Structure of the A-Domain of HMG1 and Its Interaction with DNA As Studied by Heteronuclear Three- and Four-Dimensional NMR Spectroscopy[†,‡]

Colin H. Hardman, R. William Broadhurst, Andrew R. C. Raine, Klaus D. Grasser,[§] Jean O. Thomas,* and Ernest D. Laue*

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

Received June 21, 1995; Revised Manuscript Received October 9, 1995[⊗]

ABSTRACT: HMG1 has two homologous, folded DNA-binding domains ("HMG boxes"), A and B, linked by a short basic region to an acidic C-terminal domain. Like the whole protein, which may perform an architectural role in chromatin, the individual boxes bind to DNA without sequence specificity, have a preference for distorted or prebent DNA, and are able to bend DNA and constrain negative superhelical turns. They show qualitatively similar properties with quantitative differences. We have previously determined the structure of the HMG box from the central B-domain (77 residues) by two-dimensional NMR spectroscopy, which showed that it contains a novel fold [Weir *et al.* (1993) *EMBO J. 12,* 1311−1319]. We have now determined the structure of the A-domain (as a Cys → Ser mutant at position 22 to avoid oxidation, without effect on its DNA-binding properties or structure) using heteronuclear three- and four-dimensional NMR spectroscopy. The A-domain has a very similar global fold to the B-domain and the *Drosophila* protein HMG-D [Jones *et al.* (1994) *Structure 2,* 609−627]. There are small differences between A and B, in particular in the orientation of helix I, where the B-domain is more similar to HMG-D than it is to the A-domain; these differences may turn out to be related to the subtle differences in functional properties between the two domains [Teo *et al.* (1995) *Eur. J. Biochem. 230,* 943−950] and will be the subject of further investigation. NMR studies of the interaction of the A-domain of HMG1 with a short double-stranded oligonucleotide support the notion that the protein binds via the concave face of the L-shaped structure; extensive contacts with the DNA are made by the N-terminal extended strand, the N-terminus of helix I, and the C-terminus of helix II. These contacts are very similar to those seen in the LEF-1 and SRY−DNA complexes [Love *et al.* (1995) *Nature 376,* 791−795; Werner *et al.* (1995) *Cell 81,* 705−714].

The closely related "high mobility group" non-histone chromosomal proteins HMG1 and HMG2 in vertebrates are abundant, non-sequence-specific, DNA-binding proteins with a well conserved primary structure (Johns, 1982) which have been implicated in replication, transcription, and chromatin assembly (Bustin *et al.*, 1990). The proteins comprise two folded domains, A and B [termed "HMG boxes" when homologous regions were identified in the RNA polymerase I transcription factor UBF (Jantzen *et al.*, 1990)], and an acidic C-terminal tail. HMG boxes have since been identified in a number of transcription factors that bind DNA with sequence specificity, as well as in several abundant non-sequence-specific chromosomal proteins in lower eukaryotes. For recent reviews of HMG box proteins, see Ner (1992), Landsman and Bustin (1993), Grosschedl *et al.* (1994), and Baxevanis and Landsman (1995). Recent studies suggest that HMG1 and HMG-2 are able to interact with other DNA-binding proteins and modulate transcription, either positively in the case of the progesterone receptor and HMG1 (Oñate *et al.*, 1994) and Oct1 and HMG2 (Zwilling *et al.*, 1995) or negatively in the case of TBP and HMG1 (Ge & Roeder, 1994; Stelzer *et al.*, 1994).

HMG1 binds preferentially to distorted DNA structures such as four-way junctions (Bianchi *et al.*, 1989), *cis*-platin−DNA adducts (Pil & Lippard, 1992; Locker *et al.*, 1995) and DNA with an enforced bend (Wolfe *et al.*, 1995), and bends linear-double-stranded DNA in a sequence-independent manner (Pil *et al.*, 1993; Paull *et al.*, 1993). LEF-1 and SRY also bend DNA that contains their cognate binding sites (through ~135° and 85°, respectively), and for all three proteins the bending is a property of the HMG-box domains. Biochemical (Giese *et al.*, 1991; van de Wetering *et al.*, 1993) and NMR[1] studies (King & Weiss, 1993; Peters *et al.*, 1995; Churchill *et al.*, 1995) suggested that HMG boxes contact DNA mainly through the minor groove. This has recently been confirmed directly in the structure determinations, by NMR spectroscopy, of the HMG boxes of SRY

[†] This work was supported by grants from the Science and Engineering Research Council/Biotechnology and Biological Sciences Research Council (BBSRC) of the U.K. and the Royal Society to J.O.T. and E.D.L. and through support for the Cambridge Centre for Molecular Recognition by the BBSRC and the Wellcome Trust. K.D.G. was the recipient of a Long-Term Fellowship from the European Molecular Biology Organisation.

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank (codes 1AAV and 1AAB).

* Authors to whom correspondence should be addressed.

[§] Present address: Institut für Biologie III, Albert-Ludwigs-Universität Freiburg, Schänzlestrasse 1, D-79104, Freiburg, Germany.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1995.

[1] Abbreviations: NMR, nuclear magnetic resonance; 2D, 3D, and 4D, two, three, and four dimensional, respectively; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; COSY, correlated spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum correlation.

0006-2960/95/0434-16596$09.00/0

(Werner *et al.*, 1995) and LEF-1 (Love *et al.*, 1995). The NMR studies indicate a striking similarity between the mode of DNA binding by the HMG-box proteins and the TATA-binding protein (TBP), the core component of the general transcription factor TFIID (Kim, Y., *et al.*, 1993; Kim, J. L., *et al.*, 1993); in both cases partial intercalation of exposed hydrophobic amino acid side chains into the minor groove leads to widening of the groove and bending of the DNA toward the major groove, away from the protein. The HMG-box proteins may thus have an architectural role in bending DNA which may facilitate the assembly of higher order nucleoprotein complexes [for a review, see Grosschedl *et al.* (1994)].

Although the differences in primary structure between the A- and B-domains (~30% identity) of HMG 1 are very well conserved, the significance of the two HMG boxes is at present unclear, particularly since many of the properties of the whole protein are also exhibited by the individual HMG-box domains. Most other HMG-box proteins, whether sequence-specific or non-sequence-specific, have a single box, the exceptions being mtTF1 (two boxes) and UBF (4−6 boxes depending on species) in which the different boxes may play different roles (Leblanc *et al.*, 1993). Comparison of the biochemical properties of the individual A- and B-domains of HMG1 showed that although they have broadly similar properties there are quantitative differences (Teo *et al.*, 1995a) which may be significant within the intact molecule.

We earlier reported the structure of the HMG-box motif in the B-domain of HMG1 which showed a novel and distinctive L-shaped structure comprising three α-helices and an N-terminal extended strand, with an angle of ~80° (defined as the angle between helices II and III) between the two arms (Weir *et al.*, 1993). A second structure of the same domain (as a 2-mercaptoethanol adduct at the single cysteine residue) (Read *et al.*, 1993) showed a similar fold but had a sharper angle (~40° between helices II and III) between the arms, giving a more globular shape. (The overall angle between helices I/II and helix III was given as 70°, compared with an angle of ~100° in our structure.) This raised the possibility that the structures of the isolated HMG boxes might be flexible in solution, where the elbow serves solely to tether the two arms to each other; alternatively, the structures (albeit of domains with the same amino acid sequence) might be genuinely different. In an attempt to resolve this issue, as well as with the hope of being able to identify the structural differences between the A and B HMG-box domains, we embarked on a study of the structure and backbone dynamics of the homologous A-domain of HMG1.

We now report the structure of the A-domain (as a Cys → Ser mutant at position 22 to prevent spurious oxidation) determined using heteronuclear three- and four-dimensional (3D/4D) NMR spectroscopy. In the accompanying paper (Broadhurst *et al.*, 1995) we describe ^{15}N NMR spectroscopic studies of the dynamics of the backbone nuclei in the A-domain. The structure reported here is very similar to that of the B-domain that we determined previously (Weir *et al.*, 1993), to the HMG box from HMG-D (Jones *et al.*, 1994), and to the structures of the DNA-bound forms of the sequence-specific HMG-boxes of SRY and LEF-1 (Werner *et al.*, 1995; Love *et al.*, 1995). Studies of the interaction of the non-sequence-specific A-domain HMG box with double-stranded DNA support the idea that the protein binds via the concave face of the L-shaped molecule as suggested previously (Weir *et al.*, 1993) and as found more recently for the SRY (King & Weiss, 1993; Peters *et al.*, 1995; Werner *et al.*, 1995), HMG-D (Churchill *et al.*, 1995), and LEF-1 (Love *et al.*, 1995) HMG boxes. It remains to be seen whether small differences between our A- and B-domain structures or, for example, amino acid sequence differences in surface residues are responsible for the subtle differences in properties between the A- and B- domains (Teo *et al.*, 1995a).

MATERIALS AND METHODS

Construction of pT7-7 HMG1-A (C22S). A cysteine to serine mutation was introduced at position 22 of the HMG1 coding sequence in the plasmid pRNHMG1 [from R. Cortese (Bianchi *et al.*, 1989)]. Single-stranded uracil-containing pRNHMG1 was recovered from *Escherichia coli* RZ1032 using the M13K07 helper phage. Using *in vitro* strand synthesis with the primer 5′ GTGCAAACCTCCCGGGAG-GAGC 3′, the codon change TGC → TCC (Cys 22 to Ser) was introduced and selected for in *ung*$^+$ *E. coli* TG1 cells (Kunkel, 1985; Kunkel *et al.*, 1987). The resulting plasmid pRNHMG1 (C22S) was verified by DNA sequencing. The region corresponding to amino acid residues 1−83 of rat HMG1 was then subcloned from either pRNHMG1 or pRNHMG1 (C22S) into pT7-7 (Tabor & Richardson, 1985) in two steps. First, the polymerase chain reaction, with primers 5′ GGATCCATATGGCAAAGGAGATCC 3′ and 5′ TCTAGAATTCATCACTCCCCTTTGGGGGGGA 3′, was used to introduce an *Nde*I restriction site containing the ATG start codon at the 5′-end of the coding sequence and two in-frame stop codons (TGA) followed by an *Eco*RI site at the 3′-end of the domain. Because the coding sequence of the A-domain contains an internal *Nde*I site, the 3′ *Nde*I−*Eco*RI fragment and the small 5′ *Nde*I−*Nde*I fragment were introduced in separate sequential subcloning steps. The final constructs, designated pT7-7 HMG1-A and pT7-7 HMG1-A (C22S), were verified by DNA sequencing on both strands.

Expression, Purification, and Characterization of the HMG1 A-Domain Fragment (Residues 1−83). *E. coli* BL21 (DE3) cells were transformed with pT7-7 HMG1-A or pT7-7 HMG1-A (C22S) which direct the expression of the wild type or the C22S mutant A-domain of rat HMG1 (residues 1−83), respectively. Cultures were grown at 37 °C in 2 × TY medium containing ampicillin (50 μg/mL); expression was induced at OD_{600} ~ 0.8 with 0.4 mM isopropyl thiogalactoside followed by incubation at 37 °C for 2 h. For the preparation of isotopically labeled proteins, cultures were grown on a minimal medium (Neidhardt *et al.*, 1974) containing [$^{13}C_6$]glucose and/or $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. The proteins were purified on S-Sepharose (Fast S) and Phenyl Sepharose columns essentially as described by Weir *et al.* (1993). The yield was ~4−6 mg/L of culture.

N-Terminal sequence analysis was carried out on ~ 500 pmol of the A-domain fragment using an Applied Biosystems 477 pulsed liquid sequencer. The molecular masses of the wild-type and mutant proteins were determined on ~250 pmol by electrospray ionization mass spectrometry using a VG BioQ quadrupole instrument (50% aqueous methanol/2% acetic acid as solvent; one injection of 20 μL). Circular

dichroism (CD) spectra from 195 to 260 nm (1 mm path length) were recorded at room temperature (~23 °C) using a Jobin-Yvon CD6 spectropolarimeter. Samples were at 0.1 mg/mL in 10 mM sodium phosphate, pH 5.0, 0.15 M NaCl, 1.0 mM dithiothreitol (DTT), and 0.1 mM EDTA.

Gel Retardation Assays. DNA-binding assays with ^{32}P-labeled four-way junction DNA were carried out as described previously (Teo *et al.*, 1995b). For assays with the 14-mer duplex DNA binding site studied by NMR (5′ GC-TATAAAAGGGCA 3′) (Kim, J. L., *et al.*, 1993), the salt concentration was lowered to 50 mM.

NMR Sample Preparation and NMR Spectroscopy. Pooled Phenyl Sepharose fractions were desalted by dialysis against 10 mM sodium phosphate, pH 5.0, and 0.2 mM DTT, vacuum-concentrated to 0.5 mL in a SpeedVac concentrator (Savant), and then dialyzed into the degassed final buffer of 10 mM sodium phosphate, pH 5.0, 0.15 M NaCl, 50 μM DTT, and 0.02% NaN_3 (buffer A). Preliminary 1D and 2D ^{1}H NMR spectra of the HMG1 A-domain were recorded in buffer A containing 10% D_2O at temperatures between 283 and 303 K. For the identification of slowly exchanging amide protons, the protein was dialyzed against 10 mM sodium phosphate, lyophilized, and dissolved in the original volume of buffer A (in 99.8% D_2O) lacking phosphate, immediately before recording spectra. All the spectra used in the structure determination were recorded at 293 K at a sample concentration of 2−5 mM on a Bruker AMX 600 spectrometer.

2D NOESY and *z*-filtered TOCSY (Rance, 1987) spectra were acquired, on the unlabeled protein sample in 90% H_2O/10% D_2O, with 256 (t_1) × 1024 (t_2) complex points to give acquisition times of 31.7 and 127.0 ms in t_1 and t_2, respectively. The NOESY spectra were acquired with mixing times of betweeen 30 and 200 ms while the TOCSY spectrum was acquired using a DIPSI-2 (Shaka *et al.*, 1988) mixing time of 46 ms.

The 3D ^{15}N-separated (Marion *et al.*, 1989a) and the 4D ^{13}C/^{15}N-separated (Kay *et al.*, 1990) NOESY spectra were acquired with a mixing time of 150 ms on either the ^{15}N-labeled or the ^{13}C/^{15}N-labeled protein in 90% H_2O/10% D_2O as appropriate. The 3D ^{15}N-separated spectrum was acquired with a total of 128 (t_1) × 32 (t_2) × 512 (t_3) complex points to give acquisition times of 15.9, 31.5, and 63.5 ms in t_1, t_2, and t_3 respectively. The 4D ^{13}C/^{15}N-separated NOESY spectrum was acquired with a total of 64 (t_1) × 8 (t_2) × 8 (t_3) × 512 (t_4) complex points to give acquisition times of 10.5, 2.7, 7.9, and 63.5 ms in t_1, t_2, t_3, and t_4, respectively.

2D ^{1}H−^{15}N HMQC spectra to study the interaction of the protein with DNA were recorded, on the ^{15}N-labeled protein sample, essentially as described by Sklenár and Bax (1987); spectra were acquired on a Bruker AM 500 instrument, without presaturation, using jump return pulses for ^{1}H, with a total of 64 (t_1) × 1024 (t_2) complex points to give acquisition times of 64 and 127 ms in t_1 and t_2, respectively.

3D ^{13}C-separated HCCH-COSY, HCCH-TOCSY (Bax *et al.*, 1990a,b), and NOESY (Ikura *et al.*, 1990; Zuiderweg *et al.*, 1990) spectra as well as the 4D ^{13}C/^{13}C-separated NOESY spectrum (Clore *et al.*, 1991; Zuiderweg *et al.*, 1991) were acquired on the ^{13}C/^{15}N-labeled protein in 100% D_2O. The HCCH-COSY and TOCSY spectra (24 ms mixing time) were acquired with a total of 80 (t_1) × 32 (t_2) × 512 (t_3) complex points to give acquisition times of 9.9, 10.7, and 63.5 ms in t_1, t_2, and t_3, respectively. The NOESY spectra (150 ms mixing time) were acquired with a total of 128 (t_1) × 26 (t_2) × 256 (t_3) complex points (3D) and 128 (t_1) × 8 (t_2) × 8 (t_3) × 256 (t_4) complex points (4D) to give acquisition times of 15.9 (t_1), 8.9 (t_2), and 31.7 (t_3) ms (3D) and 15.9 (t_1), 2.7 (t_2), 2.7 (t_3), and 31.7 (t_4) ms (4D), respectively.

4D HCANNH and HCA(CO)NNH spectra (Boucher *et al.*, 1992a,b) were acquired with a total of 64 (t_1) × 16 (t_2) × 8 (t_3) × 512 (t_4) complex points to give acquisition times of 10.5, 5.3, 7.9, and 63.5 ms in t_1, t_2, t_3, and t_4, respectively. 4D HCCNNH and HCC(CO)NNH spectra were acquired and processed as described previously (Richardson *et al.*, 1993; Clowes *et al.*, 1993).

The ^{1}H carrier was in all cases positioned on the water resonance whose intensity, in the spectra recorded in 90% H_2O/10% D_2O, was reduced by presaturation (20−25 Hz radio-frequency field strength) and baseline correction in the t_3 (for 3D NMR spectra) or t_4 (for 4D NMR spectra) time domain (Marion *et al.*, 1989b). 3D and 4D NMR spectra were typically processed first using conventional Fourier transforms in t_3 (3D) or t_4 and t_3 (4D), respectively; where appropriate, the low-frequency portion of the spectrum was discarded after the first transform. Processing in the t_1 and t_2 dimensions was typically achieved using a two-dimensional maximum-entropy algorithm (Laue *et al.*, 1986). All spectra were processed using the AZARA package (W. Boucher, unpublished; available by anonymous ftp from ftp.bio.cam.ac.uk:pub/azara) and analyzed using a 3D/4D version of the ANSIG program (Kraulis *et al.*, 1994) running on a Silicon Graphics Indigo computer.

Structures were calculated from the experimental constraints using the program X-PLOR (Brünger, 1992), employing the YASAP protocol (M. Nilges, personal communication, and Brünger, 1990) which starts from an initial structure (with random φ and ψ values) and assigns random initial velocities to each atom. Minor modifications to the protocol parameters, which increased the oxygen atom sigma value, were made as suggested by Bagby *et al.* (1994). All calculations were performed on Silicon Graphics Indigo computers.

RESULTS AND DISCUSSION

We initially studied the wild-type A-domain of HMG1 but found that its NMR spectra were badly overlapped with broad lines indicative of aggregation. Analysis of the protein by SDS gel electrophoresis under nonreducing conditions suggested the presence of more than one species in solution, including a dimer, suggestive of oxidation. We reasoned that, of the two cysteine residues, Cys 22 was most likely to be solvent accessible, as this residue is not conserved and is often charged in other HMG boxes whereas that corresponding to Cys 44 is usually hydrophobic. We therefore mutated Cys 22 to serine in order to prevent dimerization. Gel retardation assays (see Figure 1a) showed that the wild type and C22S mutant bound to the four-way junction DNA (as well as to a short duplex discussed later; Figure 1b) with similar affinities. 2D ^{1}H−^{15}N correlation spectra of the C22S mutant and of the wild-type A-domain (in the spectrum of the wild-type AB didomain) were also very similar (data not shown). Consistent with this finding, the circular dichroism (CD) spectra of the two proteins are very similar, with the mutant showing, if anything, slightly more α-helix

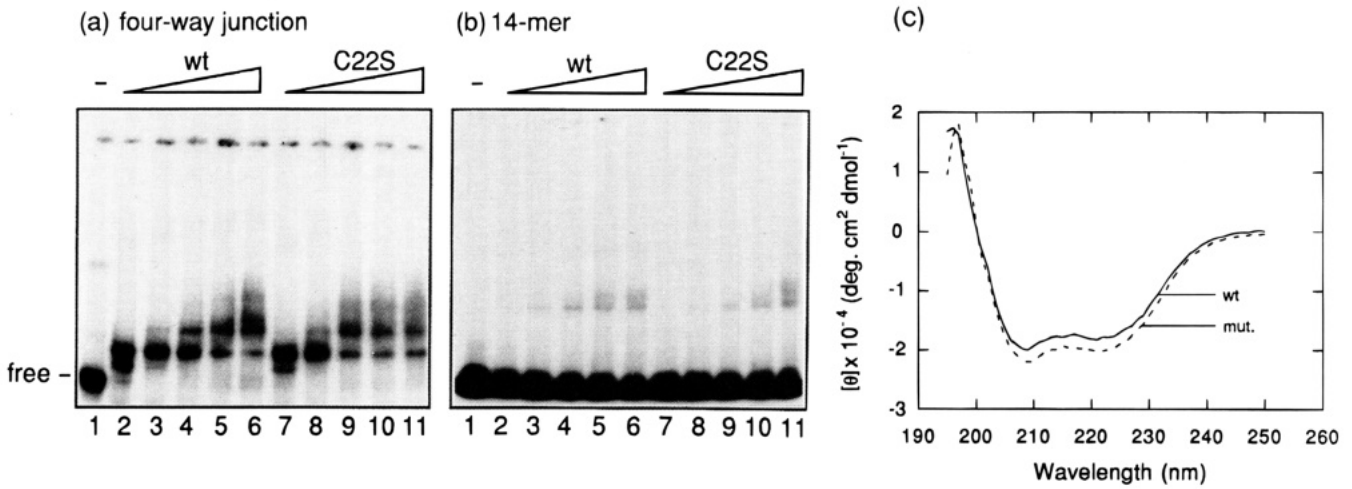

FIGURE 1: Comparison of the wild-type (wt) and mutant (C22S) A-domains. (a) Gel retardation assay showing binding to the four-way junction DNA. Lanes 2−6 and 7−11 contained 2.5 nM four-way junction and, respectively, 100, 200, 300, 400, and 500 ng of HMGA or the C22S mutant; lane 1 shows the free four-way junction DNA. (b) Gel retardation assay showing binding to the 14-mer duplex DNA. Lanes 2−6 and 7−11 contained the 14-mer (5 nM) and the same amounts of HMGA or the C22S mutant as in (a). (c) Far-UV circular dichroism spectra of the wild type (wt; solid line) and C22S mutant (mut; broken line). Both proteins were homogeneous as judged by SDS gel electrophoresis; the N-terminal sequence and molecular mass determined by mass spectrometry were as expected.

(see Figure 1c), suggesting no major structural perturbation. NMR spectra of the C22S mutant were much improved, in comparison with the isolated wild-type A-domain, and the mutant was used in all the studies reported in this and the accompanying paper (Broadhurst *et al.*, 1995).

Assignment of the NMR Spectra. The strategy for the assignment of the resonances in the A-domain of HMG1 was broadly similar to that used for the ras p21.GDP protein (Kraulis *et al.*, 1994) and is not repeated in detail here. The backbone resonances of most residues were assigned using a combination of the HCANNH and HCA(CO)NNH 4D NMR spectra (Boucher *et al.*, 1992a,b). In ras p21.GDP, most side-chain resonances could be assigned from the 3D HCCH-COSY and TOCSY spectra (Bax *et al.*, 1990a,b). However, with the A-domain this was not possible due to significant overlap in the aliphatic region of the spectra of this highly α-helical protein. This problem was exacerbated by the many long side-chain lysine (17), proline (5), and arginine (5) residues. It was necessary to develop new 4D NMR experiments in which the side-chain (^{1}H and ^{13}C) resonances are resolved according to their backbone (^1H_N and ^{15}N) resonances via through-bond interactions. An illustration of the way in which these 4D NMR spectra simplify the assignment of the aliphatic ^{1}H and ^{13}C resonances of Trp 48 and Lys 49 is shown in Figure 2. The panel shows overlaid 2D ^{1}H−^{13}C planes at the ^1H_N and ^{15}N shifts of Lys 49. Cross-peaks (blue) from both side chains are seen in the HCCNNH experiment (Richardson *et al.*, 1993) whereas only the interresidue cross-peaks (green) are seen in the HCC(CO)NNH experiment (Clowes *et al.*, 1993; Logan *et al.*, 1992). Although these side-chain experiments are less sensitive than the HCCH-COSY and TOCSY experiments, the extra resolution made it possible to assign the side-chain resonances of most residues in this protein, even though it has a correlation time of ~10 ns [see accompanying paper, Broadhurst *et al.* (1995)]; however, the HCCH-COSY and TOCSY experiments were used in particular instances to check assignments.

As discussed in more detail previously (Kraulis *et al.*, 1994), with our 4D NMR strategy the assignment of glycine and proline residues requires extra experiments, because cross-peaks from glycine residues are not observed in the 4D NMR spectra when recorded with the usual parameters, and proline residues lack an amide proton. However, the new side-chain experiments were useful for the assignment of glycine residues although, in the N-terminus of the protein, NOESY experiments were also employed due to the large number of glycine and proline residues. Although the HCCH-COSY and TOCSY experiments, surprisingly, did not prove useful for assignment of proline residues in ras p21, we were able to use them for the A-domain of HMG1.

The assignment of aromatic residues was also carried out in a manner similar to that described for ras p21.GDP, although for the A-domain we found that the 4D ^{13}C/^{13}C-separated NOESY spectrum was useful for making complete assignments in the ring spin systems. An example of the strategy employed is given in Figure 3. The assignment begins with the identification of potential $^1H_\delta$ and $^{13}C_\delta$ shifts from the 4D ^{13}C/^{15}N-separated NOESY spectrum, in this case for Tyr 15 (Figure 3a). Once these had been obtained, the appropriate plane of the 4D ^{13}C/^{13}C-separated NOESY spectrum could be displayed (Figure 3b); usually the strongest peak was due to the $^1H_\epsilon$−$^{13}C_\epsilon$ cross-peak allowing that plane to be displayed etc. (Figure 3c). By checking the relative intensities in the 4D spectrum and the ^{1}H−^{1}H connectivities in the 2D COSY spectra, a consistent assignment could be made. The 4D ^{13}C/^{15}N-separated NOESY spectrum was also employed to assign the side-chain $-NH_2$ spin systems of asparagine and glutamine residues. Methionine residues were assigned using a strategy similar to that discussed previously (Kraulis *et al.*, 1994).

Analysis of the NOESY Spectra for Distance Restraints. The 2D NOESY spectrum and the 3D ^{13}C- and ^{15}N-separated NOESY spectra, as well as the 4D ^{13}C/^{13}C- and ^{13}C/^{15}N-separated NOESY spectra, were all useful for obtaining the necessary interresidue restraints for the structure determination and were analyzed as described previously (Kraulis *et al.*, 1994). Because the structure of the HMG box is likely to be sensitive to the correct calibration of distance restraints in the structure calculations, and particularly given the discrepancy between our structure of the B-domain of HMG1 (Weir *et al.*, 1993) and that originally reported by Read *et*

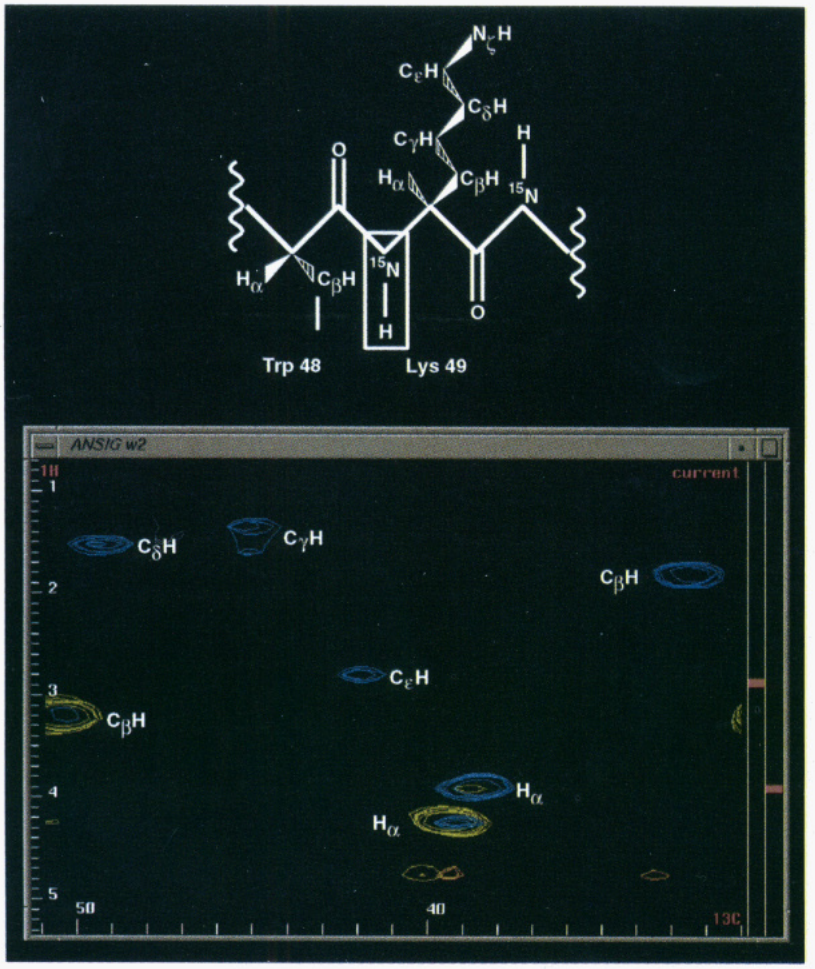

FIGURE 2: Illustration of the assignment of the side-chain ^1H and ^{13}C resonances in the aliphatic residues of the A-domain from HMG1. The panel shows the overlaid 2D planes ($f_1 = {}^1H$ and $f_2 = {}^{13}C$), at the $f_3 = {}^{15}N$ and $f_4 = {}^1H_N$ chemical shifts of Lys 49, from the HCCNNH (blue) and HCC(CO)NNH (green) 4D NMR spectra. The assignment strategy is discussed in the text.

al. (1993), we paid particular attention to this aspect of the work. The assigned NOESY cross-peaks were volume-integrated, by summing the values of the spectrum matrix data points within a box around the center of the cross-peak; the values for cross-peaks involving methyl resonances were divided by 3. Initially we took only unambiguous restraints that were free of the effects of overlap in either the 2D or, where necessary, the 3D ^{15}N-separated NOESY spectra. In the latter, there are fewer problems of differences in magnetization transfer than in ^{13}C- or $^{13}C/^{15}N$-separated NOESY spectra because $^1H-^{15}N$ couplings are uniform and the relevant nuclei in most residues have similar T_2 values. Although, in principle, faster rates of amide proton exchange can lead to weaker cross-peaks, and a more conservative estimate of the $^1H-^1H$ distance, we considered that at pH 5.0 this would not be a major problem. [This contrasts with the situation in the ^{13}C-separated NOESY spectra (see below).] The NOE intensities were converted into three distance-restraint classes: strong (0.0−2.7 Å), medium (0.0−3.3 Å), and weak (0.0−5.0 Å). Those from the 2D spectra were initially fitted to a build-up curve from a set of 2D NOESY spectra recorded with different mixing times; only restraints showing a good fit and an absence of effects of spin diffusion were used. The NOE intensities were calibrated separately for the two spectra (as well as for the 4D $^{13}C/^{15}N$-separated spectrum; see below) by reference to known distances (NH−NH and NH−C_αH) in well-defined regions of α-helix. Center averaging of degenerate or nonstereospecifically assigned protons (equivalent to the pseudoatom approximation) was used in the structure calculations, necessitating the addition of appropriate distance corrections to the upper bounds (Wüthrich *et al.*, 1983). We also computed structures using r^{-6} averaging which gave very similar results (data not shown), but in the presence of significant spin diffusion we believe center averaging is more appropriate (Kraulis *et al.*, 1994). The first list of NOE-derived constraints contained a total of 1011 distance restraints, of which 225 were long-range (i to $i \geq 5$). In the second step of the structure calculations (see below) further restraints were obtained from the 3D ^{13}C-separated, 4D $^{13}C/^{13}C$-separated, and 4D $^{13}C/^{15}N$-separated NOESY spectra. Again, only restraints based on unambiguous cross-peaks were used; the calibration of the 3D ^{13}C-separated and 4D $^{13}C/^{13}C$-separated NOESY spectra was based on the $C_\alpha H-C_\beta H$ (in Ala) and $C_\alpha H-C_\gamma H$ (in Thr) distances. Because of differences in the efficiency of magnetization transfer in these experiments, resulting from the different $^1H-^{13}C$ couplings and relaxation rates, looser restraints were used for cross-peaks analyzed in these spectra. In addition, an attempt to allow for effects of spin diffusion was made. A full relaxation matrix calculation of NOESY cross-peak intensities suggested that, for a protein with an overall correlation time of 10 ns, with a mixing time of 150 ms, significant NOEs (of similar intensity to some of those from protons

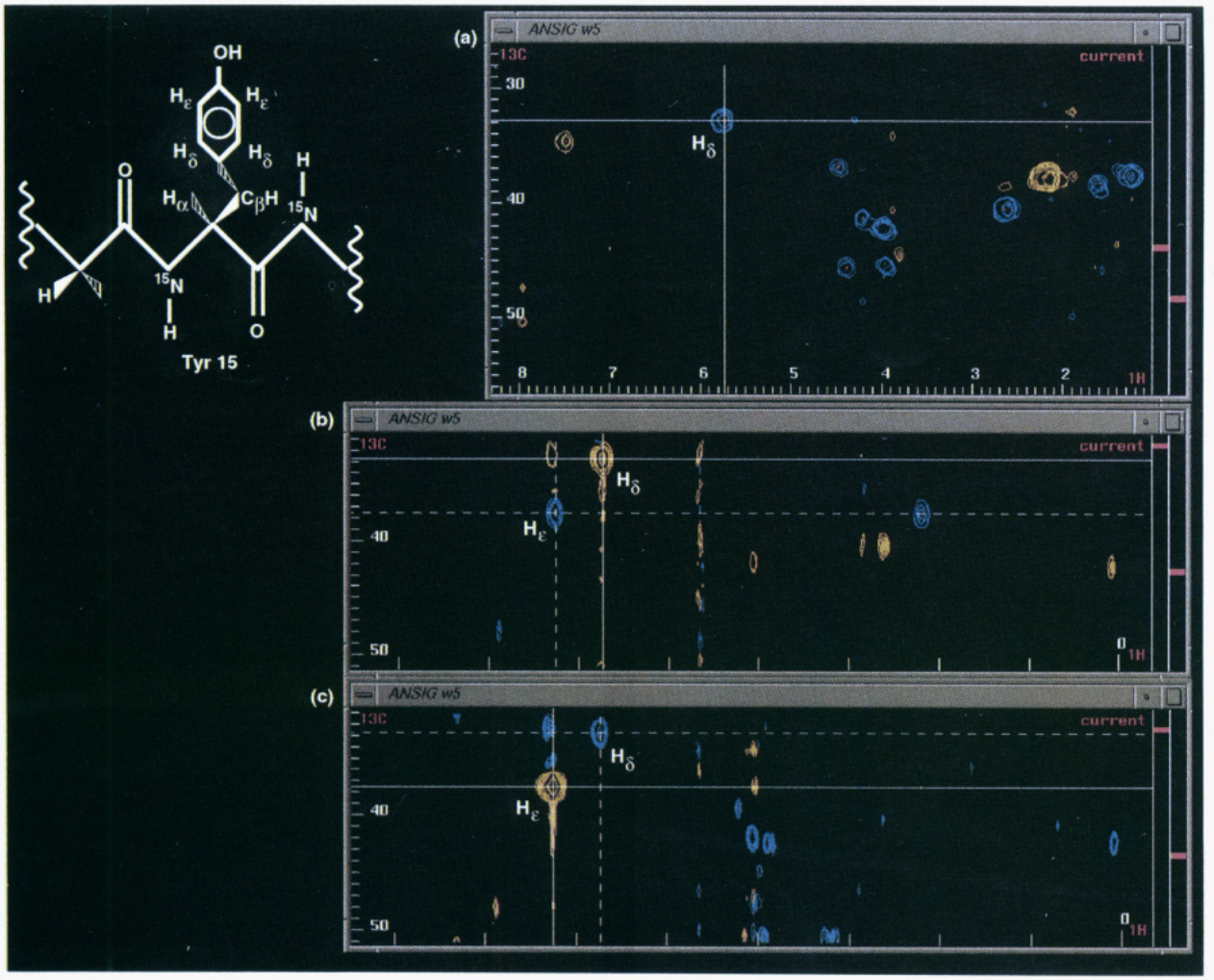

FIGURE 3: Illustration of the assignment strategy for the aromatic residues of the A-domain of HMG1. (a) A 2D plane (f_1 = ^{1}H and f_2 = ^{13}C), at the f_3 = ^{15}N and f_4 = ^1H_N chemical shifts of Tyr 15, from the 4D ^{13}C/^{15}N-separated NOESY spectrum (cross-peaks aliased an odd number of times are in blue, others are in orange). (b) The 2D plane (f_1 = ^{1}H and f_2 = ^{13}C), at the f_3 = ^{13}C and f_4 = ^{1}H chemical shifts identified from (a), from the 4D ^{13}C/^{13}C-separated NOESY spectrum. (c) The 2D plane (f_1 = ^{1}H and f_2 = ^{13}C), at the f_3 = ^{13}C and f_4 = ^{1}H chemical shifts identified from (b), from the 4D ^{13}C/^{13}C-separated NOESY spectrum. In the 4D ^{13}C/^{13}C-separated NOESY spectrum, cross-peaks aliased an odd number of times are also in blue, with others in orange.

3.0 Å apart), resulting from spin diffusion, could be observed for protons 7.0 Å apart. In order to avoid problems of bias, arising from too tight a classification of the restraints from these spectra, they were classified into three classes of 0.0−5.0, 0.0−6.0, and 0.0−7.0 Å. The final list of NOE-derived restraints, based purely on unambiguous cross-peaks, contained a total of 1659 restraints, of which 406 were intraresidue, 874 medium range (i to i < 5) and 379 long range (i to $i \geq 5$).

Other Restraints Used in the Structure Calculations. From inspection of the structures computed using the NOE restraints alone we identified 22 hydrogen bond restraints on the basis of slowly exchanging amide protons (residues 19, 21−24, 46−49, 58−69, and 72) where the acceptor atom was unambiguous; each ^1H_N hydrogen bond was represented by a single distance restraint of 1.8−2.5 Å from the amide proton to acceptor oxygen atom.

Structure Calculations. The initial structures were computed solely from the distance restraints derived from NOEs that could be assigned unambiguously in either the 2D or 3D ^{15}N-separated NOESY spectra. The calculations used the computer program X-PLOR (Brünger, 1992) and the YASAP protocol (M. Nilges, personal communication, and Brünger, 1990). They resulted in structures that displayed a unique fold and which had a root mean square deviation (RMSD) about the mean of 1.54 Å for the backbone atoms in the well-defined part of the molecule (see below and Table 2). In the second step, structures were calculated after addition of restraints that could be unambiguously assigned in the 3D ^{13}C-separated, 4D ^{13}C/^{13}C-separated, and 4D ^{13}C/^{15}N-separated NOESY spectra. A plot of the number of restraints per residue against the amino acid sequence is shown in Figure 4; there are 20 per residue on average, and they are fairly uniformly distributed along the sequence. The final set of structures was computed after addition of the hydrogen bond restraints.

3D Solution Structure. A total of 33 final structures that were consistent with the experimental restraints (see Table 1 for the structural statistics) were selected out of 60 computed (see Figure 5). The structures that were rejected all had higher energies, between one and eight NOE violations above 0.5 Å, and an unusual kink at the beginning of helix II. The final structures had no violations >0.5 Å and are well defined with an RMSD about the average structure of 0.62 Å for the backbone atoms and 1.07 Å for all the heavy atoms; in Table 2 these structures are compared

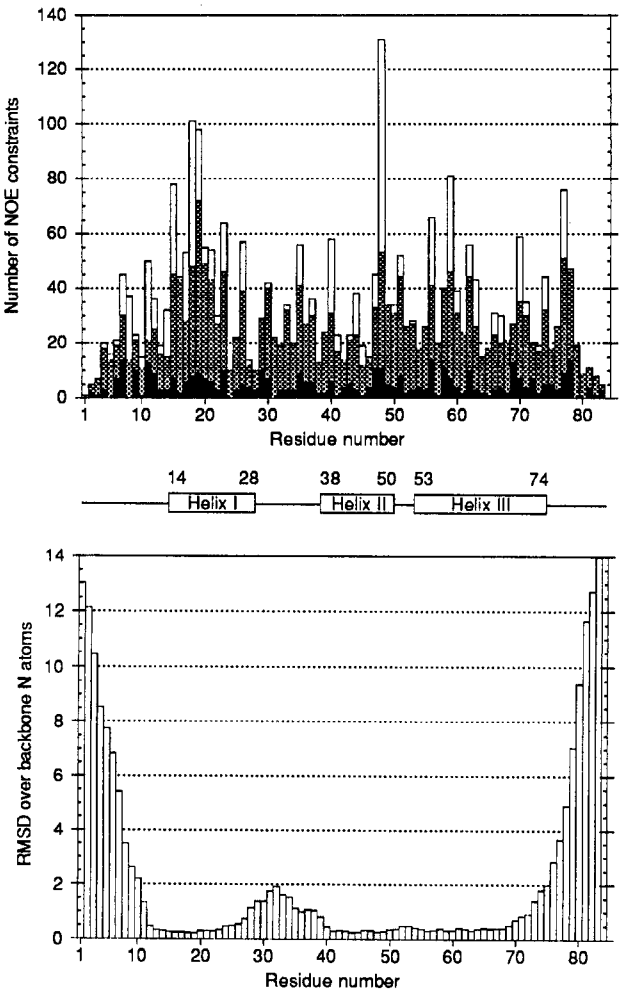

FIGURE 4: Plots of (upper panel) the number of experimental restraints [intraresidue, black; medium range (i to $i < 5$), shaded; and long range (i to $i \geq 5$), white] and (lower panel) the root mean square deviation of the backbone amide nitrogen atoms from the average structure, against the amino acid sequence.

Table 1: Structural Statistics[a]

structural statistic	⟨SA⟩	SA_{ctm}
RMS deviation from the experimental distance restraints (Å) (1659 restraints)	0.0458	0.0445
E_{L-J} (kcal/mol)	−301.6 ± 11.9	−295.6
RMS deviations from ideal covalent geometry		
bonds (Å)	0.0104	0.0105
angles (deg)	2.12	2.08
impropers (deg)	1.04	1.04

[a] The notation is as follows: ⟨SA⟩ are the average values for the final 33 simulated annealing (SA) structures; SA_{ctm} is the value for the structure closest to the mean structure; E_{L-J} is the Lennard-Jones van der Waals energy calculated with the CHARMm (Brooks *et al.*, 1983) force field (not included in the structure calculations).

with those computed from restraints obtained solely from the 2D and 3D ^{15}N-separated NOESY spectra. A plot of the RMSD of the backbone amide nitrogen atoms from the average structure (see Figure 4) and the superposition of the computed structures (see Figure 5a) shows that, in addition to the N- and C-termini, the region between residues 28 and 38 is not well defined (see below). If this region is excluded from the analysis, the corresponding RMSD values are reduced to 0.49 Å for the backbone atoms and 1.01 Å for all the heavy atoms (see Table 2). The positions of the side chains are in many cases well determined, particularly those in the core of the L-shaped structure (see Figure 5b). The stereochemistry of the structures, as judged by the ϕ and ψ angles in a Ramachandran plot is good (data not shown).

A schematic view of the structure is shown in Figure 5c. In the L-shaped structure, helices I and II pack against each other, whereas the N-terminal extended strand packs against helix III. The ill-defined region between residues 28 and 38 is seen to be the loop between helices I and II; in the accompanying paper we show that this loop is genuinely mobile on the sub-nanosecond time scale (Broadhurst *et al.*, 1995).

In the final structures we find that the hydroxyl group of Ser 22 (which replaced Cys 22 of the wild-type protein) is solvent-exposed, consistent with the fact that Cys 22 in the A-domain forms intermolecular disulfide bonds. Both intramolecular and intermolecular disulfide bonds are formed upon oxidation (data not shown), and examination suggests that a disulfide bond could be formed between Cys 22 and Cys 44 (buried) in the wild-type A-domain with minimal distortion. It seems rather surprising, therefore, that the reduced and oxidized forms of HMG1 have such different DNA- and histone H1-binding activities (Kohlstaedt & Cole, 1994a,b). (It is conceivable that in the oxidized form of HMG1 a disulfide bond forms between Cys 22 in the A-domain and Cys 106 in the B-domain, but we view this as less likely.)

Comparison with the Structures of Other HMG Boxes. We have compared the structure of the A- domain of HMG1 with that of the B-domain (Weir *et al.*, 1993) and that of the HMG box from HMG-D (Jones *et al.*, 1994). Figure 6a shows a comparison between the families of structures, and Figure 6b shows a comparison between the structures nearest to the average structure for each of the three families. Clearly the structures agree well. In particular, the global folds of the proteins are very similar. The pairwise RMSD of the C_α atoms (compared over the three α-helices for the structures nearest to the average structure from each of the families) for the A- and B-domains of HMG1 is 2.00 Å and for the A-domain and the HMG box of HMG-D is 1.69 Å (see Table 2). This gives us confidence that the structure of the B-domain we determined previously is correct and that it was indeed possible to determine the structure of such a non-globular protein using NMR spectroscopy, despite the fact that only essentially short-range restraints were used in the structure determination. The similarity in the three structures also suggests that there is no substantial flexibility between the two arms of the L-shaped structure as proposed by Falciola *et al.* (1994). This conclusion is borne out by the analysis of the ^{15}N relaxation data reported in the accompanying paper (Broadhurst *et al.*, 1995).

Comparison of the structures of the A- (this work) and B- (Weir *et al.*, 1993) domains of HMG1 indicates that the relative positions of the two arms in the L-shaped structure are better determined in the A-domain (see Figure 6a), although in general it is helix III and not the N-terminal strand that is better defined. Analysis of the NOE restraints and their subsequent classification in the structure calculations suggests some possible reasons for this. In A but not B, NOEs are observed between residue 12 and residues 19, 20, and 21, as well as between residue 17 and residues 66 and 67 (see Figure 7); it is possible that this results from the

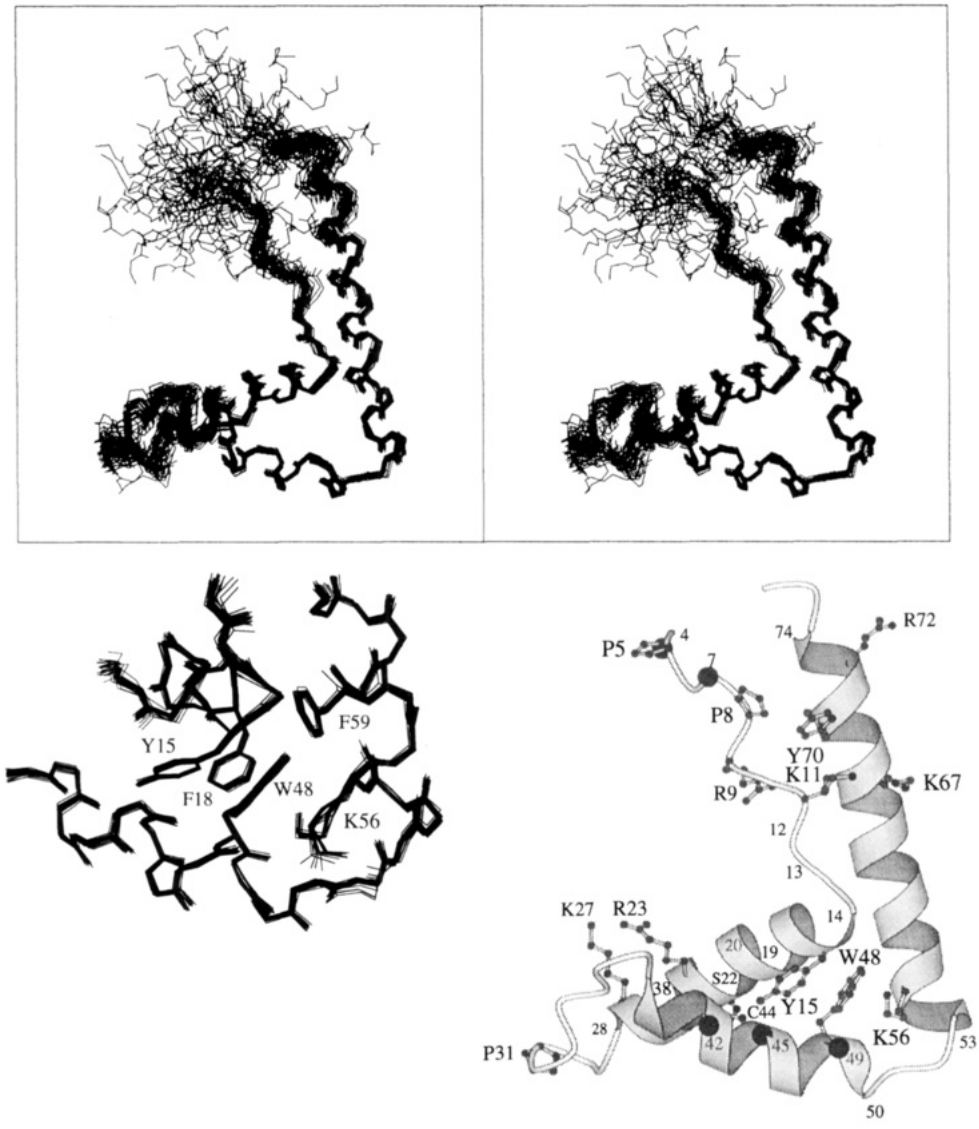

FIGURE 5: Structure of the A-domain from HMG1. (a, top) Superposition of the 33 final structures, showing the backbone, after a least squares fit over the C_α atoms in the helices (see Table 2). (b, bottom left) Superposition of some of the side chains in the core of the protein. (c, bottom right) Schematic representation of the structure; the backbone amides of residues whose $^1H/^{15}N$ resonances are most shifted ($\Delta > 100$, see Figure 8) on interaction with DNA are indicated by a solid dot. This figure and Figure 6 were generated using MOLSCRIPT (Kraulis, 1991).

Table 2: Atomic RMS Deviations[a]

	residues	backbone atoms (Å)	non-hydrogen atom
$\langle SA \rangle$ vs $\langle SA \rangle_{av}$[b]	11−74	2.16	2.64
	14−28, 38−50, 53−74	1.54	2.05
$\langle SA \rangle$ vs $\langle SA \rangle_{av}$[c]	11−74	0.62	1.07
	14−28, 38−50, 53−74	0.49	1.01

	residues	C_α atoms (Å)
SA_{ctm} vs HMG-B[d]	14−28, 38−50, 54−74	2.00
SA_{ctm} vs HMG-D[e]	14−28, 38−50, 54−74	1.69
HMG-B vs HMG-D	14−28, 38−50, 54−74	1.57

[a] The notation is as in Table 1, with the addition that $\langle SA \rangle_{ave}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other. [b] Refers to structures computed using structural restraints from homonuclear and ^{15}N-separated spectra only. [c] Refers to the final data including hydrogen bond restraints. [d] Refers to the structure of the B-domain from HMG1 (PDB codes 1HME and 1HMF; Weir *et al.*, 1993). [e] Refers to the structure of the HMG box from HMG-D (PDB code 1HMA; Jones *et al.*, 1994).

fact that we have been able to observe and assign more weak NOEs in the 3D/4D ^{13}C-separated NOESY spectra of the A-domain. Nonetheless, the weaker classification of restraints from these spectra (0.0−5.0, 0.0−6.0, and 0.0−7.0 Å) may also be responsible for the poorer definition of the N-terminal strand in the A-domain, because they were used to obtain most of the restraints between helix III and the N-terminal strand.

When the families of structures of the A- and B-domains of HMG1 are compared with each other and with the HMG box from HMG-D, other differences are also evident. First, the orientation of helix I, in particular the C-terminus relative to the rest of the helix, is different in the A-domain when compared with the B-domain and HMG-D (see Figure 6a). This difference is strongly supported by studies of the backbone dynamics and is discussed in more detail in the accompanying paper (Broadhurst *et al.*, 1995). Second, although the angle between helices I and III is very similar, the angle between helices II and III is somewhat smaller in the A-domain than in the B-domain (by ~14°) or HMG-D (by ~9°). This difference is less noticeable if the angle of only those residues in the elbow (or core) of the structure

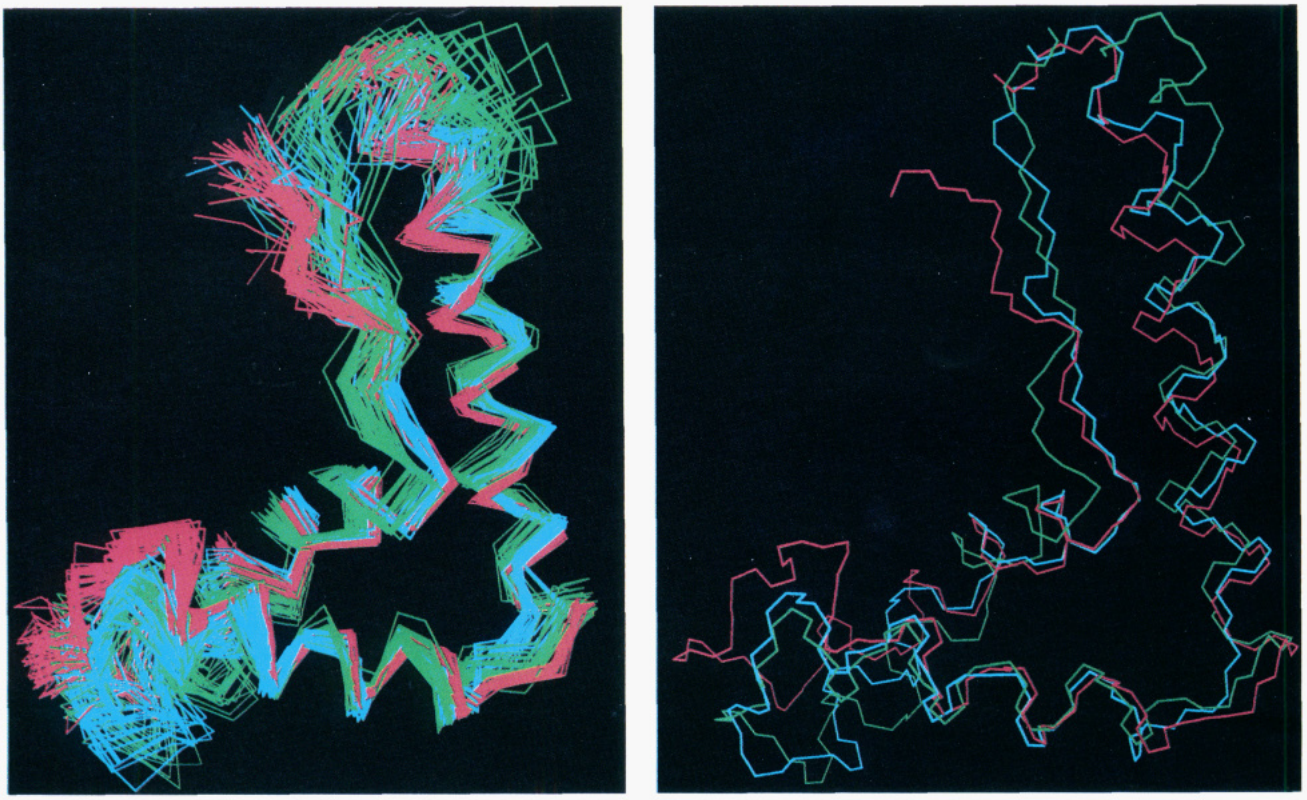

FIGURE 6: Comparison of the structure of the A-domain of HMG1 with that of the B-domain (Weir *et al.*, 1993) and that of the HMG box of HMG-D (Jones *et al.*, 1994), after a least squares fit over the C_α atoms in the helices (see Table 2). (a, left) Superposition of the three families of structures. (b, right) Superposition of the structure nearest to the average structure from each of the three families.

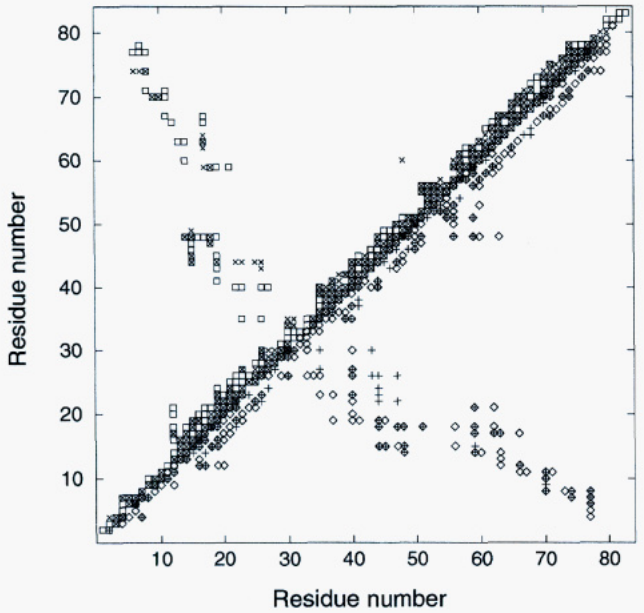

FIGURE 7: Plot illustrating the NOEs observed in the A- and B-domains of HMG1. Main-chain−main-chain or main-chain−side chain (represented by ◇ and +) and side-chain−side-chain (represented by □ and ×) NOEs are plotted for the A- and B-domains, respectively.

(as opposed to the entire helices) are compared. All the structures are also similar to a recent refinement of another structure determination of the B-domain (Read *et al.*, 1995). As discussed above, we believe that the sharper angle between helices II and III may result from the better definition of helix III in the structure of the A-domain. Finally, although others (Adzhubei *et al.*, 1995) have suggested that there is evidence for a type II polyproline helix in the N-terminal strand of both the B-domain (residues 9,10, 11, and 13) and HMG-D (residues 6, 8, 9, and 11), this is not the case in the A-domain. This could well reflect the lack of proline residues at equivalent positions in the A-domain (residues 10 and 11 are proline in the B-domain but lysine and methionine in the A-domain), although again it may also reflect the poorer definition of the N-terminal strand in the A-domain. Overall, the A-domain shows some differences from the B-domain and HMG-D which are rather similar (Jones *et al.*, 1994). These differences may prove to be important in the slightly different DNA-binding properties of the A- and B-domains (Teo *et al.*, 1995a), the most distinctive of which is the ability to discriminate between four-way DNA junctions ("static cross-overs") and competitor supercoiled DNA (which contains "dynamic cross-overs"). (The A-domain is resistant to competition whereas the B-domain, even with a basic C-terminal extension, is not, possibly because it forms more stable interactions on supercoiled DNA.)

Interaction of the A-Domain of HMG1 with DNA. HMG1 binds to double-stranded DNA without sequence specificity, and it binds preferentially to distorted or bent DNA structures (e.g., bulges or four-way junction DNA; see the introduction). However, as a simple starting point for NMR experiments we have deliberately chosen to study a short oligonucleotide (14-mer), on the assumption that binding to duplex DNA and four-way junctions have many features in common, the preference for the junctions arising from a predistorted DNA structure. Strong support for this notion has recently come from NMR studies and mutagenesis of the SRY HMG box (Peters *et al.*, 1995). After some preliminary work, the

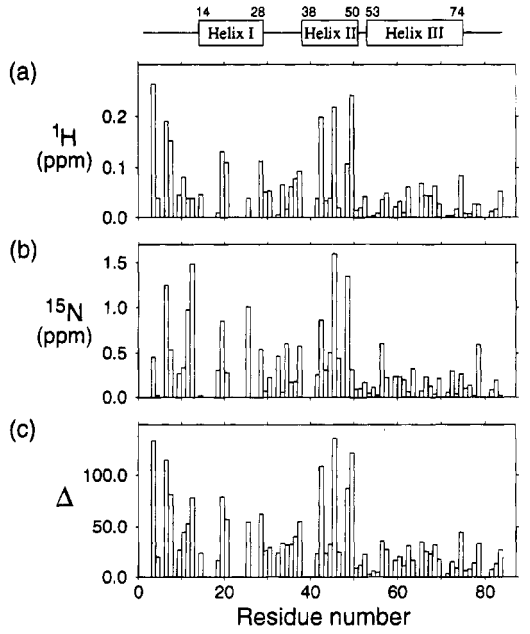

FIGURE 8: Plots of the chemical shift differences, in the $^1H-^{15}N$ HMQC spectrum, between the A-domain and its complex with a 14-mer DNA duplex (sequence 5′ GCTATAAAAGGGCA 3′). (a) Plot of the unsigned ^1H chemical shift differences. (b) Plot of the unsigned ^{15}N chemical shift differences. (c) Plot of the square root of the sum of the squares of the differences in the ^1H and ^{15}N chemical shifts: $\Delta = [(\delta_H \times 500.13)^2 + (\delta_N \times 50.68)^2]^{1/2}$.

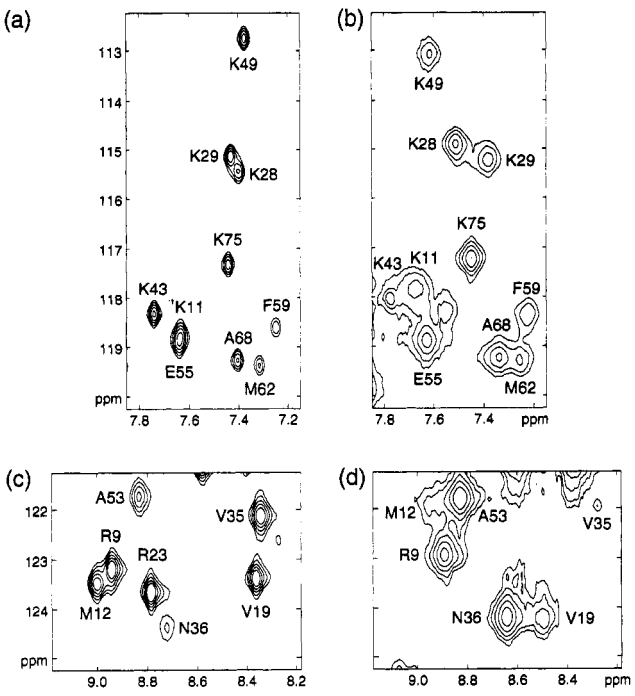

FIGURE 9: Comparison of regions of the $^1H-^{15}N$ HMQC spectra of the A-domain (a and c) and its complex with the 14-mer DNA duplex (b and d).

oligonucleotide that we chose to study further was a 14-mer consisting of the TBP-binding site, with flanking and stabilizing GC base pairs (Kim, J. L., *et al.*, 1993). Previous work had suggested a similarity in the mode of DNA binding, and hence bending, by HMG boxes and TBP (see the introduction), and it seemed possible that the A-domain might bind stably to the 14-mer sequence. Gel retardation assays (Figure 1b) showed that the 14-mer gave a shifted complex at similar protein concentrations for both the wild-type protein and the C22S mutant used for the structure determination.

Titration of DNA into protein, and *vice versa,* showed that the free and bound forms of the protein were in fast exchange on the NMR time scale. We were therefore able to follow changes in the $^1H-^{15}N$ HMQC spectrum with increasing amounts of DNA up to a 1:1 (mol/mol) mixture. In order to minimize problems arising from the binding of more than one protein molecule to a single DNA duplex, we also carried out a titration in the reverse direction and monitored changes in the DNA spectrum. The final $^1H-^{15}N$ HMQC spectra of the complex in the two cases were very similar. In Figure 8, the changes in $^1H/^{15}N$ chemical shifts in the protein on formation of the complex are plotted against the amino acid sequence; plots of two representative regions of the $^1H-^{15}N$ HMQC spectra are shown in Figure 9. The changes occur mainly in the N-terminal extended strand and helices I and II (see Figures 8, 9, and 5c), the largest shifts being in the N-terminal strand (residues 4 and 7) and the C-terminus of helix II (residues 42, 45, and 49). Significant broadening of those peaks that are most shifted on complex formation suggests that either the free and bound forms, or perhaps different bound forms, of the protein are in intermediate exchange on the NMR time scale. Interestingly, large shifts are not seen for helix III.

Various lines of evidence for HMG1 and other HMG-box proteins all point to the involvement of the concave face of the HMG box, in particular the longer arm consisting of the N-terminal strand and helix III, in DNA binding, at least to duplexes. These include mutagenesis studies (Giese *et al.*, 1991; Harley *et al.*, 1992; Faciola *et al.*, 1994; Teo *et al.*, 1995b), a "domain swap" experiment (Read *et al.*, 1994), and consideration of the HMG-box structure (Weir *et al.*, 1993; Read *et al.*, 1993; Jones *et al.*, 1994; this work), together with recent NMR studies on SRY (Peters *et al.*, 1995). In particular, the domain swap experiment showed that the N-terminal strand and helix III of TCF-1a, when combined with helices I and II of the B-domain of HMG1, was sufficient for sequence-specific DNA-binding activity. The large chemical shift changes that we observe for the N-terminal strand of the A-domain on binding to DNA, and the absence of appreciable shifts in helix III, suggest that it is the N-terminal strand that is mainly involved in the interaction. This is entirely consistent with mutational analysis. For example, the residues (A-domain numbering) at position 8 in the A-domain, LEF-1 and SRY (Teo *et al.*, 1995b; Giese *et al.*, 1991; Harley *et al.*, 1992), position 9 in the A-domain (Falciola *et al.*, 1994), positions 9 and 10 in LEF-1 (Giese *et al.*, 1991), and position 12 in LEF-1 and SRY (Giese *et al.*, 1991; Harley *et al.*, 1992) are all implicated in DNA binding.

A combination of NMR studies and site-directed mutagenesis of SRY (King & Weiss, 1993; Haqq *et al.*, 1994; Peters *et al.*, 1995) and HMG-D (Churchill *et al.*, 1995) suggests that residues 15 and 16 in the N-terminus of helix I (A-domain numbering) recognize a core element (TTG) in the DNA-binding site. Peters *et al.* (1995) suggest further that the intercalation of residue 16 at the TT base step is important for bending DNA and, therefore, for sequence-specific recognition of duplex DNA, but not for binding to intrinsically bent DNA molecules such as four-way junctions.

The large shifts that we see at the C-terminus of helix II suggest that this may be an important site of interaction in the binding to DNA, consistent with immunological evidence that helix II (in contrast to helices I and III) is buried in the *Chironomus* HMG1 protein−DNA complexes (Wisniewski *et al.*, 1994). In particular, the pattern of changes in the chemical shifts in helix II, with larger shifts for residues three and four apart in the sequence, suggests changes on one face of the helix due to its interaction with the DNA (see Figure 5c). Site-directed mutagenesis is also consistent with these results. The residues (A-domain numbering) at positions 45 in SRY (Harley *et al.*, 1992) and 56 in LEF-1 and SRY (Giese *et al.*, 1991; Harley *et al.*, 1992; Pontiggia *et al.*, 1994) are both implicated in binding to DNA duplexes. However, these interactions may not be as important in binding to DNA four-way junctions; mutations at position 45 have no effect on binding of either the A-domain or SRY (Falciola *et al.*, 1994; Pontiggia *et al.*, 1994), and mutation at position 56 has no effect on SRY binding (Pontiggia *et al.*, 1994). Recently, in further studies of site-directed mutants of the A domain, we have found that some residual structure remains in a W → R mutant at position 48 that is substantially unfolded (Teo *et al.*, 1995b). This mutant is still able to bind selectively to the four-way junction DNA (Falciola *et al.*, 1994) without refolding (Teo *et al.*, 1995b). It is tempting to speculate that in the W → R mutant some secondary structure is retained by mutual stabilization between the N-terminus of helix I and the C-terminus of helix II.

Thus collectively, mutational studies, the domain swap, and immunological experiments, as well as structural studies using NMR, have suggested that HMG proteins bind in the minor groove of DNA via the N-terminal strand, the N-terminus of helix I, and the C-terminus of helix II. The domain swap experiment (Read *et al.*, 1994) suggests that the N-terminal strand/helix III are responsible for making the sequence-specific contacts, whereas helices I and II mediate nonspecific DNA binding. It seems likely that the binding to four-way junctions and duplex DNA will be fundamentally similar, with additional interactions required for the distortion of B-form double-stranded DNA.

After this paper was submitted, the structures of the LEF-1 and SRY HMG box−DNA complexes were reported (Love *et al.*, 1995; Werner *et al.*, 1995). The nature of the interaction with DNA of these sequence-specific HMG boxes is very similar to that predicted for the structure-specific A domain from HMG1 (this work). In both structures the protein binds to a widened minor groove of the DNA via the concave face of the HMG box, as we suggested previously (Weir *et al.*, 1993) and as reported more recently for the SRY (King & Weiss, 1993; Peters *et al.*, 1995) and HMG-D (Churchill *et al.*, 1995) HMG boxes. The DNA is bent about the N-terminal strand and helix II (which are buried in the minor groove), providing a structural basis for DNA bending by these transcription factors. The structures of the LEF-1 and SRY−DNA complexes are generally similar, differing mainly toward the C-terminus of the protein where, in LEF-1, an adjacent basic region binds across a narrowed major groove contributing to DNA recognition. [The C-terminal part of this basic region is missing in the SRY fragment studied by Werner *et al.* (1995). In addition, a somewhat smaller DNA fragment was used for the structure determination of the SRY−DNA complex, suggesting that the differences between the LEF-1 and SRY−DNA complexes may reflect, to some extent, the particular fragments of the proteins/DNA binding sites studied.] It was suggested (Love *et al.*, 1995) that the differences in DNA-binding activity between the two families of HMG-box proteins, sequence- (e.g., LEF-1, SRY) and structure- (e.g., HMG1) specific, may be due to disruption of helix III in the sequence-specific proteins by a proline (which would, if present, be at position 75 in the A-domain); in LEF-1 this proline produces a kink in helix III, allowing the basic C-terminal region of the protein to make extensive contacts with DNA (Love *et al.*, 1995). This proline and the adjacent basic region are highly conserved in the sequence-specific proteins but are not present in the structure-specific class of HMG boxes (Ner, 1992). However, similar basic regions, which do contribute to DNA binding (Wisniewski & Schulze, 1994; Teo *et al.*, 1995a), are present on the C-terminal side of both the A and B HMG boxes (in the linker regions between the A/B and B/C domains) of HMG1. Nevertheless, our results for the A-domain of HMG1, in which we do not find evidence for strong interactions between helix III of the HMG box and the DNA, are entirely consistent with the proposal made by Love *et al.* (1995).

ACKNOWLEDGMENT

We thank Dr. R. Cortese for plasmid pRNHMG1, Dr. M. Nilges for protocols for the simulated annealing calculations, Dr. S. Neidle and Prof. C. Crane-Robinson for preprints of Adzhubei *et al.* (1995) and Read *et al.* (1995), respectively, J. Jacoby for amino acid analysis, M. Weldon for DNA synthesis, J. Lester for DNA sequencing, Drs. W. Boucher and R. T. Clowes for processing many of the NMR spectra, and Dr. P. J. Kraulis for help with the assignment.

SUPPORTING INFORMATION AVAILABLE

A virtually complete list of NMR assignments (^{13}C, ^{15}N, and ^{1}H) (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Adzhubei, A. A., Laughton, C. A., & Neidle, S. (1995) *Protein Eng.* (in press).

Bagby, S., Harvey, T. S., Eagle, S. G., Inouye, S., & Ikura, M. (1994) *Structure 2*, 107−122.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., & Kay, L. E. (1990a) *J. Magn. Reson. 87*, 620−627.

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Magn. Reson. 88*, 425−431.

Baxevanis, A., & Landsman, D. (1995) *Nucleic Acids Res. 23*, 1604−1613.

Bianchi, M. E., Beltrame, M., & Paonessa, G. (1989) *Science 243*, 1056−1059.

Bianchi, M. E., Falciola, L., Ferrari, S., & Lilley, D. M. J. (1992) *EMBO J. 11*, 1055−1063.

Boucher, W., Laue, E. D., Campbell-Burk, S., & Domaille, P. J. (1992a) *J. Am. Chem. Soc. 114*, 2262−2264.

Boucher, W., Laue, E. D., Campbell-Burk, S. L., & Domaille, P. J. (1992b) *J. Biomol. NMR 2*, 631−637.

Broadhurst, R. W., Hardman, C. H., Thomas, J. O., & Laue, E. D. (1995) *Biochemistry 34*, 16608−16617.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

Brünger, A. T. (1990) *X-PLOR Manual V2.1*, Yale University, New Haven, CT.

Brünger, A. T. (1992) *X-PLOR Manual V3.0*, Yale University, New Haven, CT.
Bustin, M., Lehn, D. A., & Landsman, D. (1990) *Biochim. Biophys. Acta 1049*, 231−243.
Churchill, M. E. A., Jones, D. N. M., Glaser, T., Hefner, H., Searles, M. A., & Travers, A. A. (1995) *EMBO J. 14*, 1264−1275.
Clore, G. M., Kay, L. E., Bax, A., & Gronenborn, A. M. (1991) *Biochemistry 30*, 12−18.
Clowes, R. T., Boucher, W., Hardman, C. H., Domaille, P. J., & Laue, E. D. (1993) *J. Biomol. NMR 3*, 349−354.
Faciola, L., Murchie, A. I. H., Lilley, D. M. J., & Bianchi, M. E., (1994) *Nucleic Acids Res. 22*, 285−292.
Ge, H., & Roeder, R. G. (1994) *J. Biol. Chem. 269*, 17136−17140.
Giese, K., Amsterdam, A., & Grosschedl, R. (1991) *Genes Dev. 5*, 2567−2578.
Grosschedl, R. Giese, K., & Pagel, J. (1994) *Trends Genet. 10*, 94−100.
Haqq, C. M., King, C.-Y., Ukiyama, E., Falsafi, S., Haqq, T. N., Donahoe, P. K., & Weiss, M. A. (1994) *Science 266*, 1494−1500.
Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkowitz, G. D., Sockanathan, S., Lovell-Badge, R., & Goodfellow, P. N. (1992) *Science 255*, 453−456.
Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659−4667.
Jantzen, H.-M., Admon, A., Bell, S. P., & Tjian, R. (1990) *Nature 344*, 830−836.
Johns, E. W. (1982) *The HMG Chromosomal Proteins*, Academic Press, London.
Jones, D. N. M., Searles, M. A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., & Neuhaus, D. (1994) *Structure 2*, 609−627.
Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M., (1990) *Science 249*, 411−414.
Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993) *Nature 365*, 520−527.
Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993) *Nature 365*, 512−520.
King, C.-Y., & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11990−11994.
Kohlstaedt, L. A., & Cole, R. D. (1994a) *Biochemistry 33*, 570−575.
Kohlstaedt, L. A., & Cole, R. D. (1994b) *Biochemistry 33*, 12702−12707.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kraulis, P. J., Domaille, P. J., Campbell-Burk, S. L., Van Aken, T., & Laue, E. D. (1994) *Biochemistry 33*, 3515−3531.
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
Landsman, D., & Bustin, M. (1993) *BioEssays 15*, 539−546.
Laue, E. D., Mayger, M. R., Skilling, J., & Staunton, J. (1986) *J. Magn. Reson. 68*, 14−29.
Leblanc, B., Read, C., & Moss, T. (1993) *EMBO J. 12*, 513−525.
Locker, D., Decoville, M., Maurizot, J. C., Bianchi, M. E., & Leng, M. (1995) *J. Mol. Biol. 246*, 243−247.
Logan, T. M., Olejniczak, E. T., Xu, R. X., & Fesik, S. W. (1992) *FEBS Lett. 314*, 413−418.
Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., & Wright, P. E. (1995) *Nature 376*, 791−795.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 1515−1517.
Marion, D., Ikura, M., & Bax, A. (1989b) *J. Magn. Reson. 84*, 425−430.
Neidhardt, F. C., Bloch, P. L., & Smith, D. F. (1974) *J. Bacteriol. 119*, 736−747.
Ner, S. (1992) *Curr. Biol. 2*, 208−210.
Oñate, S. A., Prendergast, P., Wagner, J. P., Nissen, M., Reeves, R., Pettijohn, D. E., & Edwards, D. P. (1994) *Mol. Cell. Biol. 14*, 3376−3391.
Paull, T. T., Haykinson, M. J., & Johnson, R. C. (1993) *Genes Dev. 7*, 1521−1534.
Peters, R., King, C.-H., Ukiyama, E., Falsafi, S., Donahoe, P. K., & Weiss, M. A. (1995) *Biochemistry 34*, 4569−4576.
Pil, P. M., & Lippard, S. J. (1992) *Science 256*, 234−237.
Pil, P. M., Chow, C. S., & Lippard, S. J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9465−9469.
Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R., & Bianchi, M. E. (1994) *EMBO J. 13*, 6115−6124.
Rance, M. (1987) *J. Magn. Reson. 74*, 557−564.
Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., & Norman, D. G. (1993) *Nucleic Acids Res. 21*, 3427−3436.
Read, C. M., Cary, P. D., Preston, N. S., Lnenicek-Allen, M., & Crane-Robinson, C. (1994) *EMBO J. 13*, 5639−5646.
Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., Carrillo, M. O. M., & Norman, D. G. (1995) *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) Vol. 9, pp 222−250, Springer, New York.
Richardson, J. M., Clowes, R. T., Boucher, W., Domaille, P. J., Hardman, C. H., Keeler, J., & Laue, E. D.(1993) *J. Magn. Reson., Ser. B 101*, 223−227.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
Sklenár, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469−479.
Stelzer, G., Goppelt, A., Lottspeich, F., & Meisterernst, M. (1994) *Mol. Cell. Biol. 14*, 4712−4721.
Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074−1078.
Teo, S.-H., Grasser, K. D., & Thomas, J. O. (1995a) *Eur. J. Biochem. 230*, 943−950.
Teo, S.-H., Grasser, K. D., Hardman, C. H., Broadhurst, R. W., Laue, E. D., & Thomas, J. O. (1995b) *EMBO J. 14*, 3844−3853.
van de Wetering, M., Oosterwegel, M., van Norren, K., & Clevers, H. (1993) *EMBO J. 12*, 3847−3854.
Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., & Thomas, J. O. (1993) *EMBO J. 12*, 1311−1319.
Werner, M. H., Huth, J. R., Gronenborn, A. M., & Clore, G. M. (1995) *Cell 81*, 705−714.
Wisniewski, J. R., & Schulze, E. (1994) *J. Biol. Chem. 269*, 10713−10719.
Wisniewski, J. R., Ghidelli, S., & Steuernagel, A. (1994) *J. Biol. Chem. 269*, 29261−29264.
Wolfe, S. A., Ferentz, A. E., Grantcharova, V., Churchill, M. E. A., & Verdine, G. L (1995) *Chem. Biol. 2*, 213−221.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W., & Fesik, S. W. (1990) *J. Magn. Reson. 86*, 210−216.
Zuiderweg, E. R. P., Petros, A. M., Fesik, S. W., & Olejniczak, E. T. (1991) *J. Am. Chem. Soc. 113*, 370−372.
Zwilling, S., König, H., & Wirth, T (1995) *EMBO J. 14*, 1198−1208.

BI951405C