

Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding

Frederic H.-T.Allain¹, Yi-Meng Yen²,
James E.Masse¹, Peter Schultze¹,
Thorsten Dieckmann¹, Reid C.Johnson^{2,3,4}
and Juli Feigon^{1,3,4}

¹Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90095-1569, ²Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, CA 90095-1737 and ³Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1570, USA

⁴Corresponding authors
e-mail: feigon@mbi.ucla.edu and rcjohnson@mednet.ucla.edu

NHP6A is a chromatin-associated protein from *Saccharomyces cerevisiae* belonging to the HMG1/2 family of non-specific DNA binding proteins. NHP6A has only one HMG DNA binding domain and forms relatively stable complexes with DNA. We have determined the solution structure of NHP6A and constructed an NMR-based model structure of the DNA complex. The free NHP6A folds into an L-shaped three α -helix structure, and contains an unstructured 17 amino acid basic tail N-terminal to the HMG box. Intermolecular NOEs assigned between NHP6A and a 15 bp ¹³C,¹⁵N-labeled DNA duplex containing the SRY recognition sequence have positioned the NHP6A HMG domain onto the minor groove of the DNA at a site that is shifted by 1 bp and in reverse orientation from that found in the SRY–DNA complex. In the model structure of the NHP6A–DNA complex, the N-terminal basic tail is wrapped around the major groove in a manner mimicking the C-terminal tail of LEF1. The DNA in the complex is severely distorted and contains two adjacent kinks where side chains of methionine and phenylalanine that are important for bending are inserted. The NHP6A–DNA model structure provides insight into how this class of architectural DNA binding proteins may select preferential binding sites.

Keywords: chromatin/DNA bending/DNA recognition/HMG box/NMR

Introduction

Non-histone protein 6A (NHP6A) is one of a number of HMG box proteins found in *Saccharomyces cerevisiae*. The HMG box is a conserved domain of ~80 amino acids which mediates DNA binding of many proteins. Proteins which contain the HMG domain are divided into two subfamilies based upon differences in amino acid sequence and specificity of DNA binding (Grosschedl *et al.*, 1994; Bustin and Reeves, 1996). The first class are generally transcription factors that bind to DNA in a sequence-specific fashion, contain one HMG box and are only expressed in a few cell types. Examples include the human sex-determining factor SRY (Gubbay *et al.*, 1990; Sinclair

et al., 1990), the lymphoid enhancer binding factor LEF1 (Travis *et al.*, 1991) and the T-cell factor TCF-1 (van de Wetering *et al.*, 1991).

The second class of HMG box proteins is more abundant, often contain two or more tandem HMG boxes, and bind DNA with little or no sequence specificity. This second class of HMG proteins, typified by vertebrate HMG1 and HMG2, which contain two HMG boxes (Wen *et al.*, 1989), are present at a level of one molecule per two or three nucleosomes and thus are a major constituent of eukaryotic chromatin (Kuehl *et al.*, 1984). Although their biological functions are just beginning to be revealed, they have been shown to participate in reactions as diverse as DNA recombination, repair, activation and repression of transcription as well as nucleosome assembly and disassembly (Paull *et al.*, 1993, 1996; Ge and Roeder, 1994; Shykind *et al.*, 1995; Nightingale *et al.*, 1996; Ura *et al.*, 1996; van Gent *et al.*, 1997). The HMG1/2 proteins strongly distort DNA upon binding and can stabilize bent and supercoiled DNA. This DNA architectural role in facilitating the formation of higher order nucleoprotein complexes is believed to be a critical component of their activity in many of these reactions. In addition, direct interactions with other proteins may be important in some cases (Zwilling *et al.*, 1995; Zappavigna *et al.*, 1996; Jayaraman *et al.*, 1998).

The *S.cerevisiae* HMG box protein NHP6A is present in the nucleus at levels similar to HMG1/2 but contains only one HMG box, which is 45% identical to rat HMG1 Box B and 80% identical to the HMG box of yeast NHP6B (Kolodrubetz and Burgum, 1990). NHP6B is a functional homologue of NHP6A which is present in lower amounts. In common with other non-sequence-specific HMG box proteins, NHP6A binds linear DNA with little sequence specificity, induces a large bend when it binds, and displays higher affinity for distorted DNA structures such as microcircular and cisplatinated DNA (Paull and Johnson, 1995; Yen *et al.*, 1998). NHP6A forms a more stable complex with DNA than does mammalian HMG1/2. NHP6A/B both contain a highly basic amino acid region that precedes the HMG box, and the 16 amino acid N-terminal basic segment of NHP6A has been shown to be essential for high affinity binding and the formation of monomeric DNA complexes (Yen *et al.*, 1998). *In vivo*, neither NHP6A nor NHP6B is essential since $\Delta nhp6A$ and $\Delta nhp6B$ mutants grow normally, but the double mutant grows slowly at 30°C and is non-viable at 38°C (Costigan *et al.*, 1994). The $\Delta nhp6A/B$ mutants display a variety of morphological changes and are defective in activated transcription of a subset of genes (Costigan *et al.*, 1994; Paull *et al.*, 1996). An NHP6A mutant lacking the N-terminal segment is incapable of rescuing growth or co-activating transcription in a $\Delta nhp6A/B$ background, demonstrating the critical importance of this region for its biological activity (Yen *et al.*, 1998).

To date, four structures of the HMG box domain, including three from the non-sequence-specific subfamily: HMG1 Box A (Hardman *et al.*, 1995), HMG1 Box B (Read *et al.*, 1993; Weir *et al.*, 1993), HMG-D (Jones *et al.*, 1994) and one from the sequence-specific subfamily, SOX4 (van Houte *et al.*, 1995), have been determined in the absence of DNA by NMR. Each of these structures has the same general shape, an L-shaped fold of three α -helices. The NMR structures of two sequence-specific HMG protein–DNA complexes, SRY and LEF1, bound to their cognate DNA recognition sequence, have also been determined (Love *et al.*, 1995; Werner *et al.*, 1995b). In these complexes, the minor groove of the DNA is bound to the concave face of the L-shaped protein and is greatly distorted, generating an overall curvature of $\sim 75^\circ$ and 90° for SRY and LEF1, respectively. The DNA is severely underwound, resulting in a widened and shallow minor groove and a highly compressed major groove. In conjunction with the helical underwinding, large positive roll angles are induced by numerous DNA–protein contacts which include a partial insertion of an amino acid side chain into the minor groove of the DNA. Extensive hydrophobic interactions as well as specific hydrogen bonds between the proteins and the DNA sequence mediate the specificity of this class of HMG1/2 proteins. Although these structures provide valuable insight into the interaction of the HMG domain with DNA, they do not fully account for how HMG1/2 proteins can interact with DNA in a non-sequence-specific manner.

In this study, we present the solution structure of the yeast HMG box protein NHP6A. The structure most closely resembles the fold of its *Drosophila* equivalent HMG-D, including a kink in the middle of helix 3 (Jones *et al.*, 1994). The complexes of NHP6A with two different 15 bp DNA duplexes containing the recognition sequence for LEF1 and SRY were also studied by NMR, with the latter complex being investigated most extensively. We present a model for the complex between NHP6A and the DNA^{SRY} based on the structure of the free protein, observed intermolecular interactions and analogy with the LEF1 protein–DNA complex (Love *et al.*, 1995). A series of mutant NHP6A proteins was constructed based on the structure of the free protein and the model structure of the complex, and their DNA binding properties were investigated using gel mobility shift and ligation assays. Taken together, the data reveal both similarities and unexpected differences between the NHP6A–DNA complex and the structures of the sequence-specific HMG box–DNA complexes, and provide insight into DNA target selection by the non-specific HMG box proteins.

Results

Solution structure of NHP6A

NHP6A is a 93 residue protein with the sequence given in Figure 1A. The amino acid sequence of the protein differs from the other HMG domain proteins in having a basic N-terminal tail of ~ 16 amino acids, which has been shown to be required for high affinity binding. At 37°C the free protein appears to be unfolded, based on the narrow chemical shift dispersion in the ^1H - ^{15}N TROSY spectra (Czisch and Boelens, 1998; Salzman *et al.*, 1998) compared with that observed at 20°C (Figure 2A and C).

Therefore, the structure of NHP6A was solved at 20°C (10 mM NaPO_4 pH 5.5, 100 mM NaCl) by multidimensional heteronuclear NMR spectroscopy. Nearly complete ^1H , ^{15}N and ^{13}C resonance assignments for all residues except the N-terminal tail (1–16) were obtained using standard resonance assignment procedures (see Materials and methods). Input restraints and structure statistics for the ensemble of 30 converged structures are given in Table I.

NHP6A (Figures 2 and 3A) is unstructured up to residue 17 (Asp17) and adopts the typical L-shaped fold of an HMG domain protein (Read *et al.*, 1993; Weir *et al.*, 1993; Jones *et al.*, 1994; Hardman *et al.*, 1995; Love *et al.*, 1995; van Houte *et al.*, 1995; Werner *et al.*, 1995b) from Pro18 to Leu92 (Figure 3). The three α -helices comprise 63% of the protein sequence: helix 1 (Ala27–Glu42), helix 2 (Phe48–Lys60) and helix 3 (Thr63–Leu92) (Figure 3C). Helix 2 is straight, while both helices 1 and 3 are kinked in their centers, at Arg36 and Asp77, respectively. These kinks create bends of 40° for helix 1 and 30° for helix 3, respectively. The kink in helix 1 is also found in most other HMG domain proteins, while the kink in helix 3 is only present in NHP6A and HMG-D, and plays an important role in DNA binding (see Discussion).

The L-shaped fold is stabilized by three hydrophobic cores (Figure 3B). Helix 1 and helix 2 are positioned antiparallel to each other and form the short arm of the L-shape. They interact via a hydrophobic core formed by amino acid side chains from helix 1, helix 2 and the loop between helix 1 and helix 2. Part of the N-terminus (Pro18–Ala24) is extended and interacts with helix 3 to form the long arm of the L-shape. These two regions of the protein are packed by hydrophobic interactions between the side chains from residues in the extended N-terminus and in helix 3 (Figure 3B). The two arms of the L are oriented at an angle of $\sim 80^\circ$ to each other via the third hydrophobic core located at the vertex of the L.

In addition to the 35 hydrogen bonds in the protein backbone maintaining the α -helices, several salt bridges and one hydrogen bond contribute to the stability of the structure. There are five possible salt bridges: between Glu42 and Lys58 stabilizing the interaction between helix 1 and helix 2, between Glu57 and either Lys54 or Lys60 stabilizing helix 2, and between Lys67 and Glu71, Lys79 and Glu82, Arg80 and Glu84 stabilizing helix 3. Similar salt bridges are present in the structures of other HMG proteins. In addition, a hydrogen bond between the Tyr28 hydroxyl group and Gly52 carbonyl oxygen is likely to be present considering the distance of $2.6 \pm 0.2 \text{ \AA}$ between the two oxygens over the ensemble of converged structures.

Formation of NHP6A–DNA complexes

In order to study the interaction of NHP6A with DNA and to compare it directly with the interaction of the sequence-specific HMG domain proteins SRY and LEF1 (King and Weiss, 1993; Haqq *et al.*, 1994; Love *et al.*, 1995; Werner *et al.*, 1995a,b), we prepared two complexes with different 15 bp DNA oligonucleotides. One sequence includes the recognition site for the SRY protein (DNA^{SRY}) (King and Weiss, 1993; Haqq *et al.*, 1994) and the second sequence includes the recognition sequence for the LEF1 protein (DNA^{lef}) (Love *et al.*, 1995) (Figure 1B). Complex formation was easily monitored by observing the changes

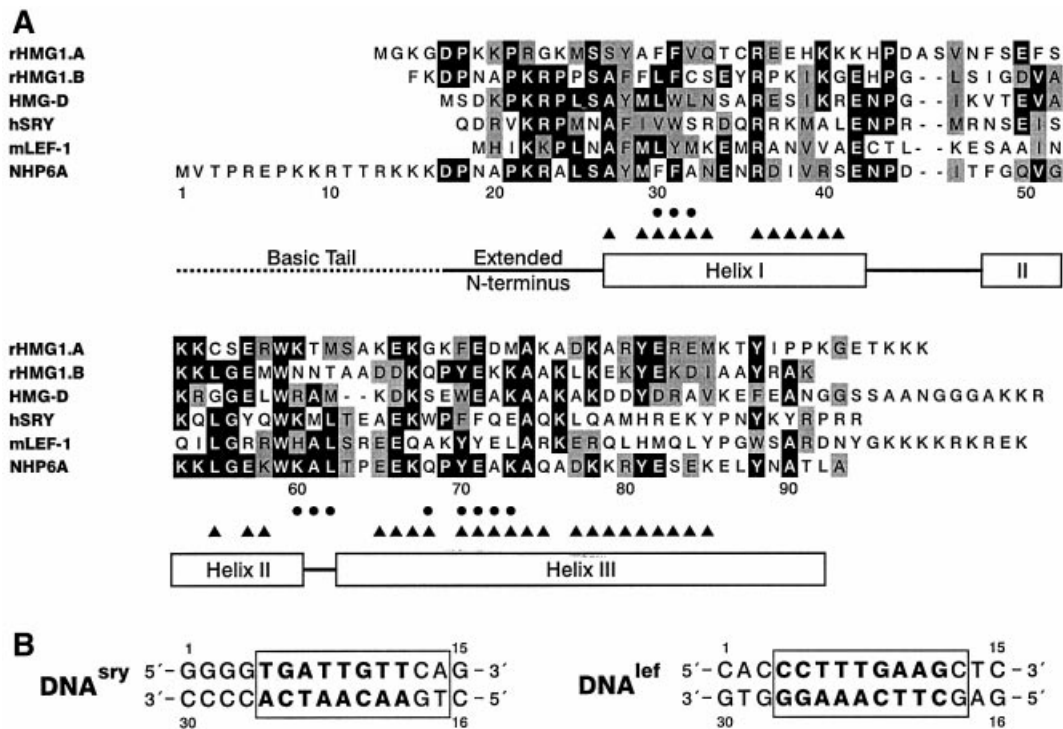


Fig. 1. (A) Amino acid sequence and alignment of rat HMG1A, rat HMB1B, *Drosophila melanogaster* HMG-D, human SRY, mouse LEF1, and *S.cerevisiae* NHP6A, based on the CLUSTAL-W program (Higgins *et al.*, 1996). Identical residues and conserved residues are highlighted in dark gray and light gray, respectively. The location of the three α -helices of NHP6A as defined by PROCHECK (Laskowski *et al.*, 1996) are shown below the sequences. The 34 residues which show a C_{α} chemical shift typical of an α -helix (Wishart and Sykes, 1994) are indicated by filled triangles and the 12 slowly exchanging amides are indicated by filled circles. (B) Sequences of the DNA duplexes used in this study. The SRY (King and Weiss, 1993) and LEF1 (Love *et al.*, 1995) recognition sequences are shown in bold and the binding site of NHP6A on these DNAs, deduced from the NMR studies, are boxed.

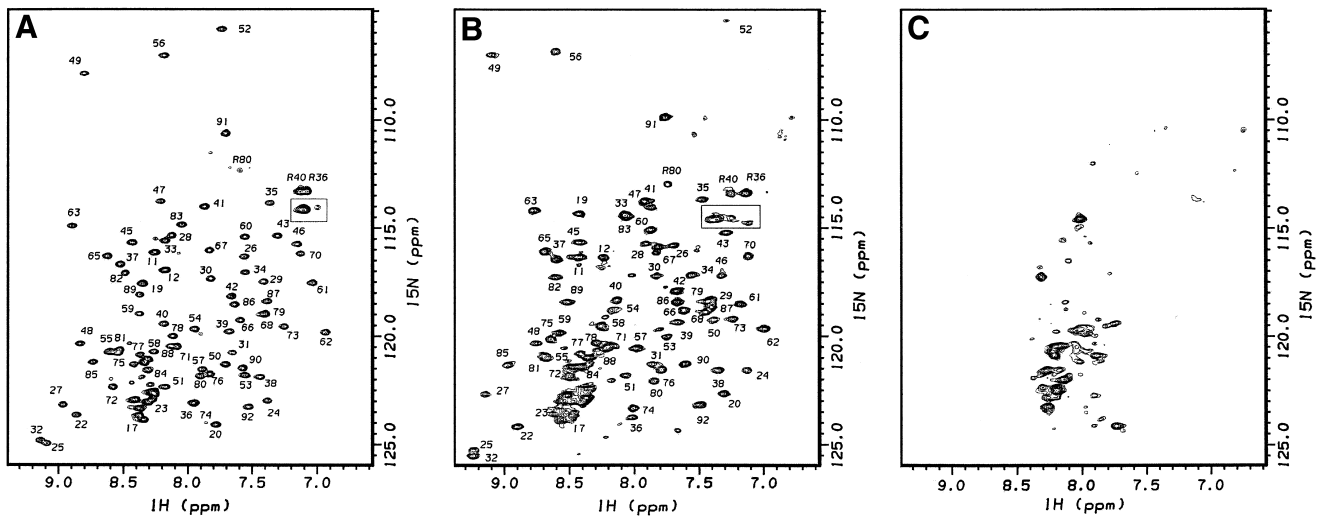


Fig. 2. ^1H - ^{15}N TROSY spectra of (A) NHP6A at 20°C, (B) NHP6A–DNA^{sry} complex at 37°C, and (C) NHP6A at 37°C. The amide ^1H - ^{15}N correlation assignments are indicated by the residue number in (A) and (B). The arginine H_{ϵ} s are indicated as R36, R40 and R80. The boxed region contains the correlations from Arg5, 10, 13 and 23 H_{ϵ} . For all three spectra the ^{15}N carrier was set at 116.5 p.p.m. and spectral widths in F1 and F2 were 1824.8 and 8992.8 Hz, respectively. 175 t1 complex points were acquired in echo anti-echo mode with 16 scans each in the ^{15}N dimension and 1024 complex points in the ^1H dimension. Spectra were processed in 1024 \times 1024 complex points after apodization with a shifted square sinebell.

in the imino proton spectra of the DNA and the appearance of the NH_{ϵ} resonance from Trp59 upon addition of NHP6A. Titration of the DNA by adding NHP6A showed that the DNA in both complexes is in slow exchange on the NMR time scale (spectra not shown) under low salt conditions.

The two NHP6A protein–DNA complexes were very

sensitive to salt concentration and temperature. We found that optimal spectra were obtained for the complexes in low salt (10 mM NaCl, 2 mM NaPO_4 pH 5.5) and at 37°C. We note that NMR studies of the sequence-specific complexes were also done at similar salt concentrations (using KCl) and temperature (Love *et al.*, 1995; Werner *et al.*, 1995b). The ^1H - ^{15}N TROSY spectra of the two

Table I. NMR and structure determination statistics for NHP6A

Input constraints		
Total number of NOE derived distances:		1393
intra residue		241
sequential		373
medium range ($i + 2-i + 4$)		524
long range ($>i + 4$)		255
Number of NOEs per residue (res. 17–93)		18.1
Total number of hydrogen bond constraints		24
Total number of dihedral angle constraints		68
Total number of constraints		1485
Final structure statistics		
Number of refined structures		30
Average of DYANA target function		0.68 ± 0.17
Number of violations >0.2 Å		none
R.m.s.d. to the average structure (Å)		
25–75 bb		0.47 ± 0.23 Å
25–75 sc		0.92 ± 0.21 Å
18–90 bb		1.01 ± 0.48
18–90 sc		1.48 ± 0.47
Comparison between the lowest energy of NHP6A and the previously determined HMG box proteins		
NHP6A (25–75) vs HMG1A (12–31, 34–64)	2.21 Å	Hardman <i>et al.</i> (1995)
NHP6A (25–75) vs HMG1B (11–61)	1.24 Å	Weir <i>et al.</i> (1993)
NHP6A (25–63, 66–75) vs HMG-D (9–57)	1.46 Å	Jones <i>et al.</i> (1994)
NHP6A (25–75) vs LEF1 (6–56)	1.60 Å	Love <i>et al.</i> (1995)
NHP6A (25–75) vs SRY (9–59)	2.74 Å	Werner <i>et al.</i> (1995b)

complexes at 37°C are very similar to one another and to that of the free protein at 20°C, showing that NHP6A adopts essentially the same structure in the two complexes as in its free form at 20°C (Figure 2, NHP6A–DNA^{lef} complex spectrum not shown). The most substantial differences are for the resonances of the N-terminus, which are highly overlapped in the free protein (Figure 2A) and become more resolved and have broader line widths in the complex (Figure 2B).

It is interesting that a stable protein–DNA complex forms at 37°C, since the free protein is apparently largely unfolded at this temperature based on the NMR spectra (Figure 2C) and CD data (Yen *et al.*, 1998). Titration of the DNA into the protein at 37°C also shows formation of the protein–DNA complex (not shown). Furthermore, the same binding affinity as measured by polyacrylamide gel shifts is obtained for protein–DNA complexes formed and run at 37°C as for those at 23°C (data not shown). Therefore, the folded structure of NHP6A is not only stabilized in the complex, it appears that the protein is actually folding on the DNA at 37°C.

Assignments of the protein and DNA in the NHP6A–DNA^{sry} complex

The complex of NHP6A with DNA^{sry} was chosen for more extensive NMR analysis. Despite the general line broadening due to the increase in molecular weight (19 kDa), almost complete resonance assignments of the protein (Figure 2B) and of the DNA in the complex were obtained. Protein assignments were facilitated by the generally small chemical shift differences (maximum of 0.5 p.p.m. for ¹H, i.e. Ala20 HN and Phe48 H_α, 3 p.p.m. for ¹⁵N, i.e. Gln19, and 1.5 p.p.m. for ¹³C, i.e. Phe48 and Met29 C_α) that were observed between the resonances of the free and the bound form and by the presence of the same sequential connectivities. Based on these results, it appears that the protein in the complex adopts almost the

same structure in its free and bound form, with the exception of the N-terminal extension (Lys7–Asp17) which becomes ordered in the complex.

For the extensive NMR study of the NHP6A–DNA complex, uniformly ¹³C,¹⁵N-labeled DNA^{sry} was also enzymatically synthesized (Masse *et al.*, 1998), and protein–DNA complexes with both strands as well as each of the individual labeled strands were prepared and studied. The use of individually labeled strands along with the labeled duplex greatly facilitated the assignments of the DNA in the complex, and more importantly, the assignment of intermolecular NOEs (Masse *et al.*, 1999). The pattern of sequential NOE connectivities and cross-peak intensities seen for the DNA^{sry} is similar to that reported for the sequence-specific SRY–DNA complex and consistent with a bent and underwound DNA structure in the complex (Werner *et al.*, 1995a).

Chemical shift mapping and line broadening of the protein–DNA interface

An analysis of the chemical shift differences (backbone and side chain) between the free and the bound forms of the protein leads to a first ‘footprint’ of the protein surface in contact with the DNA. The regions of the protein showing the largest chemical shift difference between the free and the bound forms of the protein are the N-terminal region (Asp17–Ser26), helix 1 (Ala27–Gln33, Arg36 and Arg40), the whole helix 2, and a few residues of helix 3 (Lys78, Tyr81 and Lys85) (Figure 4A). In addition to the chemical shift changes created by the new chemical environment, the resonances of some residues at the interface show a large line broadening in the complex, probably because of dynamics at the protein–DNA interface. Severe line broadening is observed in particular for the side chain resonances of Thr11, Thr12, Tyr28, Met29, Asn33 (not shown), Thr47, Phe48 (Figure 4B) and the amide resonances of Gly49 and 52 (Figure 1A). A change

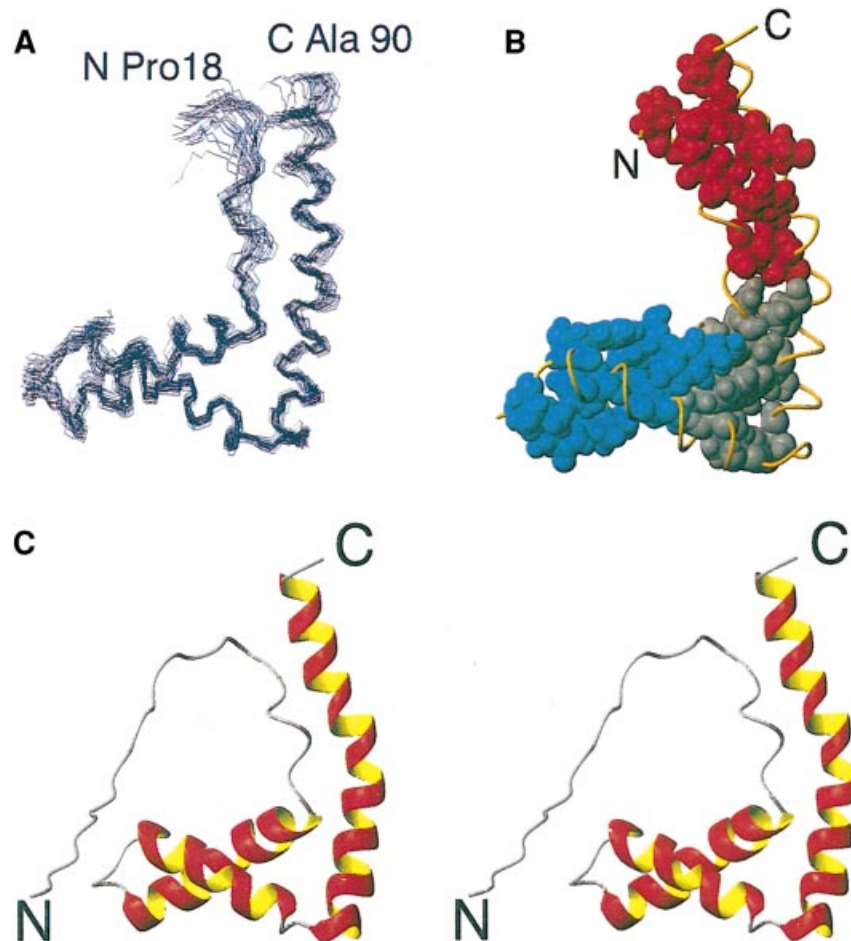


Fig. 3. Three-dimensional structure of NHP6A. (A) Superimposition of the backbone (residues 18–90) of the 30 converged NHP6A structures on the lowest energy structure. The structures are superimposed from residue 18 to residue 90, r.m.s.d. of 1.01 ± 0.48 Å. Residues 1–17 and 91–93 are not displayed because they are less ordered. (B) The three hydrophobic patches creating the fold of the protein. The van der Waals radii of the side-chain atoms packing helix 1 against helix 2 are in blue (Tyr28, Ala32, Asn35, Arg36, Val39 and Arg40 from helix 1, Phe48, Val51 and Leu55 from helix 2 and Asn43 plus Ile46 from the connecting loop). Pro18, Ala20, Pro21 and Ala24 which form the extended N-terminus are packed against Asp77, Tyr81, Lys85, Tyr88 and Thr91 from helix 3 (red). The third hydrophobic core which orients the two arms of the L relative to one another (composed of Ala27, Tyr28, Phe30 and Phe31 from helix 1, Trp59 from helix 2, Glu66, Tyr70 and Ala74 from helix 3 and Val62 from the loop between helix 2 and helix 3) is in gray. The backbone of the protein is shown as a gold ribbon. (C) Stereoview of the backbone ribbon of the lowest energy structure of NHP6A. Helical regions are shown in red/yellow. Note that helix 3 is kinked in its center by 30° . The N- and C-termini are indicated on each structure. This and subsequent structure figures were generated using MOLMOL (Koradi *et al.*, 1996).

in chemical shifts and increase in the line width is also observed for part of the N-terminal region (Lys8–Asp17) in the complex, e.g. the amides of Thr11 and Thr12 and the H_ϵ of Arg5, 10 and 12 (Figure 2A and B). These chemical shift and line width changes are consistent with the N-terminal tail becoming ordered in the complex, probably by interacting with the DNA.

Similarly, chemical shift changes and severe line broadening are observed in some regions of the DNA in the complex (not shown). Resonance broadening is particularly clear for most of the thymine imino protons and most adenine H2s of the DNA in complex. These data map the protein binding site on the DNA to the 11 bp G4–A14/T17–C27, centered at the T9–G10/C21–A22 dinucleotide step (Figure 1B).

Intermolecular NOEs in the protein–DNA complex

Despite the severe line broadening observed for many resonances at the protein–DNA interface, a large number of intermolecular NOEs were observed in 3D double half

filtered HMQC–NOESY spectra (Lee *et al.*, 1994) taken with ^{13}C , ^{15}N -labeled protein and unlabeled DNA as well as with ^{13}C , ^{15}N -labeled DNA (Figure 5). The use of labeled DNA proved essential for confirming the presence of weak cross-peaks between protein and DNA as well as for obtaining the few unambiguous NOE assignments at the protein–DNA interface (Masse *et al.*, 1999). Eight NOEs between Leu25 and the DNA were unambiguously assigned as follows: Leu25 C_βH_3 s to the H1' (weak), H3' (weak), H4', H5' and H5'' of T9 and G10 (Figure 5A and Masse *et al.*, 1999). Because of the severe spectral overlap of arginine and lysine protein side chains and the DNA H4', H5' and H5'', the majority of the intermolecular NOEs could only be assigned qualitatively. For example, Asn33, Tyr28, Phe48, Trp59 and several arginine and lysine side-chains resonances (H_γ and H_δ of Arg, H_γ , H_δ and H_ϵ of Lys) give NOEs to the unresolved deoxyribose H4', H5' and H5'', indicating contact to the DNA backbone from the minor groove of the DNA (Figure 5B). NOE correlations are observed from lysine and arginine side

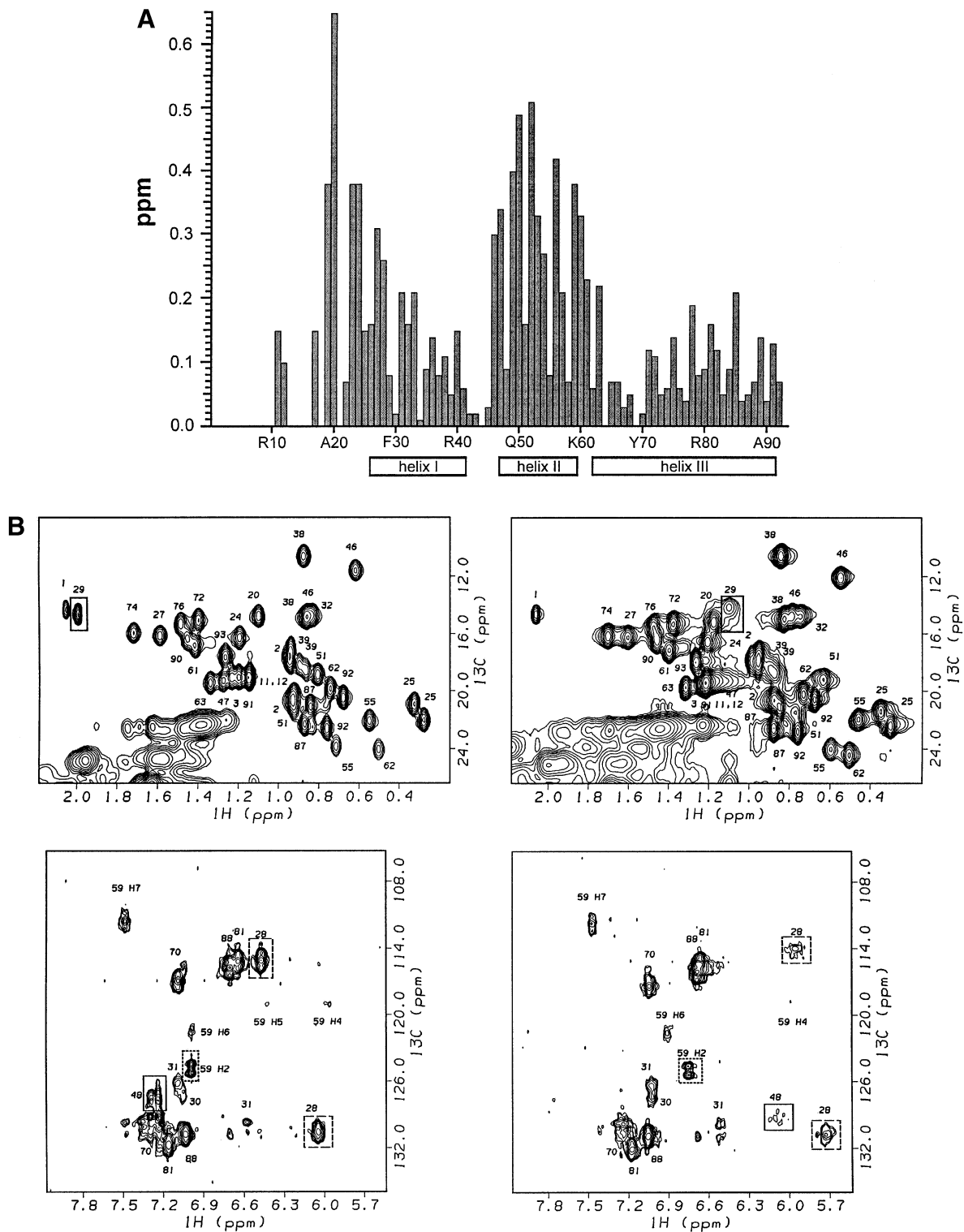


Fig. 4. (A) Plot of the amide ^1H and ^{15}N chemical shift differences between the free and the bound protein. The absolute value of the difference between the ^1H chemical shift of the free (at 20°C , Figure 2A) and bound form (at 37°C , Figure 2B) of NHP6A plus one-tenth of the absolute value of the difference between the ^{15}N chemical shift of the free and bound form of NHP6A (the tenth is taken considering the gyromagnetic ratio of ^{15}N compared with ^1H) is plotted versus residue number (Hardman *et al.*, 1995). (B) Portion of the 600 MHz ^1H - ^{13}C sensitivity-enhanced HSQC spectra of the free protein (20°C , left panels) and the NHP6A-DNA complex (37°C , right panels) with the uniformly ^{13}C , ^{15}N -labeled protein. The top two panels show the ^1H - ^{13}C correlations involving the protein methyls (Leu, Val, Ile, Thr, Ala and Met) and the bottom two panels show the ^1H - ^{13}C correlation involving the aromatics protons (Tyr, Phe and Trp). Residues which have the largest chemical changes between free and bound protein are Met29, Tyr28, Phe48 and Trp59 which are boxed. Residue numbers are indicated. The top two spectra were recorded with the following parameters: ^{13}C carrier at 40 p.p.m., spectral widths in F1 and F2 of 10563.6 and 6009.6 Hz, respectively, and 256 complex point in t1. The bottom two panels were recorded with the ^{13}C carrier at 120 p.p.m., spectral widths in F1 and F2 of 6036.4 and 6009.6 Hz, respectively, and 128 complex points in t1. All spectra were acquired with 12 scans each in the ^{13}C dimension and 1024 complex points in the ^1H dimension and processed in 1024×1024 complex points after apodization with a shifted square sinebell.

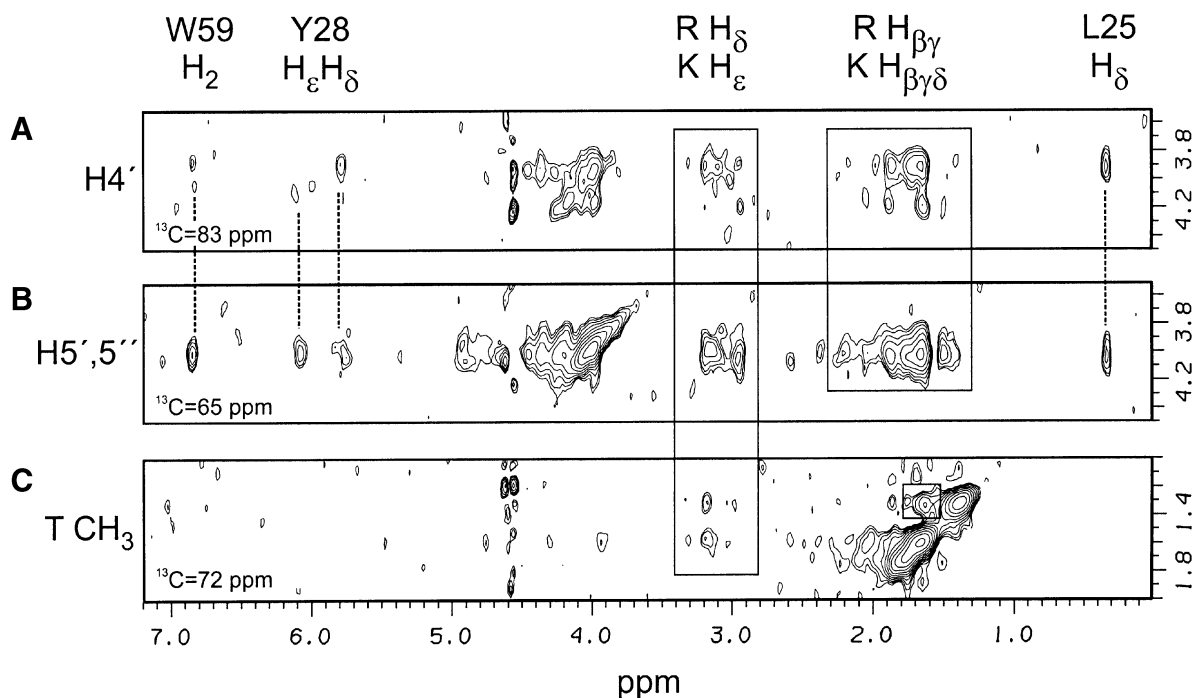


Fig. 5. Slices from ^1H - ^1H planes of a 3D ^1H - ^{13}C double half-filtered HMQC-NOESY recorded on the complex between NHP6A and the ^{13}C , ^{15}N -labeled DNA^{STY} taken at different ^{13}C chemical shifts: (A) 83 p.p.m., showing intermolecular NOEs from DNA C4'-H4', (B) 65 p.p.m., showing intermolecular NOEs from C5'-H5', H5'', and (C) 72 p.p.m., showing intermolecular NOEs edited from thymine methyl (folded in this spectrum). The protein ^1H resonance assignments are labeled on the figure. The 3D spectrum was recorded on a 600 MHz spectrometer on a sample containing 2 mM ^{13}C , ^{15}N -labeled NHP6A bound to ^{13}C , ^{15}N -labeled DNA^{STY} at 37°C in 10 mM NaCl, 2 mM phosphate pH 5.5 in 99.99% D₂O. The spectral widths in F1, F2 and F3 were 6000, 9100 and 5900 Hz, respectively. 168 increments were acquired in t1, 68 in t2, both in States-TPPI mode, with 16 scans and 512 complex points in t3. The spectrum was processed with 512×128×512 complex points after apodization with a shifted squared sinebell.

chains to T8 and T9 methyls (Figure 5C) and the deoxyribose H3' region, indicating that interactions in the major groove are also taking place.

Modeling the NHP6A–DNA complex

Since only a small number of intermolecular NOEs could be assigned unambiguously, we calculated an NMR-based model structure of the protein–DNA complex rather than a high resolution structure. Our model of the NHP6A–DNA complex is based on the free NHP6A structure, the previously published structure of the sequence-specific LEF1–DNA complex (Love *et al.*, 1995), the few assigned intermolecular NOEs and consistency with the partially assigned intermolecular NOEs in the NHP6A–DNA^{STY} complex (Material and methods). We used the LEF1–DNA complex (Love *et al.*, 1995) as the initial template for the model because: (i) the LEF1 protein is bound to a 15 bp DNA with contacts to both its major and minor grooves which is also the case for our NHP6A–DNA^{STY} complex; (ii) the protein backbones of NHP6A (residue 25–75) and of LEF1 (residues 6–56) are the most similar of the complexes (r.m.s.d. 1.6 Å; Table I); and (iii) initial docking of the NHP6A structure on the LEF1–DNA template gave a much better fit than docking of the NHP6A structure onto the less bent SRY–DNA template. The free NHP6A structure was used to model the bound protein since the NMR spectroscopic evidence indicated that only the unstructured N-terminus changes significantly upon DNA binding. The eight unambiguously assigned intermolecular NOEs between Leu25 and the DNA made

it possible to localize the protein precisely on the DNA^{STY} sequence. In the LEF1–DNA complex the residue equivalent to Leu25 (Leu6 in LEF1) contacts the sugar rings of two consecutive adenine residues (A23 and A24) (Love *et al.*, 1995). In the NHP6A complex, Leu25 contacts the sugar rings of T9 and G10 (Figure 5A and Masse *et al.*, 1999). Since the initial modeling resulted in Leu25 of NHP6A being positioned between A23 and A24 on the LEF1 DNA, we converted those nucleotides to T9 and G10 respectively, and then changed the rest of the nucleotides accordingly to correspond to the DNA^{STY} sequence.

The model structure was refined using a series of restrained molecular dynamics steps with added constraints from the free protein structure and between the side chains and the phosphodiester backbone as described in the Materials and methods. After the model structure was calculated, the NMR spectra were examined for predicted NOE cross-peaks based on short interproton distances between NHP6A and the DNA. Based on this analysis, numerous ambiguous intermolecular NOEs could be assigned. For example, Trp59 H2 and Tyr28 Hδ are <5 Å away from A22 H4', H5' and H5'' and Tyr28 Hε is less than 5 Å away from C21 H4', H5' and H5'', consistent with some of the intermolecular NOEs shown in Figure 5B. Based on the model, the intermolecular NOEs from T8 and T9 methyl to Lys and Arg side chains (Figure 5C) could be assigned to Arg10 and Arg13 respectively. The observed intermolecular NOEs from Lys and Arg side chains (Hβ, Hγ, Hδ and Hε) to the DNA H4', H5' and H5'' are also explained by the close contacts between several

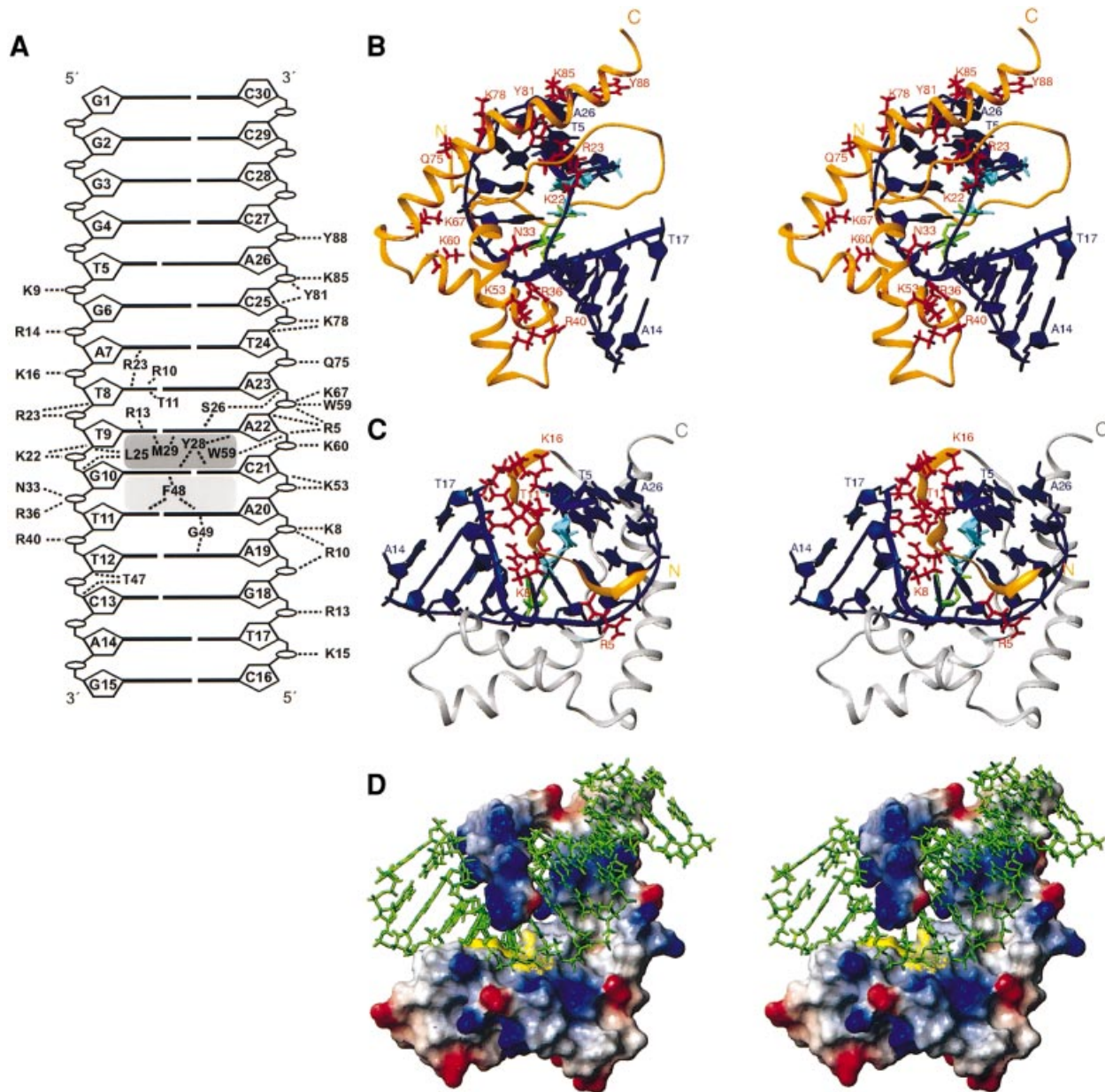


Fig. 6. (A) Schematic representation of the contacts found in the model of NHP6A–DNA complex. Contacts between the protein side chain and the DNA are indicated by dotted lines. Amino acid and DNA nucleotides are labeled with their one-letter codes. The highlighted regions between the T9–G10–T11 steps indicate the location of the kink due to Met29 and Tyr28 in the ‘hydrophobic wedge’ (dark gray) and the second major kink in the DNA due to insertion of Phe48 (light gray). (B and C) Stereo views of the model of the NHP6A–DNA complex. (B) View of the DNA minor groove illustrating the contacts from Lys22, 53, 60, 67, 78 and 85, Tyr81 and 88, Arg23, 36 and 40, Asn33 and Gln75 to the DNA phosphate oxygens and some deoxyriboses. The protein backbone is represented by the gold ribbon and labeled amino acid side chains are red. The DNA is blue. (C) View illustrating the major groove localization and contacts from the N-terminal tail (gold ribbon) (Lys8, Lys9, Arg10, Thr11, Thr12, Arg13, Lys14, Lys15 and Lys16). The protein backbone of the HMG box is represented by a gray ribbon. The DNA is blue, except for T8 and T9 (cyan) and G10 (green). (D) View showing the electrostatic surface potential of NHP6A (GRASP, Nicholls *et al.*, 1991). Blue indicates positive potential and red indicates negative potential. Met29, Phe48 and Tyr28 are colored yellow to show the side chains that insert at the T9–G10–T11 steps in the model structure of the complex. The DNA is displayed in green. Residues 1–7 of the protein are not shown.

Lys (22, 53 and 78) and Arg (5 and 23) side chains and the deoxyriboses in the model structure (Figure 6A).

The model structure is also consistent with other NMR results found in the spectra of the protein and DNA in the complex. The only significantly different chemical shifts in the complex in comparison with the free states were found for most of the residues located at the protein–DNA interface, as expected considering the change of chemical environment. The side chain resonances of Tyr28,

Met29 and Phe48 show the largest chemical shift changes in the complex (Figure 4B), and these residues are intimately associated with DNA bases. In addition, some resonances from protein side chains found in the model to contact the DNA are severely broadened (see Thr11, Thr12, Tyr28, Met29, Thr47 and Phe48 in Figure 4B) possibly because of local motion at the protein–DNA interface. Chemical shift changes for the DNA resonances are also consistent with the location of amino acid side

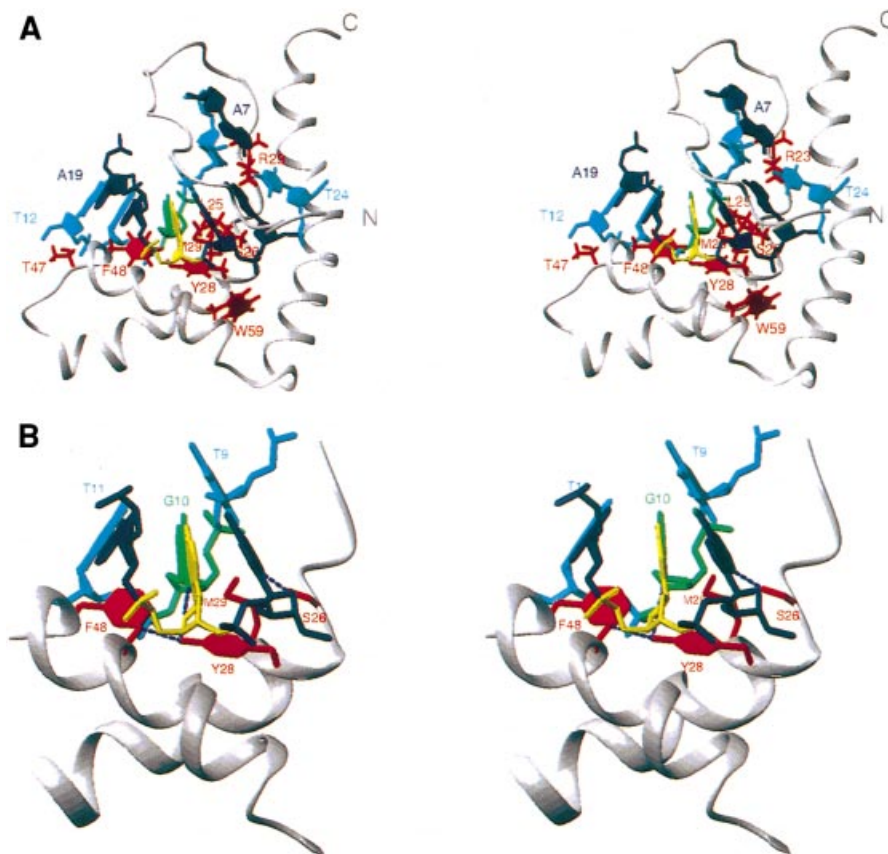


Fig. 7. Stereo views of the model of the NHP6A–DNA^{srY} complex (A) illustrating the hydrophobic contacts from Arg23, Leu25, Ser26, Tyr28, Met29, Thr47, Phe48 and Trp59 (in red) to DNA bases and deoxyriboses in the minor groove (nucleotides T8–T12, A19–A23). Nucleotides are cyan (dT), green (dG), blue (dA) and yellow (dC). The protein backbone is represented by a gray ribbon. (B) Close-up view of potential contacts between NHP6A and DNA^{srY} that may impart specificity for binding of NHP6A at the T9–G10–T11 sequence. Potential hydrogen bonds for Tyr28 hydroxyl–Gly49 carbonyl oxygen, G10 amino–Tyr28 hydroxyl oxygen, and Ser26 hydroxyl–A22 N3 are indicated by dashed lines between heteroatoms.

chains near DNA base pairs in the model structure. For example, there are significant chemical shift changes at the G10–T11 step and only minimal chemical shift changes at the T11–T12 step, consistent with partial insertion of the Phe48 between G10 and T11, as discussed below.

Description of the NHP6A–DNA^{srY} model structure

In the model structure of the complex, NHP6A is bound to the highly distorted and bent DNA through a large number of electrostatic interactions with the phosphate backbone along both the major and minor grooves combined with hydrophobic interactions along the DNA minor groove. Seventeen arginines and lysines plus the side chains of Asn33, Trp59, Gln75, Tyr81 and Tyr88 neutralize the negatively charged non-bridging phosphate oxygens (Figure 6A and B). Helices 2 and 3 follow the DNA backbone of strand 2 (C16–C30), with nine amino acid side chains contacting the minor groove via the non-bridging oxygen O1P of the DNA nucleotides 21–27. A similar set of interactions between five amino acid side chains and the DNA O1P of nucleotides 9–12 is seen in strand 1. The hydrophobic side chains exposed on the concave surface of the protein contact the DNA bases and sugar rings (Figures 6A and 7). Of particular significance are the side chains of Met29 and Phe48, which both protrude out into the DNA binding surface in the free protein structure and are found inserted between adjacent base pairs in the model of the complex (Figure 7). Met29

is inserted between T9 and G10 where a large kink is present. The aromatic ring of Phe48 is inserted almost perpendicular to the minor groove edge of G10•C21 and T11•A20 and appears to be responsible for a second large kink between these base pairs, which occurred during the refinement of the model structure (Figures 6A and 7). Only two side chains are in a position to form hydrogen bonds with bases within the minor groove. The Ser26 OH is close to the N3 of A22, and the Tyr28 OH is positioned such that a network of hydrogen bonds involving the N2 of G10 and the carbonyl oxygen of Gly52 is possible (Figure 7B). As elaborated in the Discussion, these contacts may impart some specificity to the choice of binding site.

The N-terminal tail of NHP6A is in the major groove (Figure 6C and D), with the eight arginines and lysines interacting with the O2P along the backbone of both strands. In addition, there are possible hydrophobic interactions from Arg10, Arg13 and Thr11 side chains to the DNA major groove side of T8 and T9.

Mutant NHP6A proteins with amino acid substitutions on the DNA binding surface

A series of mutations were introduced into NHP6A to assess the functional importance of the different amino acid contacts observed in the model structure. The lysines and arginines within the HMG core domain at residues 22, 23, 36, 40, 53, 60, 67, 78 and 85, as well as Asn33, Tyr81 and Tyr88, which are all in a position to contact the

Table II. DNA binding of NHP6A mutants *in vitro*

NHP6A protein	K_D for linear DNA (nM) ^a
Wild type	10 ± 2
P18A ^b	20 ± 2
K22A	33 ± 8
R23A	28 ± 9
L25D	42 ± 4
Y28A	43 ± 4
M29A ^b	21 ± 3
N33A	33 ± 5
R36A	39 ± 9
R40A	32 ± 8
F48A	19 ± 2
K53A	22 ± 3
K54A	10 ± 1
K58A	10 ± 2
K60A	19 ± 1
K67A	15 ± 2
A74D ^c	99 ± 4
K78A	20 ± 3
Y81A	14 ± 1
K85A	14 ± 2
Y88A	18 ± 3
R23A/R36A ^d	105 ± 16
K53A/K60A ^d	30 ± 4

^a K_D values were calculated from gel electrophoresis experiments as described in the Materials and methods. All of the mutants containing single amino acid changes formed discrete complexes with DNA, and the migrations of each of the mutant NHP6A–DNA complexes were the same as wild type except for Y28A, M29A and F48A (Figure 8), which displayed a faster mobility. Microcircle ligation assays on 98 bp fragments were performed on each of the single mutants using amounts of protein that compensated for any decreased binding. All of the proteins, with the exception of Y28A, F48A and A74D, efficiently formed monomeric circles. M29A is deficient in forming 75 bp microcircles (Yen *et al.*, 1998).

^bMutants described in Yen *et al.* (1998) and re-assayed under the present binding conditions.

^cThis mutant is unfolded in solution (Figure 8C).

^d K_D values for these double mutants are estimated since they were defective in forming individual protein–DNA complexes.

phosphate backbone (Figure 6A and B) were individually substituted with alanine. These mutants all displayed a 1.4- to 4-fold reduction in DNA binding as measured by gel mobility shift assays using a 98 bp DNA fragment (Table II). While some of the effects are small, they are statistically significant and reproducible. Loss of two of these contacts, as illustrated by the double mutants R23A/R36A and K53A/K60A, resulted in a loss of formation of discrete DNA complexes and up to a 10-fold reduction in binding affinity (Figure 8A and Table II). On the other hand, alanine substitutions of K54 and K58, whose side chains are directed away from DNA, have no detectable effect on binding. An aspartic acid substitution of Leu25, whose side chain is in close proximity with the T9 and G10 riboses within the minor groove as directly established by the intermolecular NOEs, resulted in a 4-fold reduction in binding affinity.

Mutations of two aromatic amino acids, Phe48 and Tyr28, along with Met29 (Yen *et al.*, 1998), which are located at the T9–G10–T11 region that is strongly deformed in the NHP6A–DNA^{SY} model structure result in proteins that have altered DNA bending properties. As shown in Figure 8A, the electrophoretic mobilities of the complexes formed with F48A are significantly faster than those produced by the wild type or most of the other

mutants. In addition, microcircle formation by F48A is impaired. Over 20 times more F48A than wild-type protein is required to generate maximum yields of 98 bp DNA fragments into circles (Figure 8B), even though the binding affinity is only reduced ~2-fold (Figure 8A, Table II). Moreover, maximum yields of circles produced with F48A is only ~40% of that formed by the wild-type protein. Because the Phe48 ring is packed against the central hydrophobic core of NHP6A, we were concerned that the alanine substitution may have affected the structure of the protein. CD analysis of the free protein, however, revealed no significant deviation from wild type (Figure 8C).

Y28A behaves similarly to F48A with respect to the faster mobilities of DNA bound complexes and inefficient microcircle formation (data not shown). In this case, the CD profile showed a demonstrable reduction in α -helical content, suggesting that the structure of Y28A is partially disrupted at 25°C (Figure 8C). The 4-fold decrease in DNA binding affinity by Y28A may therefore be a consequence, at least in part, of a folding defect. However, the faster migration of Y28A–DNA complexes and the poor microcircle formation even at high protein to DNA ratios suggest that loss of this side chain directly or indirectly alters DNA structure within the complex. In contrast to Y28A, the A74D mutant gave normal migrating DNA complexes upon gel electrophoresis and was almost as efficient as wild type in microcircle formation at high protein concentrations, even though its CD profile indicated a highly unfolded protein (Figure 8C), and it displayed a 10-fold reduction in binding affinity.

Discussion

Comparison of the NHP6A protein structure and other HMG domain proteins

We have determined the solution structure of NHP6A by multidimensional NMR spectroscopy (Figure 3). NHP6A adopts the L-shaped fold common to proteins of the HMG1/2 class. Among the HMG domain proteins whose structures have been determined, NHP6A is most similar to HMG-D (Jones *et al.*, 1994), HMG1B (Read *et al.*, 1993; Weir *et al.*, 1993) and LEF1 (Love *et al.*, 1995) (Table I). These four proteins all have a bend in helix 1 and an ~80° angle between the two arms of the L, differing slightly from SRY (Werner *et al.*, 1995b), SOX4 (van Houte *et al.*, 1995) and HMG1A (Hardman *et al.*, 1995). Helix 3 from NHP6A is seven residues longer than helix 3 from LEF1. The increased length of helix 3 in all of the non-sequence-specific HMG proteins reflects the absence of a proline at the C-terminal end of the helix. NHP6A, HMG1B and SRY each contain a kink at a proline near the N-terminal end of helix 3. Of more significance, however, is a pronounced bend in the center of helix 3 in NHP6A, which is also present in HMG-D but not in other HMG proteins. In NHP6A, this 30° kink is stabilized by a hydrophobic contact between Ala24 and Asp77 and a salt bridge ($i, i + 3$) between Lys79 and Glu82. In contrast, a hydrophobic interaction between a Pro (for Ala24) and a Leu (for Asp77) and a salt bridge ($i, i + 4$) (the equivalent residue of Lys79 is a Glu in HMG1B) stabilizes the straight helix in HMG1B. As illustrated in Figure 6B and discussed further below, the central kink in helix 3 of NHP6A plus additional curvature extending towards

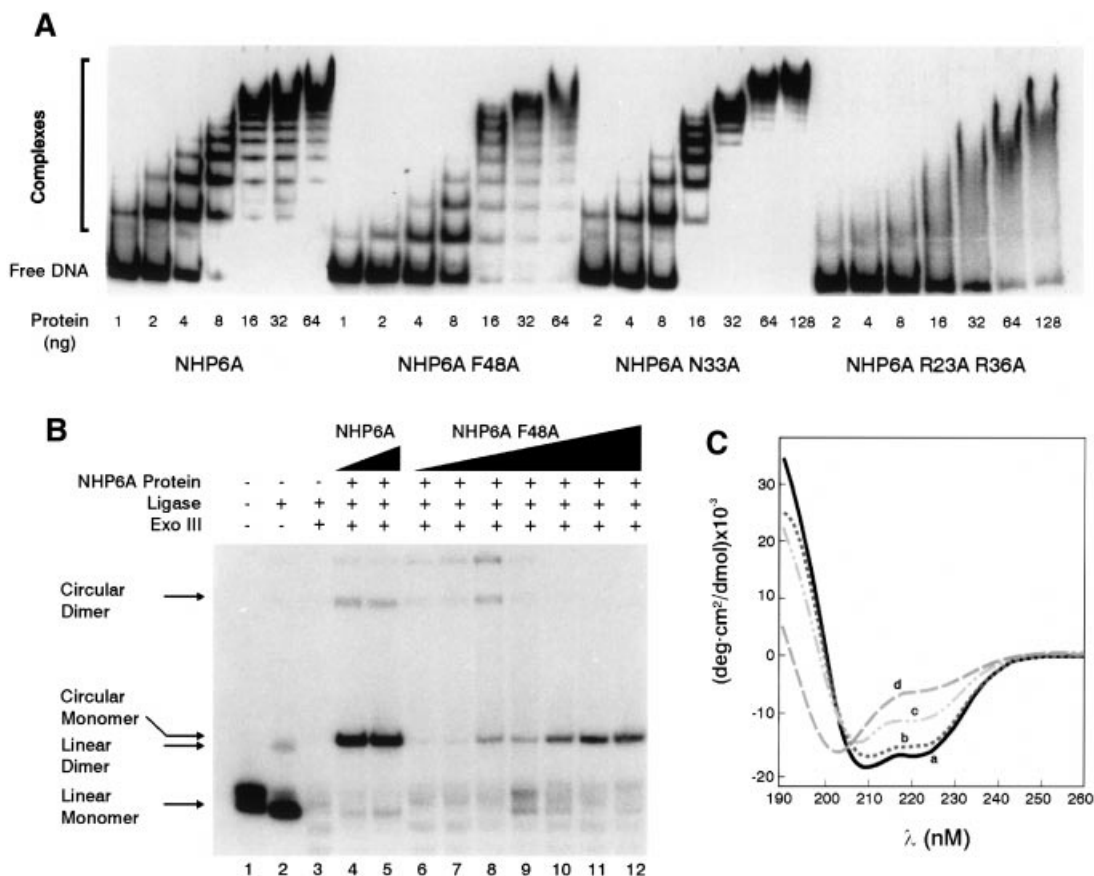


Fig. 8. (A) Gel mobility shift assays on wild-type and mutant NHP6A proteins. A ³²P-labeled 98 bp linear DNA fragment was incubated in 20 μl of buffer alone or with 2-fold increasing amounts of NHP6A wild-type, F48A, R40A or R23A R36A mutant proteins as denoted. (B) Microcircle formation by NHP6A wild type and F48A. The same 98 bp fragment with *Eco*RI ends was incubated with buffer alone (lane 1), T4 DNA ligase (lanes 2–12), and wild-type NHP6A (lanes 4–5) or NHP6A F48A (lanes 6–12). DNA–protein molar ratios were 80:1 and 160:1 for NHP6A wild type and 40:1, 80:1, 160:1, 320:1, 640:1 and 1280:1 for F48A. Exonuclease III was added to the reactions in lanes 3–12 so that the products which remain are circular species only. (C) Far UV CD spectra at 25°C of (a) wild type, (b) F48A, (c) Y28A and (d) A74D.

the C-terminus is important in positioning this helix together with the extended N-terminus along the minor groove of the DNA in the complex.

NHP6A–DNA complex

Chemical shift mapping of the protein and DNA surfaces was initially obtained after resonance assignments of the protein and DNA in the complex and free forms were completed. The protein ‘footprint’ (Figure 4A) shows that the concave face of the L-shaped NHP6A interacts with the DNA, consistent with the partial chemical shift mapping previously reported for the non-sequence-specific HMG box HMG1A bound to DNA (Hardman *et al.*, 1995). The DNA ‘footprint’ for both the SRY and LEF1 sequences show that the binding site for NHP6A extends over 11 bp (Figure 1B), which corresponds to T5–A14 within the DNA^{SRY} sequence (Figure 6A), and that the primary interface is along the minor groove of DNA. A similarly sized DNA site was deduced for HMG-D based on changes in the NMR spectra of the DNA upon binding (Churchill *et al.*, 1995). The use of isotopically labeled DNA in combination with isotopically labeled protein enabled a limited number of unambiguous NOEs between NHP6A and the DNA^{SRY} sequence to be assigned, which made it possible to precisely position the protein onto the DNA. The

structure of the DNA in the LEF1 complex provided a remarkably good initial fit when docked to the NHP6A surface at the location defined by the intermolecular NOEs, although a few clashes were present between the DNA and amino acids within the extended N-terminus. In contrast, the less bent DNA from the SRY complex did not adequately fit onto the NHP6A structure. The refined model structure of the NHP6A–DNA complex is completely consistent with all of the NMR as well as biochemical and mutational analysis generated to date.

DNA minor groove interactions by the HMG domain of NHP6A

As discussed above, the binding surface of NHP6A conforms well to the highly distorted DNA in the LEF1–DNA complex, with only small changes in the DNA structure resulting from the model structure calculations. The DNA has an overall curvature of ~90°, which is slightly lower than the value estimated from microcircle ligation and binding studies (Yen *et al.*, 1998). The DNA binding face of the HMG domain of NHP6A presents a hydrophobic surface which is associated with the interior of the wide and shallow minor groove and is flanked by a series of basic residues that are interacting with the DNA phosphodiester backbone (Figure 6B and D). While

there is striking complementarity between the protein and DNA surfaces, there is little opportunity for hydrogen bonding to the bases in the NHP6A–DNA^{sty} model, unlike the sequence-specific LEF1– and SRY–DNA structures. One strand of the phosphodiester backbone between C21 and C27 actually resides within a cleft in the protein surface and is stabilized by sequential salt bridges or hydrogen bonds involving Lys53, Trp59, Lys60, Lys67, Gln75, Lys78, Tyr81, Lys85 and Tyr88. Elimination of any of these contacts by individual alanine replacement (with the exception of Trp59 and Gln75 which were not tested) leads to a modest reduction in binding affinity (Table II). The predicted contacts from Lys78, Tyr81, Lys85 and Tyr88 from the C-terminal half of helix III to the DNA are unique to NHP6A, since the equivalent residues in SRY and LEF1 do not contact the DNA. The five sequential salt bridges or hydrogen bonds involving Lys22, Arg23, Asn33, Arg36 and Arg40 on the other phosphodiester strand between T8 and T12 may play a more important role since individual alanine substitutions at these positions have a more detrimental effect on binding. Loss of two of these contacts, represented by the double mutant R23A/R36A, reduces DNA binding 10-fold.

The largely hydrophobic DNA binding surface on NHP6A extends from Arg23 to Thr47 (Figures 6A and 7A). Several of these residues are intimately associated with the DNA near the center of the binding site at T9–G10–T11 and are responsible for the two large kinks within this region (Figure 7). The side chain of Met29 is inserted between T9 and G10. Met29 corresponds to Ile68 in the SRY and Met10 in the LEF1 HMG domains, which also insert between bases and are responsible for large kinks. M29A is defective in forming 66 and 75 bp microcircles and NHP6A M29A–DNA complexes show a slightly greater mobility than wild type, consistent with a compromised ability to induce severe bends in DNA (Yen *et al.*, 1998). M29A functions poorly as a transcriptional co-activator and only partly rescues the slow growth phenotype of $\Delta nhp6A/B$ mutants. The SRY mutant I68T binds and bends DNA very poorly, leading to a sex reversal phenotype in humans (Haqq *et al.*, 1994; Peters *et al.*, 1995).

The aromatic ring of Tyr28 is closely packed against the bases within the T9–G10 step and thus is likely to stabilize the kink. Consistent with this, Y28A is defective in DNA bending as measured by its poor ligation efficiency and faster electrophoretic migration of Y28A–DNA complexes as compared with wild type. The position of the aromatic ring of the Tyr28 is stabilized by a hydrogen bond between the side chain OH and the main chain O of Gly52. The hydroxyl of Tyr28 can also act as a hydrogen bond acceptor for the amino group of G10 (Figure 7), therefore selecting for a G at this position. The only other potential hydrogen bond within this region involves Ser26 hydroxyl to A22 N3 (base paired to T9), which is not expected to impart sequence specificity since an equivalent hydrogen bond could form with any base in this position. Both a tyrosine and a serine at these locations are highly conserved among the non-sequence-specific HMG proteins, as opposed to the sequence-specific class, which contains a Phe and an Asn, respectively (Figure 1A).

The aromatic side chain of Phe48 is partially stacked

against A20 and abuts G10 in the NHP6A–DNA^{sty} model structure (Figures 6A and 7) and thus causes a second severe kink in a manner that resembles insertions of the phenylalanines of TBP between T–A steps on either side of the TATA sequence (Kim *et al.*, 1993a,b). A similarly located kink is not present in the SRY or LEF1 DNA complexes, which have polar amino acids in their analogous positions which are hydrogen bonded to bases. Loss of Phe48 results in only a 2-fold reduction in binding affinity, but ligation and gel mobility experiments indicate that F48A–DNA complexes are less curved. The kink at the G10–T11 step induced by Phe48 in NHP6A may be a general property of the non-sequence-specific subclass of HMG proteins since these all tend to have an aromatic or branched chain amino acid at this position. A similar role for the equivalent amino acid in HMG-D was proposed based on a molecular dynamics simulation of HMG-D with the bent DNA of the TBP complex (Balaeff *et al.*, 1998).

The basic N-terminal tail of NHP6A binds in the major groove of the DNA

NHP6A differs from most HMG domain proteins in having a highly basic extension at its N-terminus. We have previously shown that the N-terminal basic tail is essential for the formation of stable NHP6A–DNA complexes (Yen *et al.*, 1998). This region, which is unstructured in the free protein, contains two basic patches of amino acids beginning at residue 8, with the segment between Arg13 and Lys16 being most important for DNA binding. In the model of the NHP6A–DNA^{sty} complex, the N-terminal basic tail crosses the phosphodiester backbone and is inserted into the compressed major groove of the DNA (Figure 6C and D), thereby accounting for a series of intermolecular NOEs on the major groove side. The N-terminal segment effectively wraps around the opposite side of the bound DNA from the expanded minor groove where the body of the protein is associated, and serves to clamp the protein onto the DNA. Every residue from Lys8 to Lys16 is positioned to contact the two phosphodiester strands in an alternating manner, thus complementing similar contacts from the minor groove side and stabilizing the bend towards the major groove. These contacts account for the critical importance of the basic extension at the N-terminus in promoting the relatively strong DNA binding by NHP6A in comparison with many other non-sequence-specific HMG proteins.

The N-terminal basic region of NHP6A in the model structurally and functionally mimics the C-terminal basic region of LEF1 (Figure 9). The four basic residues closest to the HMG domain (Arg13–Lys16 for NHP6A and Lys78–Arg81 for LEF1) are the most critical for high affinity DNA binding (Yen *et al.*, 1998). The basic tail of LEF1 is directed towards the major groove by a kink introduced at the C-terminal end of helix III by Pro67. In NHP6A, a kink at Pro18 at the very beginning of the extended N-terminus of the HMG domain directs the peptide chain into the major groove. A proline is present at the equivalent position in HMG1A and HMG1B and may play a similar role in orienting the N-termini of these two HMG domains towards the DNA major groove. We have shown previously that an alanine substitution at Pro18 reduces DNA binding without affecting the structure

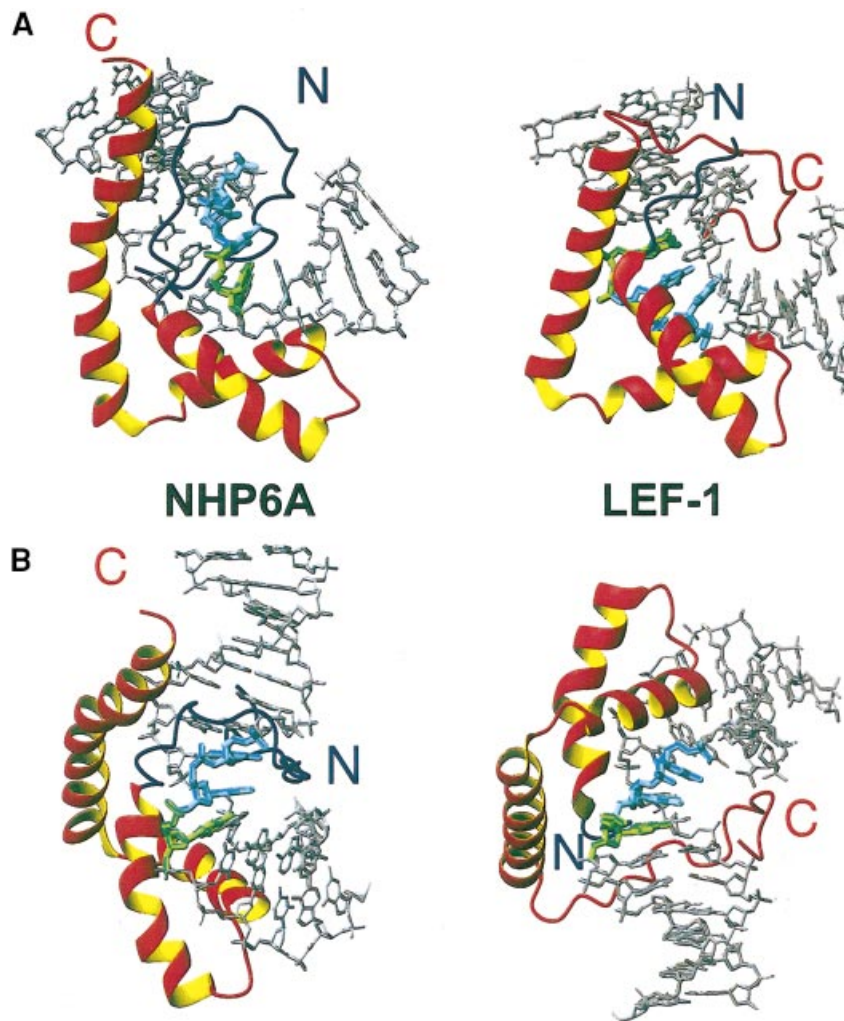


Fig. 9. Comparison between the NHP6A–DNA^{87Y} NMR-based model (left) and the LEF1–DNA structure (right) (Love *et al.*, 1995). **(A)** View of the two complexes with the two proteins oriented similarly. **(B)** View of the two complexes with their DNA oriented similarly. The DNA is gray except for the TTT sequence which is highlighted in cyan (T) and green (G). The non-helical N- and C-terminal residues are shown in blue and red, respectively.

of the HMG domain (Table II and Yen *et al.*, 1998). The path of the basic N-terminus of NHP6A is also indirectly mediated by the interaction of the extended N-terminus (residues 18–25) with helix III. The curvature of helix III, and its interactions (Lys78, Tyr81, Lys85, Tyr88) with the phosphodiester backbone of strand 2 (Figure 6B) directs the basic N-terminal residues starting with Lys16 in an orientation towards the major groove.

In a similar manner, the bend in helix 3 of HMG-D could serve to bring its C-terminal basic tail into the major groove (Jones *et al.*, 1994). The short basic linker between the two HMG boxes of HMG1 is also likely interacting with the major groove, as evidenced by the cross-link found between one of the basic residues (equivalent of Lys16 in NHP6A) and the DNA major groove of a cisplatinated DNA (Kane and Lippard, 1996). As also noted by others, interactions with a basic tail in the major groove either from the C-terminus (in LEF1 and HMG-D) or the N-terminus (in NHP6A and HMG1B) appears to be a general feature that is indispensable for promoting a relatively stable DNA complex and for DNA bending toward the major groove (Love *et al.*, 1995; Teo *et al.*, 1995; Payet and Travers, 1997; Yen *et al.*, 1998).

Cooperative folding of NHP6A and DNA upon complex formation at 37°C

A number of DNA binding proteins have been found to undergo significant conformational changes upon binding DNA, a phenomenon that has been referred to as an ‘induced fit’ model for DNA site recognition (Spolar and Record, 1994). Examples include the folding of otherwise disordered segments in *Bam*HI, *Eco*RV, the N-terminal arms of λ repressor and homeodomains, and the formation of the basic α -helix in leucine zipper and helix–loop–helix proteins upon binding DNA (Jordan and Pabo, 1988; Luisi *et al.*, 1991; Winkler *et al.*, 1993; Ellenberger *et al.*, 1994; Newman *et al.*, 1995). The binding of NHP6A to DNA at 37°C is a particularly dramatic example of cooperative folding between a protein and its DNA binding site that would be expected to be coupled with a large favorable entropic change (Spolar and Record, 1994). NHP6A is largely unfolded at 37°C as measured by either NMR or CD (Yen *et al.*, 1998), and the DNA must undergo a large structural change to conform to the NHP6A surface. Nevertheless, NHP6A readily forms DNA complexes that are stable to gel electrophoresis at 37°C, and the most optimal NMR spectra of the complex were

obtained at this temperature. We imagine that at 37°C an unfolded NHP6A HMG domain initially associates with a DNA molecule through electrostatic forces and then a cooperative folding-assembly occurs facilitated by hydrophobic interactions between the surfaces of the HMG domain and the DNA minor groove.

NHP6A binds at a different location than SRY within the same sequence

Previous studies have suggested that the sequence-specific and non-sequence-specific HMG box proteins bind DNA in a similar manner (Churchill *et al.*, 1995; Peters *et al.*, 1995). We found, however, that NHP6A is positioned onto the DNA quite differently from the sequence-specific HMG box proteins SRY and LEF1. The recognition sequences for SRY and LEF1 contain a common 5'-TTG-3' motif (Figure 1B). In all three proteins, a series of hydrophobic residues at the convex corner of the L (Leu25, Met29, Tyr28, Trp59 for NHP6A), which have been described as forming a hydrophobic wedge in the case of SRY (Werner *et al.*, 1995b), are localized at a dinucleotide step within the TTG motif. In the NHP6A-DNA complex, Leu25 contacts the deoxyriboses of T9 and G10, whereas the equivalent residues Met64 of SRY and Leu6 of LEF1 are located between the deoxyriboses of the two adenines on the opposite strand (Love *et al.*, 1995; Werner *et al.*, 1995b). Therefore, as illustrated in Figure 9 for the NHP6A and LEF1 complexes, NHP6A binds in an opposite orientation and one dinucleotide step shifted on the DNA from the sequence-specific complexes.

The difference in binding between the sequence-specific and non-sequence-specific HMG box proteins may be biologically important because if the much more abundant HMG1/2 proteins were to bind in the same manner as SRY or LEF1, they might promote aberrant transcription of the target genes. By binding in reverse orientation, the non-sequence-specific proteins would probably not be able to activate the promoter and may even function to inhibit inappropriate expression.

Selection of DNA binding site by NHP6A and other non-specific HMG proteins

At the macrosequence level, NHP6A binds to B-DNA non-specifically as illustrated by gel mobility shift, DNase I footprint and ligase-mediated circularization assays (Paull and Johnson, 1995; Yen *et al.*, 1998). However, NHP6A, like other non-specifically binding HMG1/2 proteins, binds with increased affinity to pre-bent DNA such as microcircles and cisplatinated DNA (Pil and Lippard, 1992; Pil *et al.*, 1993; Payet and Travers, 1997; Yen *et al.*, 1998). Therefore at the microsequence level, NHP6A probably selectively binds to a DNA segment that can most easily adopt to the distorted configuration present in the complex. Indeed, in the present case, a specific site was chosen within the 15 bp SRY and LEF1 sequences. The greatest DNA deformations in the model structure occur at the center of the binding site and include the large kinks at T9-G10 and G10-T11 mediated by insertion of the Met29 and Phe48 side chains into the respective base stacks. Consistent with this, using *in vitro* selection experiments for preferred HMG-D binding sites in random DNA sequences, Churchill *et al.* (1995) found that T-G plus G-T were the most over-represented dinucleotides present

after several rounds of selection. In addition, a large number of structural studies have demonstrated that a pyrimidine-purine dinucleotide step, such as T-G, has a strong tendency for generating positive roll because of its stacking properties and is often found at positions of protein-induced kinks in DNA (Dickerson, 1998; Dornberger *et al.*, 1998). NHP6A-DNA complex formation therefore may initiate at a pre-bent pyrimidine-purine step where the protruding methionine side chain could most easily penetrate. In the NHP6A-DNA^{STY} model structure, the amino group of the G is well accommodated at the interface and may even form a hydrogen bond to Tyr28 (Figure 7). On the other hand, a C instead of a T at the pyrimidine-purine step would be disfavored because of a steric clash by the amino group of the complementary G. Thus, preferential sites on the DNA may be targeted by these proteins because of their pre-existing conformation (pyrimidine-purine dinucleotides, cisplatinated DNA, microcircular DNA and four-way junctions), their ability to support a small number of hydrogen bonds (selectivity for a T-G dinucleotide), and their ability to adapt to the concave binding surface of the protein.

Materials and methods

Construction of NHP6A mutants and protein purification

The NHP6A mutants were constructed by direct cloning of two-step PCR products using mutant oligonucleotide primers (Landt *et al.*, 1990). The mutant genes were transferred into pET11a and transformed into RJ1878 [BL21 (DE3) *hupA::cm hupB::km*] for protein overexpression (Paull *et al.*, 1996). Each mutant gene in pET11a was sequenced in its entirety. Purification of the wild-type and mutant proteins was performed as described in Yen *et al.* (1998). For ¹⁵N-labeled wild-type NHP6A, cells were grown in 15 l of minimal medium A (Miller, 1992) with 8 mM (¹⁵NH₄)₂SO₄ (Isotec) and 0.2% glucose. Uniformly ¹⁵N,¹³C-labeled samples were prepared from cells grown in 8 l of the above medium containing 0.125% [¹³C]glucose (Isotec). Each of the wild-type and mutant NHP6A preparations were judged to be >99% pure by Coomassie Blue staining of overloaded SDS-polyacrylamide gels. Protein concentrations were determined by Coomassie Blue staining of SDS-polyacrylamide gels containing known concentrations of NHP6A that were originally established by direct amino acid analysis (Yen *et al.*, 1998).

Analysis of NHP6A mutants

Electrophoretic mobility shift, DNA ligase-mediated circularization and CD analyses were performed as described in Yen *et al.* (1998) with the exception that bromophenol blue was not included in the binding reactions for gel shift assays. This modification resulted in binding affinities being ~10-fold greater for wild-type NHP6A than reported previously (Paull and Johnson, 1995; Yen *et al.*, 1998). Affinities were calculated by plotting the log of the protein concentration versus the log of (*b*/1-*b*), where *b* is the fraction of bound DNA. When the value of log (*b*/1-*b*) is zero, 50% of the DNA is bound. Dissociation constants are an average of at least four independent experiments.

NMR sample preparation

The free protein was extensively dialyzed against 10 mM NaPO₄ pH 5.5, 100 mM NaCl, and used at a concentration of 1-2 mM for the NMR. The 15 nucleotide unlabeled DNA single strands of DNA^{STY} and DNA^{EF} were chemically synthesized on an Applied Biosystems 392 DNA synthesizer using standard phosphoramidite chemistry. Formation of the duplex DNA was monitored by NMR, by titration of one strand into the other until a 1:1 duplex was formed. Uniformly ¹³C,¹⁵N-labeled DNA^{STY} was prepared enzymatically as previously described (Masse *et al.*, 1998). Three DNA duplex samples were prepared: one with both strands ¹³C,¹⁵N-labeled, one with only strand 1 labeled, and the third with only strand 2 labeled. The use of these 'half-labeled' duplexes was essential to help in the resonance assignment of the free and bound DNA and to unambiguously assign intermolecular NOEs to one DNA strand or the

other. Sample conditions for the unlabeled and labeled DNA duplexes were 50 mM NaCl, 10 mM NaPO₄ pH 6.0 and 1–2 mM DNA duplex.

NHP6A–DNA complexes were formed by titration of the protein into the DNA. The titration was monitored by 1D NMR at 310 K by adding an increasing amount of protein to the DNA until a 1:1 ratio was reached as judged by the relative height of the Trp59 N_εH to the DNA imino or by the disappearance of some imino protons of the free DNA. The sample (1–2 mM) was then dialyzed overnight against a low salt buffer (10 mM NaCl, 2 mM NaPO₄ pH 5.8).

NMR spectroscopy

NMR spectra were recorded on Bruker DRX 500, 600 and 750 MHz spectrometers. Spectra were processed with Bruker Xwinnmr and analyzed with the Felix 97 software. All the spectra of the free protein were recorded at 293 K with the exception of an ¹H-¹⁵N sensitivity-enhanced TROSY (Czisch and Boelens, 1998; Salzmann *et al.*, 1998) recorded at 310 K. For spectral assignment, 3D CBCACONH (Grzesiek and Bax, 1992a), CBCANH (Grzesiek and Bax, 1992b), HCCH-TOCSY (Bax *et al.*, 1990), TOCSY-HMQC and 2D homonuclear TOCSY (Briand and Ernst, 1991) were recorded. For the structure determination, a series of 2D homonuclear NOESY spectra at different mixing times (30, 60, 90 and 150 ms) in H₂O and D₂O, three 3D at 150 ms mixing time (a ¹⁵N NOESY-HMQC and a ¹³C HSQC-NOESY in H₂O and a ¹³C NOESY-HMQC in D₂O) were recorded at 600 MHz.

All the spectra of the free DNA and the protein–DNA complex were recorded at 310 K. For spectral assignments of the free DNA, 2D homonuclear TOCSY (50 ms mixing time) (Briand and Ernst, 1991) and NOESY (300 ms mixing time) were recorded on the unlabeled sample. 2D ¹⁵N HMQC, 2D ¹³C HSQC and 3D ¹³C NOESY-HMQC (300 ms mixing time) (Cavanagh *et al.*, 1996) and HCCH-TOCSY (Bax *et al.*, 1990) were recorded on the isotopically labeled sample. The same set of heteronuclear spectra previously taken on the free labeled protein was recorded on the protein–DNA complex where the protein was isotopically labeled, and similarly for the complex where the DNA was isotopically labeled. All the NOESY spectra (2D and 3D) were recorded with a 150 ms mixing time. In addition, 3D double half-filtered HMQC-NOESY (Lee *et al.*, 1994) (150 ms mixing time) were recorded on the complexes with only one of the two components isotopically labeled at a time to observe intermolecular NOEs (Masse *et al.*, 1999).

Protein assignment strategy

The free protein was assigned applying well established methods. The backbone amide, C_αH, C_βH were assigned using the 3D CBCACONH (Grzesiek and Bax, 1992a), CBCANH (Grzesiek and Bax, 1992b), ¹⁵N TOCSY-HMQC and ¹⁵N and ¹³C HSQC-NOESY recorded in H₂O. The side-chain assignments were completed by the analysis of the 3D HCCH-TOCSY and NOESY-HMQC recorded in D₂O. The backbone and side-chain assignment is complete except for the N-terminal region (Met1–Lys16) which is unstructured and whose resonances (with the exception of Thr11 and Thr12) overlap.

Resonance assignment of the protein in the complex was first obtained by comparison with the free protein since fairly small chemical shift changes (backbone and side chain) were observed. They were later confirmed by the sequential connectivities. Despite severe line-broadening for several residues, the resonance assignment is complete from Glu17 to Ala93. The resonances of the basic N-terminal residues (Met1–Lys16) were broader and more dispersed but could not be assigned because of spectral overlap of the many lysine (five) and arginine (three) side-chain resonances which all have very similar chemical shifts.

DNA assignment strategy

The resonances of the two DNA duplexes DNA^{sfy} and DNA^{lef1} were assigned using well established methodology. The imino and adenine H2 resonances were assigned using the sequential connectivities observed in the 2D NOESY in H₂O using a 11 echo water suppression scheme. They confirmed the published assignments of the imino resonances of the SRY DNA (King and Weiss, 1993). All the other non-exchangeable proton resonances except the H5', H5'' were assigned using the 2D homonuclear NOESY and TOCSY spectra on the unlabeled sample. Confirmation of the deoxyribose proton assignments as well as assignment of the H5', H5'' together with the ¹³C assignments were made with the 3D HCCH-TOCSY and NOESY-HMQC spectra recorded on the isotopically labeled DNA.

Most of the DNA imino proton resonances in the complex could not be assigned due to broadening of several resonances, particularly the T iminos, and spectral overlap (not shown). Initial assignments of the non-exchangeable base and deoxyribose H1', H2', H2'', and H3'

resonances of the DNA in the complex were obtained, as described above for the free DNA. These assignments were confirmed and extended to the H4' and H5', H5'' resonances from analysis of 3D ¹H-¹³C NOESY-HMQC spectra. All of the non-exchangeable resonances with the exception of the A22 and A23 H2, H4', H5', H5'' were assigned. These latter resonances showed more line broadening than the other DNA nucleotides in the complex.

Structure calculations for NHP6A

Inter-proton distance constraints were obtained from 2D homonuclear NOESY spectra at different mixing time and 3D ¹⁵N and ¹³C separated NOESY spectra at 150 ms mixing time in H₂O and D₂O. The volume of the NOE cross-peaks assigned in the 2D NOESY and the ¹⁵N NOESY-HMQC spectra were integrated with the program spscan (Glaser and Wuthrich, www.mol.biol.ethz.ch/wuthrich/software/spscan/) and converted into distance constraints using the subroutine CALIBA within the program XEASY (Guntert *et al.*, 1991). Because of spectral overlap in the ¹³C NOESY-HMQC in D₂O, the peaks were not volume integrated, so the distance constraints derived from the assigned NOEs of this spectrum were all given an upper limit of 5 or 7 Å (for the very weak correlation) plus a pseudo atom correction. This analysis resulted in 1393 relevant distance constraints (Table I). Sixty-eight loose dihedral ψ (160° ± 100°) and φ (160° ± 80°) angle constraints were added based on the deviation of C_α shift from the random coil value and hydrogen bond constraints were added based on the observation of slowly exchanging amide protons when the protein was freshly put into D₂O. Fifty structures were calculated in dihedral angle space using the program DYANA (Guntert *et al.*, 1997). The 30 lowest energy structures form the ensemble of converged structures which are further analyzed and described (Table I).

Structure calculations for the model of the NHP6A–DNA complex

The calculations were done within X-PLOR 3.1 (Brunger, 1992). The first phase of the modeling started with the lowest energy structure of the pdb structure 2LEF of the LEF1–DNA complex. The bases of the DNA in that complex were changed to the DNA^{sfy} sequence by best fit of sugar atoms using standard nucleotides from fiber diffraction data (Biosym). This change was made such that the A23–A24 step in the LEF1 DNA was replaced by T9–G10 in DNA^{sfy} and the other bases accordingly. The coordinates of the sugar and backbone atoms were left unchanged. The terminal G10C30 base pair in the DNA^{sfy} was added since it had no equivalent in the DNA^{lef1} sequence. The LEF1 protein was then replaced with the lowest NHP6A structure by best fit of backbone atoms for residues 25–75 of NHP6A with LEF1 residues 6–56 (the r.m.s. difference between these two structures is 1.6 Å). At this point some bad base pair geometry and a few clashes between the protein side chains and the DNA were corrected using the standard simulated annealing protocol in X-PLOR, except that the backbone of the DNA and residues 25–75 of the protein were restrained to their original positions with the harmonic coordinate restraint function in X-PLOR (Brunger, 1992). DNA hydrogen bonds were enforced with distance restraints, and good base pair geometry was obtained by adding planarity restraints for the base atoms in each base pair. In the next phase, the same calculation was run including distance constraints used in the free protein structure calculation and the eight unambiguous intermolecular distance constraints (upper limits 5 and 6 Å) from Leu25 H_δs to T9 and G10 sugar ring. Examination of the complex at this point showed that the N-terminus of the protein was in the major groove of the DNA and that several arginine and lysine side chains appeared to be close enough to interact with a DNA phosphate oxygen. Taking advantage of the striking sequence similarity between the stretch R₁₃K₁₄K₁₅K₁₆ in NHP6A and the stretch R₈₁K₈₀K₇₉K₇₈ in LEF1, we modeled R13–K16 similarly to the LEF1–DNA complex (Love *et al.*, 1995) by using a distance constraint of 3 Å between the amino or guanidinium protons and the contacted phosphate oxygen. K22, R23, R36, R40, K53, K60, K67, K78 and K85 were modeled with a similar constraint to the closest phosphate oxygen in the model structure. Finally, K8, K9 and R10 were modeled based on a proposed model of interaction of stretches of arginine in the major groove of DNA (Hud *et al.*, 1994) using the same type of constraints. Calculations with these latter constraints resulted in the NMR-based model presented here. No violations were found either in the intermolecular NOEs or the side chain to phosphate oxygen constraints, and no violations of more than 0.2 Å were found for the constraints coming from the free protein structure. Indeed, the protein in the complex deviates with an r.m.s.d. of 1.5 Å from the lowest energy structure of the free NHP6A (superimposition

of the backbone atoms from residue 20 to 90). Finally, the model of the complex has excellent geometry and van der Waals contacts.

Coordinate deposition

Coordinates for the 30 lowest energy structures of free NHP6A have been deposited in the Protein Data Bank (accession number 1cg7). Coordinates for the model structure of the NHP6A–DNA^{SY} complex are available upon request from the authors.

Acknowledgements

We thank Rick Fahrner for help on the early part of this work and for DNA purification, Lenore Landis for initial characterization of several NHP6A mutants, Michael Haykinson for computer support and the members of our laboratories for useful discussion. This work was supported by NIH grants GM48123 to J.F. and GM38509 to R.C.J., an American Cancer Society Faculty Research Award to R.C.J., an NIH NIGMS training grant GM08042, the Medical Scientist Training Program and the Aesculapian Fund of the UCLA School of Medicine to Y.-M.Y., and a European Molecular Biology Organization and Human Frontiers Sciences Program Organization postdoctoral fellowships to F.H.-T.A.

References

- Balaeff, A., Churchill, M.E.A. and Schulten, K. (1998) Structure prediction of a complex between the chromosomal protein HMG-D and DNA. *Proteins Struct. Funct. Genet.*, **30**, 113–135.
- Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) ¹H-¹H correlation via isotropic mixing of ¹³C magnetization, a new three-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *J. Magn. Reson.*, **88**, 425–431.
- Briand, J. and Ernst, R.R. (1991) Computer-optimized homonuclear Tocsy experiments with suppression of cross relaxation. *Chem. Phys. Lett.*, **185**, 276–285.
- Brunger, A.T. (1992) *X-PLOR Manual Version 3.1*. Yale University, New Haven, CT.
- Bustin, M. and Reeves, R. (1996) High-mobility-group chromosomal proteins: Architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.*, **54**, 35–100.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G., III and Skelton, N.J. (1996) *Protein NMR Spectroscopy: Principles and Practice*. Academic Press, Inc., San Diego, CA.
- Churchill, M.E., Jones, D.N., Glaser, T., Hefner, H., Searles, M.A. and Travers, A.A. (1995) HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J.*, **14**, 1264–1275.
- Costigan, C., Kolodrubetz, D. and Snyder, M. (1994) NHP6A and NHP6B, which encode HMG1-like proteins, are candidates for downstream components of the yeast SLT2 mitogen-activated protein kinase pathway. *Mol. Cell. Biol.*, **14**, 2391–2403.
- Czisch, M. and Boelens, R. (1998) Sensitivity enhancement in the TROSY experiment. *J. Magn. Reson.*, **134**, 158–160.
- Dickerson, R. (1998) Helix structure and molecular recognition by B-DNA. In Neidle, S. (ed.), *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, Oxford, UK, pp. 1–23.
- Dornberger, U., Flemming, J. and Fritzsche, H. (1998) Structure determination and analysis of helix parameters in the DNA decamer d(CATGGCCATG)₂: comparison of results from NMR and crystallography. *J. Mol. Biol.*, **284**, 1453–1463.
- Ellenberger, T., Fass, D., Arnaud, M. and Harrison, S.C. (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer. *Genes Dev.*, **8**, 970–980.
- Ge, H. and Roeder, R.G. (1994) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J. Biol. Chem.*, **269**, 17136–17140.
- Grosschedl, R., Giese, K. and Pagel, J. (1994) HMG domain proteins: Architectural elements in the assembly of nucleoprotein structures. *Trends Genet.*, **10**, 94–100.
- Grzesiek, S. and Bax, A. (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114**, 6291–6293.
- Grzesiek, S. and Bax, A. (1992b) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.*, **99**, 201–207.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, **346**, 245–250.
- Guntert, P., Braun, W. and Wuthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.*, **217**, 517–530.
- Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
- Haqq, C.M., King, C.Y., Ukiyama, E., Falsafi, S., Haqq, T.N., Donahoe, P.K. and Weiss, M.A. (1994) Molecular basis of mammalian sexual determination: Activation of Mullerian inhibiting substance gene expression by SRY [published erratum appears in *Science* (1995) **267**, 317]. *Science*, **266**, 1494–1500.
- Hardman, C.H., Broadhurst, R.W., Raine, A.R., Grasser, K.D., Thomas, J.O. and Laue, E.D. (1995) Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry*, **34**, 16596–16607.
- Higgins, D.G., Thompson, J.D. and Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383–402.
- Hud, N.V., Milanovich, F.P. and Balhorn, R. (1994) Evidence of novel secondary structure in DNA-bound protamine is revealed by Raman spectroscopy. *Biochemistry*, **33**, 7528–7535.
- Jayaraman, L., Moorthy, N.C., Murthy, K.G., Manley, J.L., Bustin, M. and Prives, C. (1998) High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev.*, **12**, 462–472.
- Jones, D.N., Searles, M.A., Shaw, G.L., Churchill, M.E., Ner, S.S., Keeler, J., Travers, A.A. and Neuhaus, D. (1994) The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*. *Structure*, **2**, 609–627.
- Jordan, S.R. and Pabo, C.O. (1988) Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science*, **242**, 893–899.
- Kane, S.A. and Lippard, S.J. (1996) Photoreactivity of platinum (II) in cisplatin-modified DNA affords specific cross-links to HMG domain proteins. *Biochemistry*, **35**, 2180–2188.
- Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993a) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
- Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993b) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
- King, C.Y. and Weiss, M.A. (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natl Acad. Sci. USA*, **90**, 11990–11994.
- Kolodrubetz, D. and Burgum, A. (1990) Duplicated NHP6 genes of *Saccharomyces cerevisiae* encode proteins homologous to bovine high mobility group protein 1. *J. Biol. Chem.*, **265**, 3234–3239.
- Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55, 29–32.
- Kuehl, L., Salmond, B. and Tran, L. (1984) Concentrations of high-mobility-group proteins in the nucleus and cytoplasm of several rat tissues. *J. Cell Biol.*, **99**, 648–654.
- Landt, O., Grunert, H.P. and Hahn, U. (1990) A general method for rapid site-directed mutagenesis using the polymerase chain reaction. *Gene*, **96**, 125–128.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
- Lee, W., Revington, M.J., Arrowsmith, C. and Kay, L.E. (1994) A pulsed field gradient isotope-filtered 3D ¹³C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.*, **350**, 87–90.
- Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791–795.
- Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352**, 497–505.
- Masse, J.E., Bortmann, P., Dieckmann, T. and Feigon, J. (1998) Simple, efficient protocol for enzymatic synthesis of uniformly ¹³C, ¹⁵N-labeled DNA for heteronuclear NMR studies. *Nucleic Acids Res.*, **26**, 2618–2624.

- Masse, J.E., Allain, F.H.-T., Yen, Y., Johnson, R.C. and Feigon, J. (1999) Use of ^{13}C , ^{15}N -labeled DNA in a non-sequence specific protein–DNA complex resolves ambiguous assignments of intermolecular NOES. *J. Am. Chem. Soc.*, in press.
- Miller, J.H. (1992) *A Short Course in Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of BamHI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Nightingale, K., Dimitrov, S., Reeves, R. and Wolffe, A.P. (1996) Evidence for a shared structural role for HMG1 and linker histones B4 and H1 in organizing chromatin. *EMBO J.*, **15**, 548–561.
- Paull, T.T. and Johnson, R.C. (1995) DNA looping by *Saccharomyces cerevisiae* high mobility group proteins NHP6A/B. Consequences for nucleoprotein complex assembly and chromatin condensation. *J. Biol. Chem.*, **270**, 8744–8754.
- Paull, T.T., Haykinson, M.J. and Johnson, R.C. (1993) The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.*, **7**, 1521–1534.
- Paull, T.T., Carey, M. and Johnson, R.C. (1996) Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation *in vivo* and assembly of preinitiation complexes *in vitro*. *Genes Dev.*, **10**, 2769–2781.
- Payet, D. and Travers, A. (1997) The acidic tail of the high mobility group protein HMG-D modulates the structural selectivity of DNA binding. *J. Mol. Biol.*, **266**, 66–75.
- Peters, R., King, C.Y., Ukiyama, E., Falsafi, S., Donahoe, P.K. and Weiss, M.A. (1995) An SRY mutation causing human sex reversal resolves a general mechanism of structure-specific DNA recognition: Application to the four-way DNA junction. *Biochemistry*, **34**, 4569–4576.
- Pil, P.M. and Lippard, S.J. (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin. *Science*, **256**, 234–237.
- Pil, P.M., Chow, C.S. and Lippard, S.J. (1993) High-mobility-group 1 protein mediates DNA bending as determined by ring closures. *Proc. Natl Acad. Sci. USA*, **90**, 9465–9469.
- Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C. and Norman, D.G. (1993) Solution structure of a DNA-binding domain from HMG1. *Nucleic Acids Res.*, **21**, 3427–3436.
- Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wuthrich, K. (1998) TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA*, **95**, 13585–13590.
- Shykind, B.M., Kim, J. and Sharp, P.A. (1995) Activation of the TFIID–TFIIA complex with HMG-2. *Genes Dev.*, **9**, 1354–1365.
- Sinclair, A.H. *et al.* (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240–244.
- Spolar, R.S. and Record, M.T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
- Teo, S.H., Grasser, K.D. and Thomas, J.O. (1995) Differences in the DNA-binding properties of the HMG-box domains of HMG1 and the sex-determining factor SRY. *Eur. J. Biochem.*, **230**, 943–950.
- Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. *Genes Dev.*, **5**, 880–894.
- Ura, K., Nightingale, K. and Wolffe, A.P. (1996) Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA: Structural transitions and transcriptional repression. *EMBO J.*, **15**, 4959–4969.
- van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.*, **10**, 123–132.
- van Gent, D.C., Hiom, K., Paull, T.T. and Gellert, M. (1997) Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.*, **16**, 2665–2670.
- van Houte, L.P., Chuprina, V.P., van der Wetering, M., Boelens, R., Kaptein, R. and Clevers, H. (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J. Biol. Chem.*, **270**, 30516–30524.
- Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D. and Thomas, J.O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.*, **12**, 1311–1319.
- Wen, L., Huang, J.K., Johnson, B.H. and Reece, G.R. (1989) A human placental cDNA clone that encodes nonhistone chromosomal protein HMG-1. *Nucleic Acids Res.*, **17**, 1197–1214.
- Werner, M.H., Bianchi, M.E., Gronenborn, A.M. and Clore, G.M. (1995a) NMR spectroscopic analysis of the DNA conformation induced by the human testis determining factor SRY. *Biochemistry*, **34**, 11998–12004.
- Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995b) 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.
- Winkler, F.K. *et al.* (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
- Wishart, D.S. and Sykes, B.D. (1994) The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.
- Yen, Y.M., Wong, B. and Johnson, R.C. (1998) 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.*, **273**, 4424–4435.
- Zappavigna, V., Falciola, L., Citterich, M.H., Mavilio, F. and Bianchi, M.E. (1996) HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. *EMBO J.*, **15**, 4981–4991.
- Zwilling, S., Konig, H. and Wirth, T. (1995) High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J.*, **14**, 1198–1208.

Received February 8, 1999; revised and accepted March 17, 1999