel A.
Footprinting of the Binding Domain Peptide to 3′-UTR DNA—Three different methods of footprinting, DNase I, MPE-Fe(II), and hydroxyl radical, were employed to demonstrate specific binding of the BD peptide to substrate 3′-UTR DNA. On a qualitative basis all three methods gave essentially the same result. Namely, the BD peptide (but not control peptides) was found to specifically bind to substrate DNA and give footprints similar to those of both the intact HMG-I proteins and to the dye Hoechst 33258 (data not shown).

As an illustration of the typical types of results obtained by the three different footprinting methods, Fig. 5 shows a representative example of one such experiment. This figure shows a laser densitometry scan of an autoradiogram obtained using the MPE-Fe(II) footprinting method to analyze the interactions of HMG-I protein, BD peptide, kentsin, and Hoechst dye with an A-T-rich region of 3′-UTR (control) DNA. It is evident that the BD peptide and the HMG-I protein bind similarly to the 3′-UTR DNA. This specific binding is particularly evident in the A-T-rich region between nucleotides 694 and 712, where it is seen that both HMG-I and the BD peptide (but not kentsin) greatly inhibit cleavage of substrate DNA by MPE-Fe(II). Hoechst dye also binds specifically and tightly to this same region of A-T-rich DNA. However, Hoechst is seen to characteristically bind to slightly larger areas and to sequences slightly offset from those of the HMG-I proteins or the BD peptide in keeping with the known ability of the piperazine ring portion of this dye molecule to hydrogen bond with G/C residues immediately flanking stretches of A-T-rich sequence (39). From these and other experiments, we therefore conclude that the 11-amino acid BD peptide includes the amino acid sequences necessary and sufficient for specific A-T-binding of the HMG-I proteins.

Structure of the Binding Domain Peptide—Previous circular dichroism analyses of the structure of murine HMG-I proteins in solution revealed that this molecule displayed the spectrum typical of a “random coil” with little, if any, structural components corresponding to α-helix or β-sheet formations (data not shown). Structural predictions based on the Chou-Fasman method (35) did, nevertheless, suggest that bends or turns might exist at the threonine-proline residues located at positions 53 and 78 in the murine and human HMG-I proteins. These are the same dipeptide residues found by the analyses noted above to be at the amino termini of two of the three potential DNA-binding peptides found within the HMG-I proteins (Fig. 2). Further Chou-Fasman analyses of the peptide sequences shown in Fig. 2 predicted that significant curves or arcs might also be present in the backbones of each of these peptides due to the presence of closely spaced proline residues. Molecular model building based on these predictions using CPK space filling components to model the BD peptide resulted in the construction of a planar, curved, or arced shaped peptide with a strong structural resemblance to both Hoechst dye and to the drugs netropsin and distamycin. This physical scale model of the BD peptide was also observed to fit snugly into the narrow minor groove (but not in the wide major groove) of stretches of A-T-rich B-form DNA in space filling CPK scale models (data not shown).

Further support for such a secondary structure comes from the results of computer model building studies employing the interactive MacroModel software programs (25) for energy minimization and optimization to predict and refine the most probable hydrated structures of the BD peptide. Fig. 6A shows a computer predicted sideview projection of the partially hydrogenated backbone structure of the BD peptide after 1500 iterations of the energy minimization programs. Fig. 6B shows a space filling CPK model of the same peptide structure. As evident from these figures, the predicted structure of the 11-amino acid BD peptide has a marked crescent or arc shape due to bends in the relatively planar peptide backbone imparted by proline residues at positions 5 and 9. The peptide also is predicted to have a sharp bend or turn at the amino end of the molecule as the result of an intramolecular hydrogen bond formed between the hydroxyl of the terminal threonine residue and the NH of the lysine residue in position 3 (see below). Radiating out laterally like “bristles” from the relatively flat but curved peptide backbone are the six positively charged side chains of the lysine and arginine residues at positions 3, 4, 6, 8, 10, and 11. The predicted three-dimensional structure of the BD peptide can be better visualized by observing the computer-generated
peptide. Fig. 7, A and B, show the computer predicted stereo pairs of peptide backbone and space filling CPK models space filling model of the BD peptide. That the amino acid residues at the amino terminus of the BD shown in Fig. 6, A and B, respectively. Amino acids, proline has the greatest tendency to occupy the nature. This is based on the fact that among the 20 common residues such as threonine or serine occupying the first position—turn formation (36). Additional support for the second position in a $\beta$-type turn with slightly hydrophilic residues of the BD peptide being involved in secondary structure formation comes from the above-mentioned control experiments demonstrating that a synthetic peptide lacking these amino-terminal threonine and proline residues binds much less tightly to substrate DNA than does the full-length peptide. Fig. 7, A and B, show the computer predicted structures of the first 6 residues of the BD peptide based on energy minimization considerations. Instead of a $\beta$-I type backbone turn which is stabilized by two intrachain hydrogen bonds (37, 38), the bend structure at the end of the peptide appears to be similar to an "Asx turn" (39) involving the terminal Thr-Pro-Lys residues stabilized by a single hydrogen bond (Fig. 7A). The side chain to main chain interaction of an Asx turn is characterized by a hydrogen bond between the OH of amino acid residue 1 (in this case threonine) and the NH of residue 3 (lysine) of the turn (39). Having only one intramolecular bond, the Asx turn is less stable and more easily disrupted than a $\beta$-I bend (38).

The BD Peptide, Netropsin, Hoechst 33258, and $A\cdot T$-DNA Binding—As shown in Fig. 8, the BD peptide, netropsin, and Hoechst dye all share certain common structural features. In addition to having a relatively planar backbone and a similar type of arc or crescent shape, each of these molecules also possesses on its concave edge a number of potential hydrogen bonding donor amine groups indicated by arrows in Fig. 8. The BD peptide has three such amides (Arg$^6$, Arg$^7$, and Lys$^{16}$), netropsin has three, and the Hoechst dye has two. As mentioned before, the intrinsic crescent shape of the BD peptide backbone is due to the presence of appropriately positioned proline residues, while in netropsin a similar curvature is imparted by the presence of two methylypyrrole rings, and in Hoechst 33258 by the occurrence of two benzimidazole rings. In addition, each of these molecules has positively charged groups potentially capable of ionic interactions with DNA. In the BD peptide there are six such groups including the basic side chains of Lys$^3$, Arg$^4$, Arg$^6$, Arg$^7$, Lys$^{16}$, and Lys$^{17}$. The two basic guanidinium groups at the ends of netropsin are the primary ionic bonding parts of this molecule as is the charged piperazine ring at one end of Hoechst dye.

X-ray crystallographic analyses of complexes of netropsin (29), distamycin (30), and Hoechst 33258 (31) indicate that the critical amino acid residues involved in DNA binding are those that are specifically hydrogen bonded to DNA. These hydrogen bonds are donated by the amide NH groups of these molecules bridge adenine N-3 and thymine O-2 atoms occurring on adjacent base pairs and opposite helix straights, exactly as with the spine of hydration (29–31). Only part of the specificity of these ligands for A.T base may be attributed to these hydrogen bond configurations, however, since a major contribution to specificity also appears to be dependent on the close van der Waals contacts made between the floor and walls of the narrow minor groove and the surface of the drugs themselves (29, 32). Electrostatic attractions between the cationic groups of the drugs and the anionic phosphate groups of DNA also contribute to the stability of these DNA-drug complexes (26, 33).

As is evident from its structure as shown in Fig. 8, it is likely that the BD peptide also selectively binds to the minor groove of A.T-rich DNA replacing the inner spine of hydration by the donation of bifurcated hydrogen bonds from appropriately placed amide NH residues in the peptide backbone. Also, like these drugs, the specificity of binding of the peptide may, in part, be dependent on its ability to form appropriate close van der Waals contacts with the floor and walls of the narrow minor groove of A.T-rich DNA.

Two additional structural features of the BD peptide, not shared by the other ligand, may also contribute significantly to its ability to selectively recognize and bind A.T DNA. The first of these, as shown in Figs. 6–8, is the presence of a potential sharp Asx bend at the amino-terminal end of the BD peptide that may act as a "hook" or anchor helping to precisely position this end of the peptide at the junction between the narrow minor groove of a track of A.T sequences and the wider diameter of the minor groove in an adjacent G-C base pair. Furthermore, CPK model building suggests that there is no significant degree of steric hindrance to a continuation of the flexible peptide backbone anterior to such a sharp bend in the HMG-I proteins (data not shown). Never-
A. T-DNA-binding Peptide Motif

**FIG. 7.** Computer-generated structure of the first 6 residues at the amino terminus of the BD peptide showing the predicted Asx turn containing an intramolecular hydrogen bond (see text). a, peptide structure without hydrogen atoms; b, space-filling CPK model.

Nevertheless, owing to the fact that the predicted Asx bend is stabilized by only a single intrachain hydrogen bond, it is also possible that after its initial function as a positioning element for the peptide this bend may straighten somewhat to allow for a closer apposition of the amino-terminal end of the molecule to the interior of the minor groove. An analogous situation has been suggested for the repositioning of the piperazine ring located at the end of Hoechst dye molecules (31, 33).

**FIG. 8.** Comparison of the structures of the binding domain peptide, netropsin, and Hoechst 33258. Potential hydrogen bond donor groups, are indicated by arrows (see text).

The second structural feature of the BD peptide that differs significantly from netropsin and Hoechst dye, but which may be involved with the stabilization of a peptide-DNA complex, is the occurrence of six basic Arg/Lys side chains distributed along the length of the planar peptide backbone (Figs. 8 and 9a). As shown in Fig. 9d, these laterally projecting cationic bristles are positioned in such a way that they coordinate well for ionic bonding with the anionic phosphate groups of the two antiparallel DNA stands that define the width of the
FIG. 9. Interaction of the BD peptide with the minor groove of stretches of A·T-rich DNA. a, computer predicted structure of the BD peptide rotated in space to allow alignment for "docking" with the minor groove of the model DNA structure shown in c; b, CPK space filling model of the BD peptide shown in a; c, backbone structure of a computer-generated B-form DNA with the narrow minor groove of a stretch of A·T-sequences as indicated by arrows; d, proposed mode of interaction of the CPK model of the BD peptide shown in b with the minor groove of the B-form DNA shown in c. For detailed discussion of interactions refer to text.

minor groove. The presence of such a large number of potential ionic bonding groups in the BD peptide compared with the limited number of such cationic groups in netropsin and Hoechst dye would suggest that the stability of the interaction of this peptide with DNA is more salt concentration "sensitive" than is the binding of these drugs to DNA, a prediction confirmed by experimentation (data not shown). At sufficiently low salt concentrations these charged side chains might also contribute to the formation and stability of the BD·DNA complex as a consequence of the "polyelectrolyte effect" (40) in which electrostatic interactions and counterion release act as an entropic driving force for complexation of proteins and cationic ligands with DNA.

DISCUSSION

Only a limited number of different protein structural motifs have been identified that can recognize and bind to specific sequences of DNA. Among these are: (i) the helix-turn-helix structure of various prokaryotic repressor and activator proteins (41, 42) as well as eukaryotic homeotic gene products (43); (ii) symmetry-related pairs of α-helices or a pair of β-strands as found in a bacterial Met repressor protein (44); (iii) an antiparallel β-sheet found in the G5BP protein (45); and, (iv) the Zn finger structure originally identified in the transcription factor IIIA protein (46), but subsequently found in a large number of other eukaryotic regulatory proteins (47). Due to the size and secondary structure of these DNA-binding units they all, by necessity, bind primarily to the wide major groove of duplex DNA.

Here we report the structure of a new protein DNA-binding domain different from previously reported motifs. In contrast to these other DNA-binding units which localize primarily to the major groove of DNA, the newly defined binding domain of the HMG-I proteins, which we call the A·T-hook motif, specifically binds to the narrow minor groove of A·T-rich sequences of DNA.

Both physical and computer molecular model building studies indicate that the 11-amino acid residues of the consensus HMG-I binding domain (T-P-K-R-P-R-G-R-P-K-K) assume a secondary structure similar in many ways to the A·T DNA-binding ligands netropsin, distamycin A, and Hoechst 33258 (Fig. 8). Consistent with these suggested structural similarities are the observations that both Hoechst 33258 (Figs. 3 and 4), as well as the drugs distamycin and netropsin (8, 9) effectively compete with the HMG-I proteins and the BD peptide for specific DNA binding.

Two distinctive structural features predicted for the BD peptide not shared by these other A·T-binding ligands, the Asx turn or hook at the amino terminus (Fig. 7) and the cationic "R/K-bristles" projecting laterally from the planar backbone of the peptide (Figs. 5 and 6), may also significantly contribute to both the specificity and strength of DNA binding. The importance of a potential Asx bend at the end of the BD peptide is supported by the observation that a synthetic peptide lacking the first 2 residues (Thr-Pro) of the peptide (and hence unable to form an intramolecular Asx hook) binds much less tightly to substrate DNA than does the full-length peptide. The positively charged arginine and lysine side chains that constitute the lateral bristles of the BD peptide may contribute not only to the strength of DNA binding due to electrostatic interactions and possible polyelectrolyte effects (40), but they may also help to explain the observed differences in salt-dependent DNA binding between the BD peptide, the HMG-I proteins, and the other A·T-binding ligands (data not shown). Nuclear magnetic resonance studies are currently in progress to more precisely define the three-dimensional structure of the BD peptide and its interaction with DNA to test these speculations. In particular, such studies should determine whether an Asx turn actually plays
a significant role in BD peptide structure and function. As shown in Figs. 3 and 4, the $K_d$ values for intact HMG-I proteins (1.02–1.38 µM) are several orders of magnitude less than the $K_a$ for the synthetic 11-amino acid BD peptide (9.97 µM). One explanation for this large difference is that more than one of the three potential binding domains present in an intact protein may be cooperatively involved in tight, specific binding to DNA and/or that other peptide sequences in the protein outside of the binding domains help to stabilize specific DNA interactions. Such an explanation seems reasonable given the fact that an isolated proteolytic fragment of the NH$_2$ terminal end of the l repressor protein retains the same specificity of binding to λ DNA operator sites as does the intact protein but exhibits a greatly reduced affinity of binding to these sites (41). Further support for such postulated intramolecular binding domain cooperativity comes from the observation that a single zinc finger peptide domain of either the yeast transcription activator protein ADR1 (49) or of the Xenopus transcription factor protein TFIIIA (50) by itself does not bind tightly and specifically to substrate DNA, whereas two (49) or three (51) such zinc fingers present together on the same peptide do exhibit tight specific DNA binding.

An alternative explanation, and one which we currently favor as being the most likely explanation for the large difference in $K_a$ values observed between the HMG-I proteins and the BD peptide, is that the synthetic BD oligopeptide is in reality a mixture of many different isoform molecules not all of which have the necessary A-T-hook type structural conformation required for specific binding to the minor groove of A-T-rich DNA. Owing to the presence of 3 proline residues within the BD peptide, each of which is capable of spontaneous cis-trans isomerization (52), we expect that only a subset of the synthetic BD peptide molecules are in the requisite configuration for specific binding. This proline isomerization-induced conformational variability could reduce the effective concentration of the actual BD binding peptide in solution thus resulting in a greatly reduced apparent binding affinity (53).

If intramolecular cooperativity indeed contributes to the overall affinity of specific binding of intact HMG-I proteins to DNA, it is likely that the different domains do not contribute equally to the strength of binding since each probably exhibits slightly different substrate affinities due to minor variations in sequence and structure. For example, peptide II (residues 53–63) of human HMG-I (Fig. 2) is virtually identical to the consensus synthetic BD peptide employed in most of the studies reported here and might be expected to have a greater DNA-binding affinity than peptides I and III (residues 21–31 and 78–89, respectively) of the human protein which deviate somewhat from the consensus sequence. Consistent with this suggestion is the recent observation of Karlson et al. (48) that a large thermolysin generated fragment (Th-3) of the human HMG-I protein encompassing peptide II binds more tightly to duplex poly(dA-dT) substrate than do other large proteolytic fragments containing either peptides I or III. Thus, the precise manner of binding of intact HMG-I proteins to DNA may be influenced by complex interactions involving several different regions or domains of these proteins.
The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure.
R Reeves and M S Nissen


Access the most updated version of this article at http://www.jbc.org/content/265/15/8573

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/15/8573.full.html#ref-list-1