The solution structure of an HMG-I(Y)–DNA complex defines a new architectural minor groove binding motif

Jeffrey R. Huth[1], Carole A. Bewley[1], Mark S. Nissen[2], Jeremy N.S. Evans[2], Raymond Reeves[2], Angela M. Gronenborn[1] and G. Marius Clore[1]

The solution structure of a complex between a truncated form of HMG-I(Y), consisting of the second and third DNA binding domains (residues 51–90), and a DNA dodecamer containing the PRDII site of the interferon-β promoter has been solved by multidimensional nuclear magnetic resonance spectroscopy. The stoichiometry of the complex is one molecule of HMG-I(Y) to two molecules of DNA. The structure reveals a new architectural minor groove binding motif which stabilizes B-DNA, thereby facilitating the binding of other transcription factors in the opposing major groove. The interactions involve a central Arg-Gly-Arg motif together with two other modules that participate in extensive hydrophobic and polar contacts. The absence of one of these modules in the third DNA binding domain accounts for its ~100 fold reduced affinity relative to the second one.

Members of the non-histone chromosomal high mobility group HMG-I(Y) family are required for the assembly of higher order transcription enhancer complexes which are central for transcriptional activation of a number of important genes[1], the best characterized of which is the interferon-β gene (IFN-β)[2–4]. In addition HMG-I(Y) has been implicated as a crucial host protein for HIV-1 viral integration[5], an activity that may be related to observations that fusions of HMG-I(Y) and cellular proteins are associated with specific chromosome translocations frequently found in human lymphomas and leukemias[6–9].

The HMG-I family of proteins includes the two isoform proteins HMG-I and HMG-Y and the homologous protein HMG-I(C)[1]. These proteins, which have a relative molecular mass of ~10,000, comprise variable sequences at the N terminus, an acidic C terminus, and three short DNA binding domains (DBDs) separated by linkers of 11–23 amino acids (Fig. 1). The three DBDs, which have also been termed AT-hooks, recognize a wide variety of A,T rich sequences four to eight base pairs in length. Despite its relaxed specificity, HMG-I binds to DNA with nanomolar affinity[10] and methylation interference studies have located its binding site to the minor groove of A,T-tracts[2].

In this paper, we report the structure of a truncated form of HMG-I (residues 50–91) comprising the second and third DBDs, with a DNA dodecamer comprising an HMG-I target site in the IFN-β gene promoter. The structure indicates how HMG-I recognizes multiple regulatory elements and may modulate the conformation of DNA while in a ternary complex with other transcription factors.

DNA-binding stoichiometry and affinities

Initial studies were carried out on full length HMG-I bound to an 18mer oligonucleotide with eight sequential A,T base pairs using a sequence derived from selection experiments (R.R., unpublished data). Although many intermolecular NOEs were observed for this complex, poor spectral dispersion of the protein resonances suggested an unstructured or poorly defined conformation for the majority of amino acids. This is probably due to the 18mer DNA being too short to accommodate all three DBDs, such that only one or two of the three DBDs are bound at any one time resulting in a mixture of complexes. Since two DBDs are sufficient to confer DNA binding with nanomolar affinity[11], a truncated form of HMG-I, referred to here as HMG-I(2/3), was prepared comprising the second (DBD2) and third (DBD3) DBDs and consisting of residues 50–91 of full length HMG-I (Fig. 1*a*). Complexes of $^{13}C/^{15}N$ labelled HMG-I(2/3) and DNA dodecamers containing A,T-tracts of five or eight base pairs were screened for optimal DNA spectra. Analysis of the NMR spectra of the longer A,T-tracts was hindered by the inability to unambiguously assign all the H2 resonances of the adenines, which is essential for positioning and orienting the protein in the minor groove. Consequently, we chose to study a complex of HMG-I(2/3) with a dodecamer comprising a five base pair A,T-tract from the PRDII element of the IFN-β enhancer (Fig. 1*b* and hereafter referred to as the PRDII dodecamer) for which all the adenine H2 protons were readily assigned.

The stoichiometry of binding was established by a series of titration experiments using $^{15}N/^{13}C$[Gly 11]-HMG-I(2/3) and $^{15}N/^{13}C$-[Gly 37]-HMG-I(2/3) in which a single glycine in DBD2 and DBD3 respectively was isotopically labelled (Fig. 2*a*). HMG-I(2/3) binds to the PRDII dodecamer with DBD2 on the slow side of intermediate exchange; at ratios of DNA to protein less than 1:1, the cross-peak of Gly 11 in the $^{1}H-^{15}N$ correlation spectrum is broadened out (note this may also be due to non-specific interac-

The first two authors contributed equally to this work. [1]Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-520, USA. [2]Department of Biochemistry and Biophysics, Washington State University, Pullman, Washington 99164-4660, USA.

Correspondence should be addressed to G.M.C. *email: clore@vger.niddk.nih. gov* and A.M.G. *email: gronenborn@vger.niddk.nih.gov*

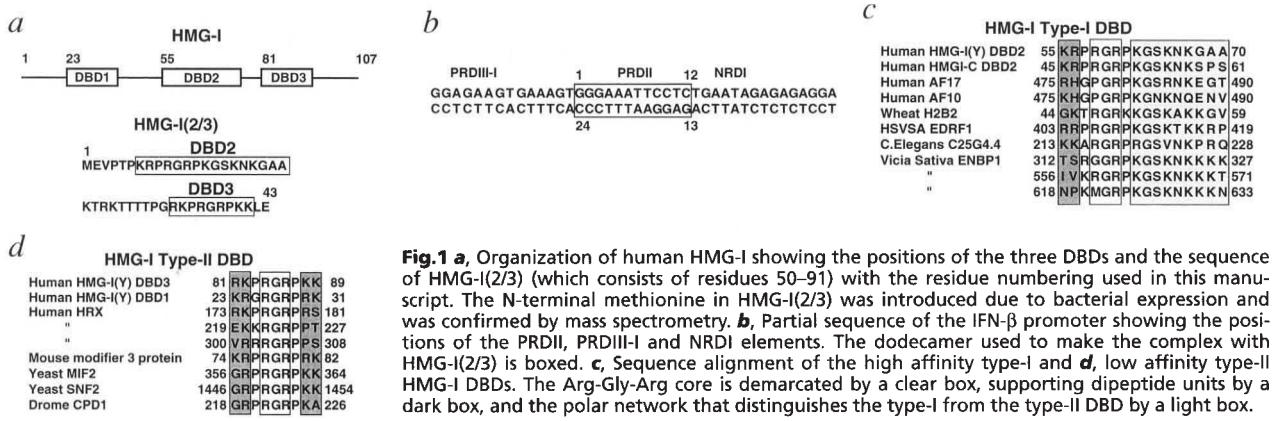
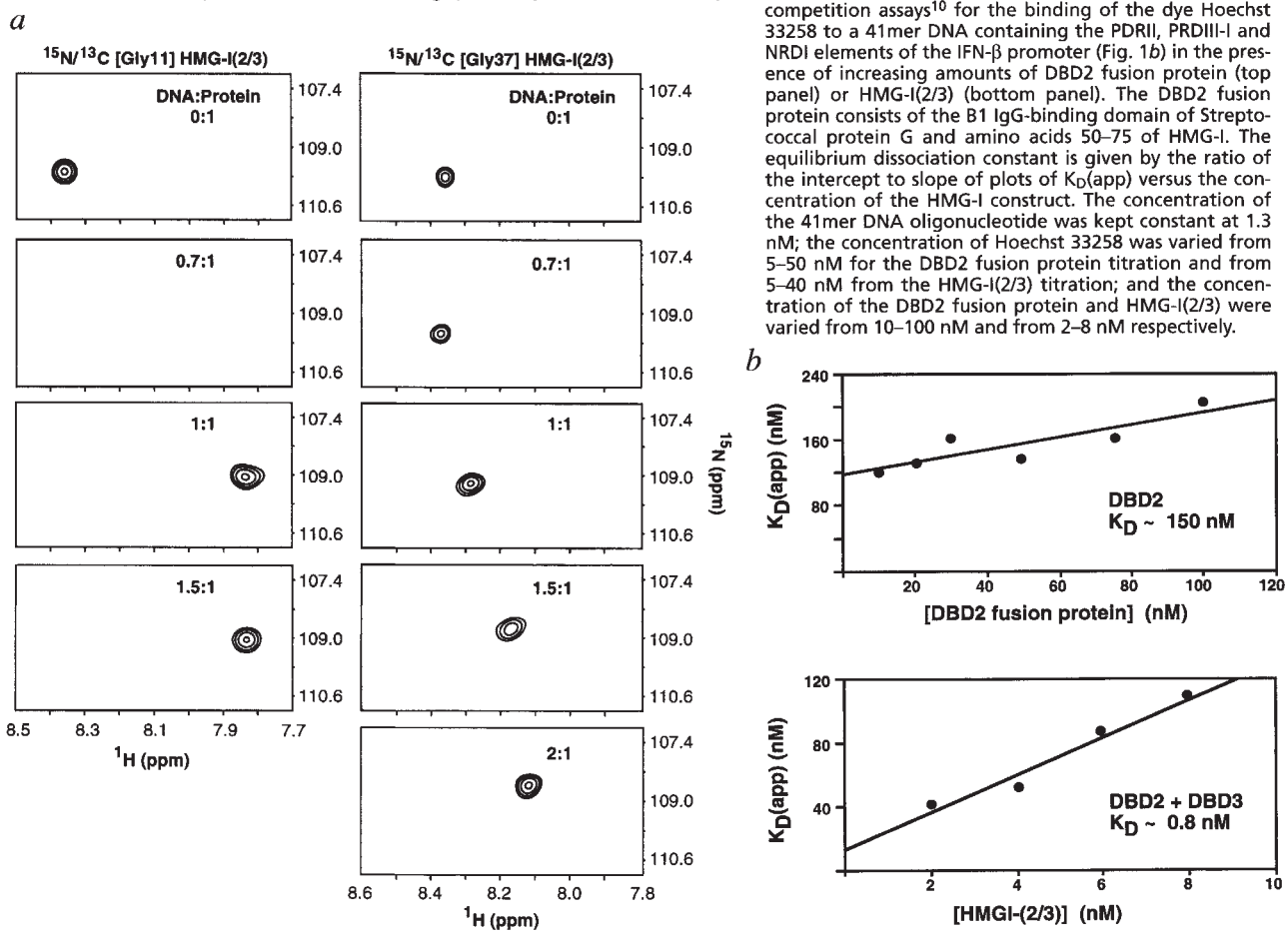


Fig.1 ***a***, Organization of human HMG-I showing the positions of the three DBDs and the sequence of HMG-I(2/3) (which consists of residues 50–91) with the residue numbering used in this manuscript. The N-terminal methionine in HMG-I(2/3) was introduced due to bacterial expression and was confirmed by mass spectrometry. ***b***, Partial sequence of the IFN-β promoter showing the positions of the PRDII, PRDIII-I and NRDI elements. The dodecamer used to make the complex with HMG-I(2/3) is boxed. ***c***, Sequence alignment of the high affinity type-I and ***d***, low affinity type-II HMG-I DBDs. The Arg-Gly-Arg core is demarcated by a clear box, supporting dipeptide units by a dark box, and the polar network that distinguishes the type-I from the type-II DBD by a light box.

tions resulting in the formation of higher order structures since some precipitation is observed at ratios of DNA to protein below 1:1 which completely clears up upon further addition of DNA); at a 1:1 ratio the cross peak is located at the position of the bound form and does not shift upon further addition of DNA, although it does sharpen up slightly. In contrast, Gly 37 of DBD3 is in fast exchange and its cross peak in the ^{1}H-^{15}N correlation spectrum only begins to shift from its position in the free state at ratios of DNA to protein greater than about 0.8:1, reaching the position of the bound form at a ratio of 2:1. Thus, under our experimental conditions, one molecule of HMG-I(2/3) binds two molecules of the PRDII dodecamer, the equilibrium association constant for the interaction of DBD3 with the PRDII dodecamer is ~$1.0(\pm0.5)\times10^{5}$ M^{-1}, and the affinity of DBD2 for the PRDII dodecamer is about one to two orders of magnitude greater than that of DBD3. These results are consistent with data from fluorescence competition assays which indicate that the overall association constants for the binding of HMG-I(2/3) and a DBD2

Fig. 2 ***a***, ^{1}H-^{15}N correlation spectra illustrating the change in position of the cross-peaks of Gly 11 and Gly 37 upon the addition of increasing amounts of DNA. The spectra were recorded using specifically labelled $^{15}N/^{13}C$-[Gly 11] and $^{15}N/^{13}C$-[Gly 37] HMG-I(2/3). ***b***, Results of fluorescence competition assays[10] for the binding of the dye Hoechst 33258 to a 41mer DNA containing the PDRII, PRDIII-I and NRDI elements of the IFN-β promoter (Fig. 1*b*) in the presence of increasing amounts of DBD2 fusion protein (top panel) or HMG-I(2/3) (bottom panel). The DBD2 fusion protein consists of the B1 IgG-binding domain of Streptococcal protein G and amino acids 50–75 of HMG-I. The equilibrium dissociation constant is given by the ratio of the intercept to slope of plots of K_D(app) versus the concentration of the HMG-I construct. The concentration of the 41mer DNA oligonucleotide was kept constant at 1.3 nM; the concentration of Hoechst 33258 was varied from 5–50 nM for the DBD2 fusion protein titration and from 5–40 nM from the HMG-I(2/3) titration; and the concentration of the DBD2 fusion protein and HMG-I(2/3) were varied from 10–100 nM and from 2–8 nM respectively.



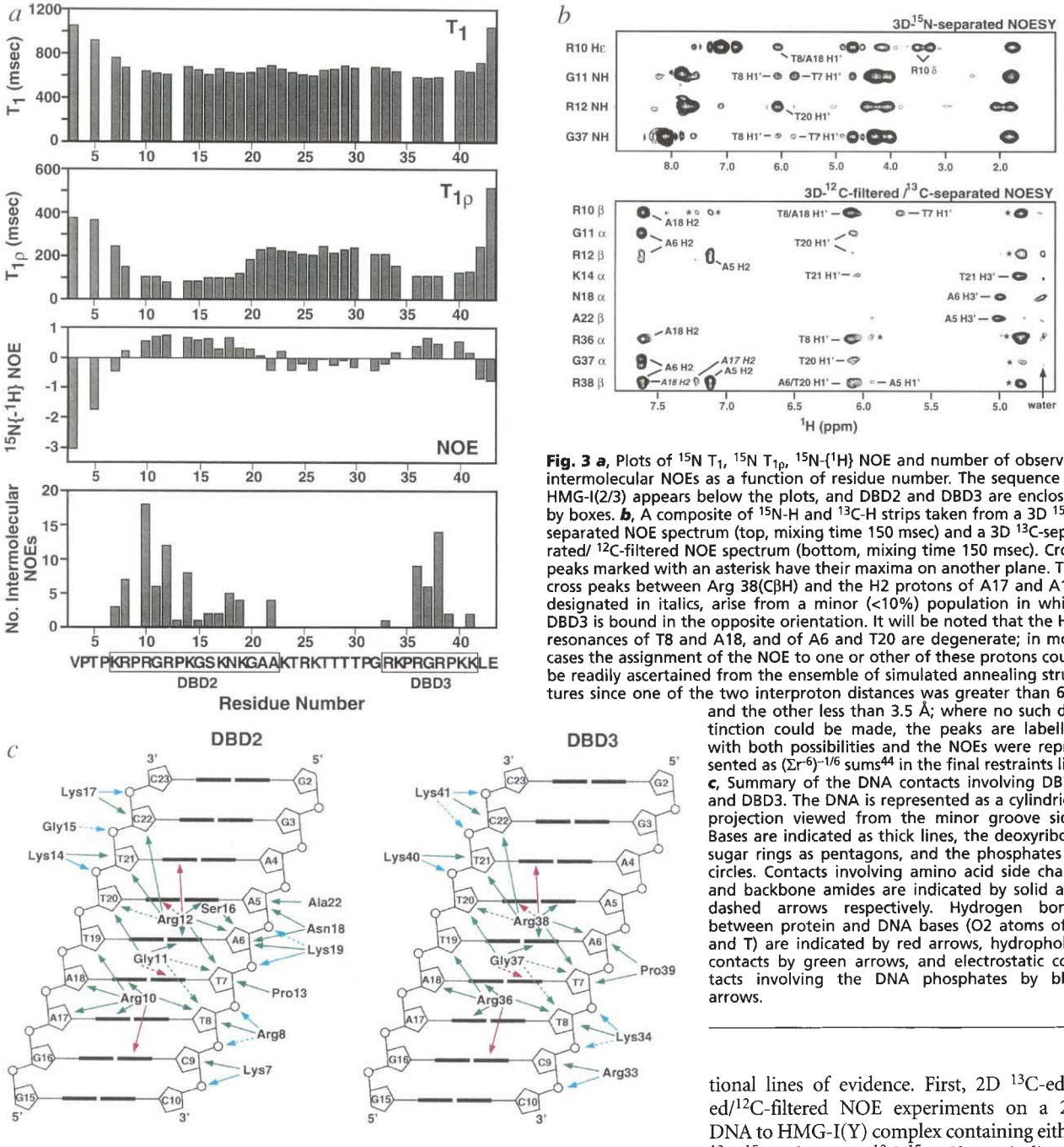


Fig. 3 ***a***, Plots of ^{15}N T_1, ^{15}N $T_{1\rho}$, ^{15}N-{^1H} NOE and number of observed intermolecular NOEs as a function of residue number. The sequence of HMG-I(2/3) appears below the plots, and DBD2 and DBD3 are enclosed by boxes. ***b***, A composite of ^{15}N-H and ^{13}C-H strips taken from a 3D ^{15}N-separated NOE spectrum (top, mixing time 150 msec) and a 3D ^{13}C-separated/ ^{12}C-filtered NOE spectrum (bottom, mixing time 150 msec). Cross peaks marked with an asterisk have their maxima on another plane. The cross peaks between Arg 38(CβH) and the H2 protons of A17 and A18, designated in italics, arise from a minor (<10%) population in which DBD3 is bound in the opposite orientation. It will be noted that the H1' resonances of T8 and A18, and of A6 and T20 are degenerate; in most cases the assignment of the NOE to one or other of these protons could be readily ascertained from the ensemble of simulated annealing structures since one of the two interproton distances was greater than 6 Å and the other less than 3.5 Å; where no such distinction could be made, the peaks are labelled with both possibilities and the NOEs were represented as $(\Sigma r^{-6})^{-1/6}$ sums[44] in the final restraints list. ***c***, Summary of the DNA contacts involving DBD2 and DBD3. The DNA is represented as a cylindrical projection viewed from the minor groove side. Bases are indicated as thick lines, the deoxyribose sugar rings as pentagons, and the phosphates as circles. Contacts involving amino acid side chains and backbone amides are indicated by solid and dashed arrows respectively. Hydrogen bonds between protein and DNA bases (O2 atoms of C and T) are indicated by red arrows, hydrophobic contacts by green arrows, and electrostatic contacts involving the DNA phosphates by blue arrows.

fusion protein to a 41mer oligonucleotide containing the PRDII, PRDIII-I and NRDI elements of the IFN-β promoter (Fig. 1*b*) are ~1.3×10^9 M^{-1} and ~7×10^6 M^{-1} respectively (Fig. 2*b*). At 600 MHz, the largest chemical shift change (Δδ) observed upon DNA binding is ~340 Hz for Gly 11(NH) of DBD2 and ~250 Hz for Arg 38(NH) of DBD3. Hence, the lifetime of the complex bound through DBD2 to DNA is >0.5 ms $[(2\pi\Delta\delta)^{-1}]$ while that for the one bound through DBD3 is <<0.6 ms $[<<(2\pi\Delta\delta)^{-1}]$. To distinguish these DNA-binding motifs, we refer to the high affinity DBD2 as an HMG-I type-I DBD and the low affinity DBD3 as an HMG-I type-II DBD.

The conclusion that one molecule of HMG-I(2/3) binds to two molecules of the PRDII dodecamer was confirmed by several additional lines of evidence. First, 2D ^{13}C-edited/^{12}C-filtered NOE experiments on a 2:1 DNA to HMG-I(Y) complex containing either $^{13}C/^{15}N$-Gly 11 or $^{13}C/^{15}N$-Gly 37 indicated that both residues contact the same DNA bases since NOEs were observed from the CαH protons of Gly 11 and Gly 37 to the H2 and H1' protons of A6. Second, no NOEs between the two DBDs were observed in a 3D ^{13}C-separated NOE spectrum recorded on both 1:1 and 2:1 DNA to protein complexes containing uniformly $^{13}C/^{15}N$ labelled HMG-I(2/3), ruling out a side-by-side orientation at one A,T-tract. Third, analysis of ^{15}N relaxation data[12] for the 2:1 DNA to protein complex (Fig. 3*a*) delineated two distinct ordered domains comprising a longer DBD2 (residues 7–22) and a shorter DBD3 (residues 33–41) separated by a highly flexible linker (residues 22–32). The two domains tumble independently of each other with apparent rotational correlation times of ~9.3 ns for DBD2 and ~7.5 ns for DBD3 (and apparent diffusion

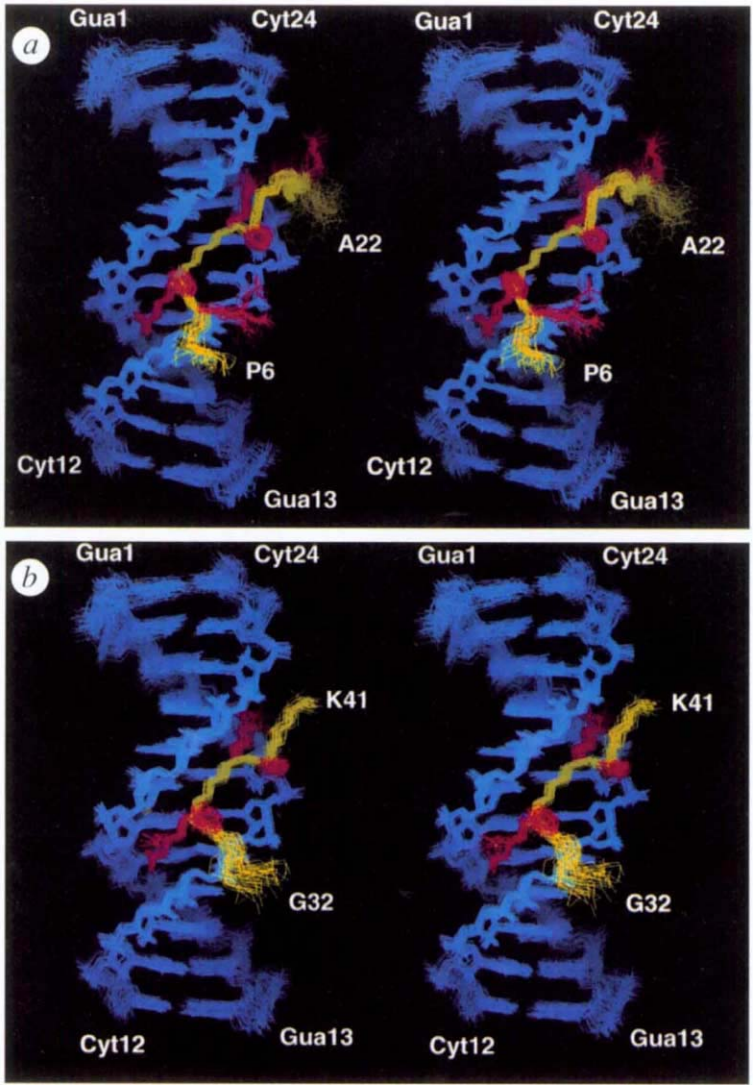


Fig. 4 Stereoviews showing best fit superpositions of the final 35 simulated annealing structures of the ***a***, HMG-I DBD2–DNA and ***b***, DBD3–DNA complexes. The backbone (N, Cα, C) atoms of the DBDs are shown in yellow; side chains of Arg 8, Pro 9, Arg 10, Arg 12, Pro 13, Lys 17 and Asn 18 for DBD2, and Pro 35, Arg 36, Arg 38 and Pro 39 of DBD3 are shown in red; and the DNA is shown in blue. Note that the side chains of Arg 10 and Arg 12 of DBD2 are well defined beyond the Cδ position, principally due to a significant number of intermolecular NOEs involving the NεH and guanidino protons. These NOEs, however, are not observed for the corresponding arginines of DBD3, Arg 36 and Arg 38, which are therefore less well defined.

anisotropies, assuming axial symmetry, of ~1.6–2.0), consistent with a flexible dumb-bell model in which the average angle between the long axes of the two DNA molecules is ~15–25° and the two domains freely diffuse in a cone with a semi-angle of ~50° and a time constant of ~1 ns.

Structure determination

The three-dimensional solution structure of HMG-I(2/3) bound to the PRDII dodecamer in a 2:1 DNA to protein complex was solved using multidimensional heteronuclear-filtered and -edited NMR spectroscopy[13–16]. Only one set of DNA resonances was observed for the 2:1 complex indicating that the chemical environments of the DNA protons in the DBD2-DNA and DBD3-DNA complexes are very similar. Hence, the same set of NMR derived restraints for the DNA was used to calculate the structures of the DBD2- and DBD3-DNA complexes. Since the linker region is highly mobile (Fig. 3*a*) and does not exhibit any non-sequential NOEs, the structures of the two halves (residues 3–27 and 32–41) of HMG-I(2/3) complexed to the PRDII dodecamer were calculated independently. The distribution of intermolecular NOEs and an example of the quality of the NMR data illustrating intermolecular contacts are shown in Fig. 3*a*,*b* respectively, and superpositions of the final 35 simulated annealing structures of the two complexes are shown in Fig. 4. A summary of the structural statistics is provided in Table 1.

Description of the structure

In the absence of DNA, the NMR spectrum of HMG-I(2/3) is indicative of a random coil. Upon binding, the two DBDs that contact the DNA become ordered and adopt a well defined conformation in the minor groove. A summary of the intermolecular contacts is provided in Fig. 3*c*. Both DBDs bind to the 5'-A_4AATT_8 sequence in a specific orientation with the N-terminal Arg of the core Arg-Gly-Arg sequence (Arg 10 of DBD2 and Arg 36 of DBD3) located near base pair 8 and the C-terminal arginine (Arg 12 of DBD2 and Arg 38 of DBD3) located near base pair 4 (Fig. 3*b*,*c*). In the case of the low affinity DBD3 domain, a minor population (<10%) is observed to bind in the reverse orientation, as evidenced by very weak NOEs between Arg 38(CβH) and the H2 protons of A17 and A18; the corresponding NOEs for the major population are much more intense and involve the H2 protons of A5 and A6 (Fig. 3*b*). No evidence of an alternate orientation for the high affinity DBD2 domain could be detected.

The DNA in the DBD2 and DBD3 complexes is essentially B-type (Figs 4, 5). A best-fit superposition of the DNA in the complexes onto classical B and A-DNA yields atomic r.m.s. differences of ~1.7 Å and ~7.1 Å respectively, for base pairs 1–12. For reference, the atomic r.m.s. difference between the mean coordinates of the DNA in the DBD2 and DBD3 complexes is 0.6 Å, which is comparable to the precision of the coordinates. The minor groove width in both complexes is on average only ~1–1.5 Å wider than that in classical B-DNA. The average values for the local helical twist and rise are 35.6° and 3.6 Å respectively. The propeller twist, local inter-base pair tilt angles and local inter-base pair roll angles vary from ~ -26° to ~ +3°, ~ -3° to ~ +2° and ~ -6° to ~ +4° respectively, with average values of ~ -10°, ~ 0° and ~ -1.5° respectively.

The HMG-I DBDs can be divided into three modular components (Figs 1*c*, 5*a*,*b*): a central Arg-Gly-Arg core that adopts an extended conformation deep in the minor groove, a pair of lysine and arginine residues at either end of the core that mediate electrostatic and hydrophobic contacts with the DNA backbone, and in the case of DBD2 (but not DBD3) a more extensive network of six amino acids C-terminal to the core that interacts with the sugar-phosphate backbone on either edge of the minor groove. The latter, which results in a significant increase in the DNA contact surface, is the distinguishing feature between the HMG-I type-I (DBD2) and type-II (DBD3) motifs and accounts for the higher DNA binding affinity of the type-I motif. The existence of

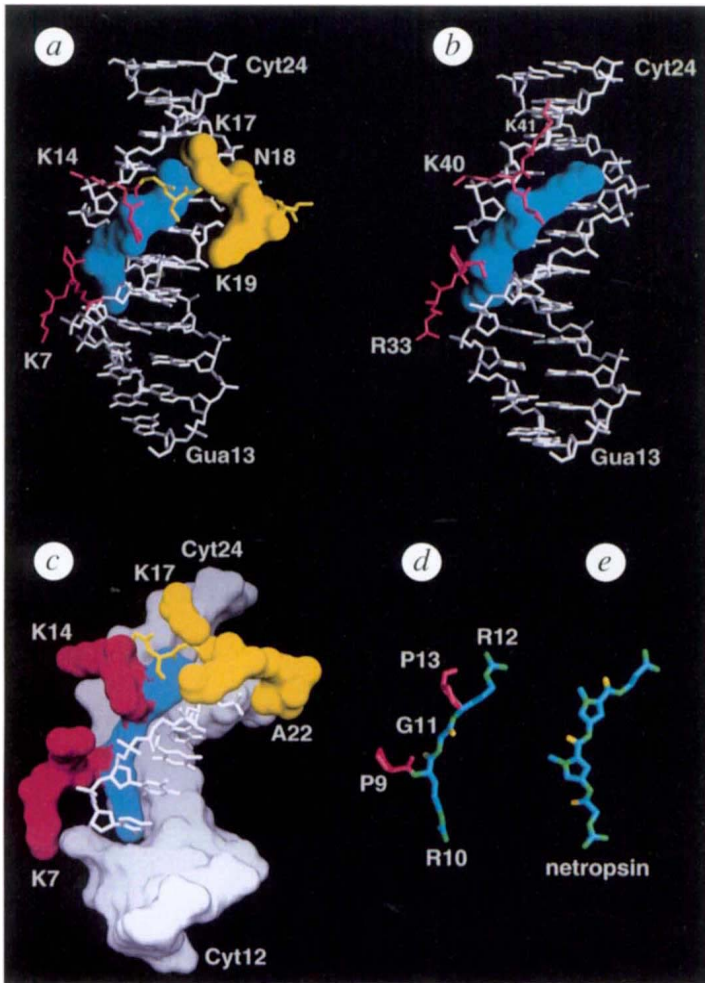


Fig. 5 Surface representations of ***a***, DBD2 and ***b***, DBD3 in which the core Arg-Gly-Arg DNA-binding unit is shown in blue, and the polar network of amino acids that accounts for the higher affinity of DBD2 relative to the DBD3 is shown in yellow. The DNA and the backbone and side chains of the proline and lysine residues immediately adjacent to the core are displayed as a bond representation in white and red respectively. ***c***, View of DBD2 where the DNA and modular components of the type-I motif are colour coded: white for the DNA, blue for the Arg-Gly-Arg core, red for amino acids that contact the phosphate backbone similar to those in the type-II motif, and yellow for the DNA-contacting residues unique to the type-I motif. ***d***, Bond representation of the Arg-Gly-Arg core and adjacent prolines. Carbons are coloured in blue, nitrogens in green, oxygens in yellow, and the prolines in red. ***e***, Bond representation of netropsin with the same colour coding as for panel (*d*). The structures shown are the restrained regularized mean structures. This figure was generated with the program GRASP[45].

an extended Arg-Gly-Arg motif was first proposed on the basis of NMR studies on a series of short peptides, 5–11 residues in length and containing the sequence Pro-Arg-Gly-Arg-Pro, which bound to DNA with an affinity only in the millimolar range[17]. Thus, the additional modules in the HMG-I DBDs are crucial determinants of affinity and specificity since the affinities of DBD3 and DBD2 for DNA are two and four orders of magnitude higher respectively, than those of the short peptides.

The Arg-Gly-Arg core presents a narrow concave surface which is perfectly suited to insert into the minor groove of A,T-tracts without causing a large perturbation in the DNA conformation (Figs 4, 5*a*,*b*). The side chains of the arginine core residues are oriented parallel to the minor groove and extend away from the central A,T base pair (base pair 6). Arg 10 (DBD2)/Arg 36 (DBD3) and Arg 12 (DBD2)/Arg 38 (DBD3) are packed against the bases of A17 and A18, and A5 and A6 respectively, and their guanidino groups are hydrogen bonded to the O2 atoms of C9 and T21 respectively (Figs 3*c*, 6*b*,*d*). The backbone of Arg 12 (DBD2)/Arg 38 (DBD3) is anchored by a hydrogen bond between their respective backbone amides and the O2 atom of T20 (Fig. 6*a*,*c*), and in the case of DBD2 the side chain orientation of Arg 12 may be further stabilized by the interaction of its guanidino group with the Oδ1 atom of Asn 18 (Fig. 6*a*). The conformation of the core arginines is facilitated by the presence of an intervening glycine residue (Gly 11 for DBD2 and Gly 37 for DBD3) whose CαH protons are packed against the base of A6 and whose backbone amide is hydrogen bonded to the O2 atom of T7 (Fig. 6*b*,*d*). This close proximity of the backbone to the bases precludes any other amino acid at this position. Moreover, the nature of these core interactions with the DNA bases excludes a G·C from the central four base pairs. The large 6-NH_2 group of guanine in place of a proton at the equivalent position for A5, A6, A17 or A18 would introduce substantial bulk into the narrow minor groove such that the snug fit of the core arginine and glycine methylene groups would be prevented. Indeed, the absence of the 6-NH_2 group in poly(dI-dC).poly(dI-dC) sequences accounts for their ability to bind HMG-I[18].

Trans prolines on either side of the Arg-Gly-Arg core (Pro 9 and 13 of DBD2, and Pro 35 and 39 of DBD3) direct the peptide backbone away from the minor groove (Figs 5*c*, 6), and position amino acids N- and C-terminal to the core near the phosphate backbone where a pair of lysine and arginine residues N-terminal to Pro 9/Pro 35 and a single lysine C-terminal to Pro 13/Pro 39 interact with the DNA in a similar manner in the two complexes (Figs 3*c*, 6). Thereafter the interactions of DBD2 and DBD3 with the DNA are different. For the weakly binding DBD3 (type-II motif) only one additional residue (Lys 41) contacts the DNA (Figs 3*c*, 5*b*, 6*c*), whereas for the strong binding DBD2 (type-I motif), a network of polar and hydrophobic contacts extends from Gly 15 to Ala 22 (Figs 3*c*, 5*a*, 6*a*). A key component of the latter module is a type II turn between Pro 13 and Ser 16 which positions the side chain of Lys 14 towards the phosphate of T21 on the top strand of the DNA and the backbone amide of Gly 15 in close proximity to the phosphate of C22, and orients the subsequent polypeptide backbone approximately orthogonal to the long axis of the DNA (Fig. 6*a*).

Determinants of binding orientation

One of the principal determinants of binding orientation arises from hydrophobic interactions of the arginine side chains of the Arg-Gly-Arg core with the adenine bases of A5, A6, A17 and A18. With the DBDs bound in the reverse orientation there would be steric clash with the bulky O2 atoms of the complementary thymidines. Further, given the extended conformation of the Arg-Gly-Arg motif with the arginine side chains located on the same side of the polypeptide chain (Fig. 5*d*), optimal van der Waals packing is achieved when the adenine bases contacting the aliphatic portions of the arginine side chains are located on opposite strands of the A,T-tracts (Fig. 6). This suggests that optimal HMG-I binding sites should contain the sequence AA(T/A)T at their centre, a prediction which appears to be borne out from an examination of numerous HMG-I binding sites[1].

Comparison with other minor groove binding motifs

HMG-I(2/3) bound to DNA represents only the fourth example of

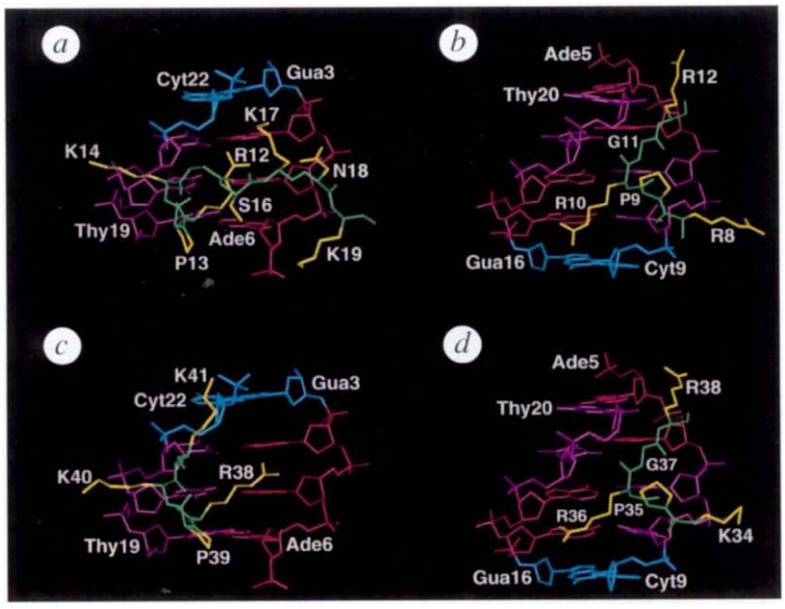


Fig. 6 Views illustrating the interactions of DBD2 and DBD3 at the 5′ (***a*** and ***c*** respectively) and 3′ (***b*** and ***d*** respectively) ends of the A,T-tract. The protein backbone is shown in green, the side chains in yellow, C and G in blue, A in red, and T in magenta. The structures shown are the restrained regularized mean structures. This figure was generated with the program GRASP[45].

a structure in which the protein recognizes DNA exclusively through the minor groove. In the case of the TATA binding protein[19,20], the HMG-1/2 box proteins SRY[21] and LEF[22], and integration host factor[23], a wedge of hydrophobic amino acids, oriented orthogonal to the long axis of the DNA, protrudes from the concave protein surface and partially intercalates between DNA base pairs, causing expansion of the minor groove, and unwinding and bending of the DNA. In contrast, the concave recognition surface of HMG-I is narrow, lacks a protruding hydrophobic amino acid (Fig. 6), and preserves a B-form DNA conformation (Figs. 5*a*,*b*).

HMG-I like motifs in other proteins

A search of the protein database[24] for sequences similar to the HMG-I type-1 or type-2 DNA-binding motifs (Fig. 1*c*,*d*) identified a number of sequence similarities. Factors such as wheat histone H2, modifier 3, SNF2, and CPD1[25] participate in chromatin organization or, in the case of MIF2[26], associate with DNA during mitosis. Other factors, such as HRX, AF17, AF10, EDRF1, and ENBP1, regulate the transcription of specific genes. Some of the chromosomal translocations associated with leukemia involve the HRX and AF17, or HRX and AF10 genes. In the resulting chimaeras, three HMG-I type-II DBDs of HRX effectively replace or add to the HMG-I type-I DBD of AF17 or AF10[6,27]. Although the functional relevance of HMG-I domain switching has not been established, one can postulate that replacement of a tightly binding type-I motif by a weakly binding type-II motif would result in loss of function.

Correlation with biochemical data

The present study demonstrates that two DBDs of HMG-I do not bind side-by-side but bind to sequential A,T-rich sites, in agreement with biochemical results that demonstrated multivalent DNA-binding to the IFN-β promoter[28]. Mutations in the five base pair A,T-tract of the PRDII element have a greater effect on HMG-I binding than mutations in the neighbouring four base pair A,T-tract of the NRDI element[28], suggesting that the PRDII element contains the higher affinity HMG-I binding site. The present results show that DBD2 binds DNA with significantly higher affinity than DBD3, and are consistent with protease protection studies in which only DBD2 was found to be protected in the presence of DNA[10]. At the micromolar concentrations used in these experiments, only a small fraction of the weakly binding type II DBDs (for example DBD3) would be bound. Rapid protease digestion of DBD1 in the presence of DNA and sequence alignments (Fig. 1*d*) suggest that, like DBD3, DBD1 is also an HMG-I type-II DBD. Thus, it seems likely that the DBD2 binds to the high affinity PRDII site, while DBD1 and DBD3 enhance the overall affinity of the complex by binding to the shorter, adjacent NRDI and possibly PRDIII-I A,T tracts (Fig. 1*b*).

It has been observed that the IFN-β enhancer exhibits a small intrinsic bend of ~ -20° towards the minor groove that is reversed upon HMG-I binding[4]. A bend in the direction of the minor groove frequently results in minor groove compression, while a bend in the direction of the major groove would involve minor groove expansion. The minor groove width (~ 7–7.5 Å) of the HMG-I DBD2- and DBD3-DNA complexes is comparable to that of canonical B-DNA (~ 6 Å), and neither DBD could interact optimally with DNA in either a compressed (as in the case of the IFN-β enhancer) or expanded minor groove. Hence, it seems likely that the principal architectural role of HMG-I probably involves reversing and preventing intrinsic distortions in DNA conformation, including bending or kinking, by binding in the minor groove, thereby facilitating specific recognition of the opposing major groove by other transcriptional factors such as NF-κB.

Drug design based on the HMG-I Type-I DBD

Several lines of evidence suggest that HMG-I(Y) is a potential cellular target for the design of anti-neoplastic and anti-HIV drugs. First, overexpression of HMGI-I(Y) proteins is often necessary for transformation of cultured cells, and high expression has been observed in human tumors of the prostate, colon, and thyroid[1]. Second, chromosomal translocations that involve HMG-I(Y) genes are hallmarks of benign tumors of the uterus, breast, and fatty tissue[1]. Third, HMG-I(Y) is an essential cellular component of a functional HIV-1 pre-integration complex[5]. Thus, specific inhibition of DNA binding by HMG-I(Y) may block the development and/or progression of a number of tumors and prevent HIV integration.

The structure of the high affinity HMG-I type-I domain (DBD2) complexed to DNA suggests a means to design novel minor groove DNA-binding drugs that may achieve these therapeutic goals. The Arg-Gly-Arg core mimics the conformation of polypyrrolecarboxyamide drugs such as netropsin bound to DNA[29] (Fig. 5*d*,*e*). Because the polar and hydrophobic network of amino acids (residues 15–22 of DBD2) bridging the minor groove in the HMG-I type I motif (Figs 1*c*, 5*a*,*c*) increases its DNA-binding affinity by about two orders of magnitude relative to that of the HMG-I type-II motif, modifications of polypyrrolecarboxamide drugs to include functional groups that mimic this network may result in significant increases in affinity. This is particularly so if these modifications can be designed to be conformationally

Table 1 Structural statistics[1]

	DBD2 <SA>	DBD3 <SA>
R.m.s. deviations from NOE interproton distance restraints (Å)[2]		
All (624) (501)	0.017±0.003	0.014 ±0.001
Protein		
interresidue sequential (|*i* - *j*| = 1) (71/31)	0.024±0.014	0.00005 ±0.0003
interresidue medium range (1 < |i - j| ≤ 5) (4/2)	0.02±0.027	0.016 ±0.025
intraresidue (64/22)	0.0006±0.007	0.007 ±0.012
DNA		
intraresidue (249/249)	0.0007±0.002	0.0004±0.001
sequential intrastrand (119/119)	0.012±0.004	0.010±0.004
interstrand (44/44)	0.011±0.007	0.011±0.006
Protein-DNA (73/34)[3]	0.031±0.007	0.042±0.008
R.m.s. deviations from DNA H-bond restraints (Å) (42/42)[4]	0.0205±0.008	0.022±0.006
R.m.s. deviations from distance restraints to phosphates (5/4)[5]	0.0008 ±0.004	0.003±0.010
R.m.s. deviations from 'repulsive' restraints (Å) (20/28)[6]	0.025±0.011	0.010±0.008
R.m.s. deviations from experimental dihedral restraints (deg) (172/153)[2]	0.093±0.057	0.190±0.096
R.m.s. deviations from experimental coupling constants		
$^3J_{HN\alpha}$ (Hz) (13/6)[2]	0.58±0.08	0.827±0.046
$^3J_{COCO}$ (Hz) (8/3)[2]	0.38 ±0.11	0.308±0.007
R.m.s. deviations from experimental ^{13}C shifts		
$^{13}C\alpha$ (p.p.m.) (21/9)	0.873±0.082	0.926±0.096
$^{13}C\beta$ (p.p.m.) (18/8)	0.534±0.072	0.511±0.109
Deviations from idealized covalent geometry		
bonds (Å) (1229/999)	0.006±0.0007	0.006±0.0005
angles (°) (2238/1814)	0.863±0.027	0.923±0.017
impropers (°) (603/496)[7]	0.497±0.057	0.511±0.062
E_{L-J} (kcal mol^{-1})[8]	-402±7	-364±7
Coordinate precision (Å)[9]		
Protein backbone plus DNA	0.62±0.12	0.63±0.15
All protein atoms plus DNA	0.72±0.11	0.77±0.14
Protein backbone	0.57±0.17	0.62±0.20
All protein atoms	0.98±0.19	1.25±0.25
DNA	0.58±0.14	0.58±0.16

[1]<SA> is the final set of 35 simulated annealing structures for each complex. $\overline{SA}$ are the mean structures for the two complexes obtained by averaging the coordinates of the individual SA structures best fitted to each other (with respect to residues 6–19 of the protein DBD2 and residues 32–41 of the protein DBD3, and base pairs 1–12 of the DNA). $(\overline{SA})r$ are the restrained regularized mean coordinates for the two complexes obtained by restrained regularized of the mean $\overline{SA}$ coordinates. The number of terms for the various restraints is given in parentheses with the first number referring to the DBD2-DNA complex and the second to the DBD3-DNA complex.

[2]None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, $^3J_{HN\alpha}$ coupling constant violations greater than 2 Hz, or $^3J_{COCO}$ coupling constant violations greater than 0.5 Hz. The torsion angle restraints comprise 36, 17, and 136 torsion angles for the DBD2, DBD3 and DNA respectively. The latter comprise broad torsion angle restraints for the DNA backbone, covering the values characteristic for both A and B-DNA, to prevent problems associated with local mirror images[36].

[3]The number of NOEs involving $(\Sigma r^{-6})^{-1/6}$ sum restraints to multiple DNA protons is 27 for the DBD2–DNA complex and 9 for the DBD3–DNA complex. The effect of such restraints is to permit a close interproton distance contact to be obtained from a proton of the protein to whichever of the protons on the DNA is specified in the restraint, the $(\Sigma r^{-6})^{-1/6}$ sum restraint being satisfied providing only that at least one of these DNA protons is close to the specified proton of the protein[36,44].

[4]The hydrogen bond restraints within the DNA were used to maintain Watson-Crick base pairing.

[5]The Nζ atom of a lysine or guanidino nitrogen atoms of arginines were restrained within 7 Å of a DNA phosphate atom when NOEs from a residue (Lys 7, Arg 8, Lys 14, Lys 17, Lys 19, Arg 33 and Lys 41) or when structure calculations and ^{15}N relaxation measurements (Lys 34 and Lys 40) indicated that that residue interacts with a DNA phosphate. $(\Sigma r^{-6})^{-1/6}$ sum restraints were used and in each case a choice of two adjacent phosphate atoms was given.

[6]In the final stages of the structure calculations two types of 'repulsive' distance restraints[36], with a lower bound of 4 Å (and an unrestrained upper bound), were introduced to facilitate convergence. First, 'repulsive' distance restraints were used to prevent energetically unfavourable proximity of hydrogen bond donors to other donors, and hydrogen bond acceptor groups to other acceptors: there were 11 such restraints for the DBD2–DNA complex and 10 for the DBD3–DNA complex. Second, a small number of 'repulsive' restraints were also used to prevent structures being generated which predicted short interproton distances (<3 Å) for which no corresponding NOEs could be observed: there were 9 such restraints for the DBD2–DNA complex and 18 for the DBD3–DNA complex.

[7]The improper torsion restraints serve to maintain planarity and chirality.

[8]E_{L-J} is the Lennard-Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein and PARNAH1ER1 DNA parameters[46] and is *not* included in the target function for simulated annealing or restrained minimization.

[9]The precision of the coordinates is defined as the average atomic r.m.s. difference between the 35 individual simulated annealing structures of each complex and the mean coordinates $\overline{SA}$. The values refer to residues 6–19 and 32–41 of DBD2 and DBD3 respectively, and to base pairs 1–12 of the DNA.

restricted. Additional increases in both affinity and specificity could be achieved by tethering two or more such drugs together using flexible linkers[30] of variable lengths, in an analogous manner to that employed by intact HMG-I. Using this approach, it should be possible to specifically target A,T-tracts of 5–7 base pairs in length separated by a variable number of base pairs, thereby increasing the selectivity of such compounds towards HMG-I binding sites *in vivo*.

Methods

Sample preparation. The coding sequence for amino acids 50–91 of human HMG-I, referred to in this paper as HMG-I(2/3), was cloned into the *E. coli* vector pET-21a (Novagen) and expressed in host strain BL21(DE3). Uniform (>95%) ^{15}N and ^{13}C labeling was obtained by growing the cells in modified minimal medium containing $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen and carbon sources respectively. The cells were grown at 37 °C, after which protein expression was induced for four hours with 1.0 mM isopropyl-—D-thiogalactoside. The cells were harvested, resuspended in 50 mM Tris buffer, pH 7.5, 5 mM EDTA, and 5 mM benzamidine, lysed by passage through a French press, and cleared by centrifugation. The supernatant was applied to a CM-Sepharose Fast Flow (Pharmacia) column equilibrated with 50 mM Tris, pH 7.2, and HMG-I(2/3) was eluted over a NaCl gradient (0–1.5 M) of five column volumes. Fractions containing HMG-I(2/3) were applied to a C_4 FPLC column equilibrated with 0.05% trifluoroacetic acid and eluted with a 0–100% gradient of acetonitrile. HMG-I(2/3) was further purified by C_4 reversed phase HPLC and lyophilized. The product was characterized by mass spectrometry and found to contain an N-terminal methionine. Based on UV spectroscopy and amino acid analyses, the extinction coefficient of HMGI(2/3) is 38,200 $mol^{-1}cm^{-1}$ at 220 nm. The DBD2 fusion protein was prepared by cloning the gene encoding residues 50–75 of human HMG-I into a pET-21a-derived expression vector that contained the gene for the B1 IgG-binding domain of Streptococcal protein G[31] (J.R.H., C.A.B, G.M.C. and A.M.G., unpublished data). The DBD2 fusion protein was purified by IgG affinity chromatography. The DNA oligonucleotides used for NMR were purchased from Midland Certified Reagent Co. (Texas), purified by anion exchange chromatography, and characterized by mass spectrometry. $^{15}N/^{13}$[Gly11]-HMG-I(2/3) and $^{15}N/^{13}$[Gly37]-HMG-I(2/3) were synthesized on an Applied Biosystems peptide synthesizer and purified by HPLC.

The HMG-I(2/3)-DNA complexes were prepared by dissolving HMG-I(2/3) in 250 µl 10 mM sodium phosphate buffer, pH 5.7. This solution was gradually added to doubled stranded PRDII DNA that was dissolved in 1.2 ml of the same buffer. The solution was concentrated using a Centricon-3 (Amicon) concentrator, and 0.02% (final concentration) sodium azide, 1 mM deuterated EDTA, and 5 µl of a 50% slurry of benzamidine-sepharose beads were added. The pH was adjusted to 6.1. Three samples were prepared that contained a 1:1 ratio of DNA to protein: 1.3 mM ^{15}N HMG-I(2/3)-DNA in 90% H_2O/10% D_2O, 1.3 mM $^{13}C/^{15}N$ HMG-I(2/3)-DNA in 90% H_2O/10% D_2O, and 2.9 mM $^{13}C/^{15}N$ HMGI(2/3)-DNA in 100% D_2O. To make 2:1 DNA to protein complexes, a second equivalent of lyophilized DNA was added to the previously prepared 1:1 samples.

NMR spectroscopy. Spectra for the complexes were recorded at 33 °C on AMX500, AMX600, DMX600, and DMX750 Bruker spectrometers equipped with z-shielded gradient triple resonance probes. Multidimensional experiments were performed as described[13–16]. 3D double and triple resonance through-bond correlation experiments were used to assign the spectra of the protein; 2D ^1H-^1H NOE, ^{12}C-filtered homonuclear Hartmann-Hahn, and ^{12}C-filtered NOE and ROE experiments were used to assign the spectrum of the bound DNA; coupling constants ($^3J_{HN\alpha}$, $^3J_{\alpha\beta}$, $^3J_{C\gamma N}$, $^3J_{C\gamma CO}$, $^3J_{COCO}$) were obtained by 2D and 3D quantitative J correlation spectroscopy[16,32,33]. Intramolecular NOEs within the protein were obtained from 3D ^{15}N-separated (40 and 150 ms mixing time) and ^{13}C-separated (40 and 120 ms mixing times) NOE spectra; intramolecular NOEs within the DNA were obtained from 2D ^1H-^1H NOE (for the imino, amino and H2 protons with a mixing time of 120 ms), ^{12}C-filtered NOE (75 and 100 ms mixing time) and ^{12}C-filtered ROE (35 ms mixing time) spectra; and intermolecular NOEs were obtained from 3D ^{13}C-separated/^{12}C-filtered NOE spectra (80 and 150 ms mixing times), as well as from a 3D ^{15}N-separated NOE spectrum (150 ms mixing time). ^{15}N T_1, $T_{1\rho}$ and ^{15}N-{^1H} NOE experiments were recorded and analyzed as described[12]. Spectra were processed with the NMRPipe package[34], and analyzed using the programs PIPP and STAPP[35].

Structure calculations. Interproton distance and torsion angle restraints were derived from the NOE and coupling constant data as described[36]. The structures were calculated by simulated annealing[37] using the program X-PLOR-31[38] modified to incorporate pseudo-potentials for $^3J_{HN\alpha}$ and $^3J_{COCO}$ coupling constants[39], secondary $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts[40] and a conformational database for both protein and nucleic acids[41,42]. Structural DNA parameters were analyzed using the program CURVES[43].

The coordinates of the 35 final simulated annealing structures of the DBD2–DNA and DBD3–DNA complexes, together with the coordinates of the restrained regularized mean structures, and the complete list of experimental NMR restraints have been deposited in the Brookhaven Protein Data Bank as 2EZD, 2EZE, 2EZF, 2EZG and R2EZEMR.

Acknowledgements

The authors thank D. Garrett and F. Delaglio for software support; R. Tschudin for hardware support; N. Tjandra for help with the analysis of the ^{15}N relaxation data; L. Pannell and P. Lecchi for MS analysis; A. Murphy for amino acid analysis; F. Bushman for communicating results prior to publication; and A. Bax, J. Kuszewski, J. Omichinski, U. Siebenlist, M. Starich, and P. Wingfield for useful discussions. C.A.B. acknowledges a fellowship from the Cancer Research Institute. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.).

Recieved 7 May 1997; accepted 30 May 1997.

1. Bustin, M. & Reeves, R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Progr. Nucl. Acids Res.* **54**, 35-100 (1996).
2. Thanos, D. & Maniatis, T. The high mobility group protein HMG I(Y) is required for NF-κB-dependent virus induction of the human IFN-β gene. *Cell* **71**, 777-789 (1992).
3. Du, W., Thanos, D. & Maniatis, T. Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements. *Cell* **74**, 887-898 (1993).
4. Falvo, J.V., Thanos, D. & Maniatis, T. Reversal of intrinsic DNA bends in the IFN-β gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83**, 1101-1111 (1995).
5. Farnet, C.M. & Bushman, F.D. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **88**, 483 492 (1997).
6. Zeleznik-Le, N.J., Harden, A.M. & Rowley, J.D. 11q23 translocations split the 'AT-hook' cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (*MLL*) gene. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10610-10614 (1994).
7. Zhou, X., Benson, K.F., Ashar, H.R. & Chada, K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* **376**, 771-774 (1995).
8. Ashar, H.R., et al. Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas of distinct transcriptional regulatory domains. *Cell* **82**, 57-65 (1995).
9. Schoenmakers, E.F.P.M., Wanschura, S., Mols, R., Bullerdiek, J., Van den Berghe, H. & Van de Ven, W.J.M. Recurrent rearrangements in the high mobility group protein gene, *HMGI-C*, in benign mesenchymal tumours. *Nature Genet.* **10**, 436-444 (1995).
10. Reeves, R. & Nissen, M.S. The A·T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. *J. Biol. Chem.* **265**, 8573-8582 (1990).
11. Claus, P., Schulze, E. & Wisniewski, J.R. Insect proteins homologous to mammalian high mobility group proteins I/Y (HMG I/Y). *J. Biol. Chem.* **269**, 33042-33048 (1994).
12. Tjandra, N., Wingfield, P., Stahl, S. & Bax, A. Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* **8**, 273-284 (1996).
13. Clore, G.M. & Gronenborn, A.M. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390-1399 (1991).
14. Bax, A. & Grzesiek, S. Methodological advances in protein NMR *Acct. Chem. Res.* **26**, 131-138 (1993).
15. Gronenborn, A.M. & Clore, G.M. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *CRC Crit Rev. Biochem. Mol. Biol.* **30**, 351-385 (1995).
16. Bax, A. et al. Measurement of homo- and hetero-nuclear J couplings from quantitative J correlation. *Meth. Enz.* **239**, 79-106 (1994).
17. Geierstanger, B.H., Volkman, B.F., Kremer, W. & Wemmer, D.E. Short peptide fragments derived from HMG-I(Y) proteins bind specifically to the minor groove of DNA. *Biochemistry* **33**, 5347-5355 (1994).
18. Solomon, M.J., Strauss, F. & Varshavsky, A. A mammalian high mobility group protein recognizes any stretch of six A·T base pairs in duplex DNA. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1276-1280 (1986).
19. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).
20. Kim, J.L., Nikolov, D.B. & Burley, S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527 (1993).
21. Werner, M.H., Huth, J.R., Gronenborn, A.M. & Clore, G.M. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705-714 (1995).
22. Love, J.J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795 (1995).
23. Rice, P.A., Yang, S.-W., Mizuuchi, K. & Nash, H.A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-1306 (1996).
24. Pearson, W.R. & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
25. Levinger, L. & Varshavsky, A. Protein D1 preferentially binds A+T-rich DNA *in vitro* and is a component of *Drosophila melanogaster* nucleosomes containing A+T-rich satellite DNA. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7152-7156 (1982).
26. Meluh, P.B. & Koshland, D. Evidence that the *MIF2* gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere. *Mol. Biol. Cell.* **6**, 793-807 (1997).
27. Prasad, R. *et al.* Leucine-zipper dimerization motif encoded by the *AF17* gene fused to *ALL-1* (*MLL*) in acute leukemias. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8107-8111 (1994).
28. Maher, J.F. & Nathans, D. Multivalent DNA-binding properties of the HMG-I proteins. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6716-6720 (1996).
29. Goodsell, D.S. Kopka, M.L. & Dickerson, R.E. Refinement of netropsin bound to DNA: bias and feedback in electron density map interpretation. *Biochemistry* **34**, 4983-4993 (1995).
30. Mrksich, M., Parks, M.E. & Dervan, P.B. A new class of oligopeptides for sequence-specific recognition in the minor groove of double helical DNA. *J. Amer. Chem. Soc.* **116**, 7983-7988 (1994).
31. Gronenborn, A.M. *et al.* A novel, highly stable fold of the immunoglobulin binding domain of Streptococcal protein G. *Science* **253**, 657-661 (1991).
32. Hu, J.-S. & Bax, A. Measurement of three-bond ^{13}C-^{13}C *J* couplings between carbonyl and carbonyl/carboxyl carbons in isotopically enriched proteins. *J. Am. Chem. Soc.* **118**, 8170-8171 (1996).
33. Hu, J.-S. & Bax, A. Determination of φ and χ_1 angles in proteins from ^{13}C-^{13}C three-bond *J* couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond. *J. Am. Chem. Soc.*, in press (1997).
34. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomolec. NMR* **6**, 277-293 (1995).
35. Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214-220 (1991).
36. Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. & Clore, G.M. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nature Struct. Biol.* **4**, 122-132 (1997).
37. Nilges, M., Clore, G.M. & Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317-324 (1988).
38. Brünger, A.T. X-PLOR Version 3.1: A system for X-ray crystallography and NMR. (Yale University Press, New Haven, Connecticut; 1993).
39. Garrett, D.S. et al. The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson. Series B* **104**, 99-103 (1994).
40. Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. The impact of direct refinement against ^{13}Cα and ^{13}Cβ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Series B* **106**, 92-96 (1995).
41. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Prot Sci.* **5**, 1067-1080 (1996).
42. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.* **125**, 171-177 (1997).
43. Lavery, R. & Skelnar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655-667 (1989).
44. Nilges, M. A calculational strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Prot. Struct. Funct. Genet.* **17**, 297-309 (1993).
45. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Prot. Struct. Funct. Genet.* **11**, 281-296 (1991).
46. Brooks, B.R. CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187-217 (1993).