Nucleic Acids Research, 1993, Vol. 21, No. 15 3427–3436

Solution structure of a DNA-binding domain from HMG1

Christopher M.Read, Peter D.Cary, Colyn Crane-Robinson*, Paul C.Driscoll[1] and David G.Norman[1,+]

Biophysics Laboratories, School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DT and [1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Received April 30, 1993; Revised and Accepted June 15, 1993

ABSTRACT

We have determined the tertiary structure of box 2 from hamster HMG1 using bacterial expression and 3D NMR. The all α-helical fold is in the form of a V-shaped arrowhead with helices along two edges and one rather flat face. This architecture is not related to any of the known DNA binding motifs. Inspection of the fold shows that the majority of conserved residue positions in the HMG box family are those involved in maintaining the tertiary structure and thus all homologous HMG boxes probably have essentially the same fold. Knowledge of the tertiary structure permits an interpretation of the mutations in HMG boxes known to abrogate DNA binding and suggests a mode of interaction with bent and 4-way junction DNA.

INTRODUCTION

The HMG box is a conserved ~80 amino acid domain that mediates the DNA binding of many proteins that are known or presumed transcription factors. Phylogenetic analysis distinguishes two subfamilies of proteins: one subfamily as exemplified by HMG1 (1) and UBF1 (2), typically contain multiple HMG boxes, and the other subfamily as exemplified by SRY (3), TCF1 (4) and LEF-1 (5) typically contain single boxes embedded in a large protein (6).

Considerable variation is shown in the specificity of DNA binding by HMG box proteins. In the HMG1 and UBF1 subfamily, the box(es) bind to DNA in a relatively non-sequence specific manner. UBF1 binds to a range of GC-rich segments in the promoter of rRNA genes (2). The yeast protein, ABF2 shows both phased sequence-specific DNA binding and non-specific binding (7). Other HMG boxes apparently lack DNA sequence specificity and bind to the distorted DNA of 4-way junctions (eg. HMG1 (8) or its individual boxes (9)), or to cisplatin-modified DNA (eg. HMG1 (10) and SSRP1 (11)). In the SRY, TCF1 and LEF-1 subfamily, the box binds to a specific AT-rich DNA sequence that includes the segment 5′-(A/T)(A/T)CAAAG-3′, making contacts primarily in the minor groove and inducing a considerable bend in the DNA (12–15). Recently, the HMG box from SRY has also been shown to bind 4-way junction DNA (15). The DNA binding of HMG boxes thus ranges from the sequence-specific to the structure-specific, the unifying feature being that DNA bound by an HMG box is always in a highly bent or otherwise distorted form. All HMG boxes may possess features which allow them to recognise already distorted DNA, whilst some HMG boxes are also capable of binding to linear DNA with the induction of considerable distortion. A brief overview of the DNA binding of HMG boxes has been given by Lilley (16).

The chromosomal high mobility group protein HMG1 (1) contains 2 related HMG boxes (17) linked to an acidic C-terminal tail through a short basic segment. The function of HMG1 is at present unclear. The activation by HMG1 in *in vitro* transcription assays suggests a role as a transcription factor (18–20). Indeed the trout homologue of mammalian HMG1, HMG T (21) binds to an AT-rich sequence in its own promoter that has a high potential for cruciform formation (22). The binding of HMG1 to 4-way junction DNA (8) however suggests a role in DNA recombination. The demonstration that HMG1 binds to cisplatin-modified DNA (10) also suggests a function in the repair of chemically modified and distorted DNA.

We have now determined the structure of box 2 from HMG1 in order to further define the HMG box fold and as a first step toward understanding the specificity of the protein-DNA interactions in this large family of proteins.

MATERIALS AND METHODS

Construction of HMG1 box 2 expression plasmid pchHMG1/5

The DNA sequence of HMG1 box 2 (residues 1–79 in our numbering system) was amplified by PCR using as template a partial cDNA clone of chinese hamster HMG1 (plasmid pCH1 (23)), and primers of sequence: 5′-CGGCGCGGGATCC-AATGCACCCAAGAGGCCTC-3′ and 5′-GCGCGAATT-CTTACGCTGCATCGGGTTTTCC-3′.

BamH1 and EcoR1 cut amplified DNA was ligated into BamH1 and EcoR1 cut pGEX-2T vector DNA (24) and transformed into *E.coli* DH5α to give pchHMG1/5. Dideoxynucleotide sequencing

* To whom correspondence should be addressed

+ Present address: Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK

of both strands confirmed the correct inserted DNA sequence. This plasmid was also transformed into the prototrophic *E.coli* strain BL21.

Purification of HMG1 box 2 protein

E.coli DH5α transformed with pchHMG1/5 were grown overnight and then inoculated 1:10 into fresh ampicillin-containing L-broth and grown for 3 hr. (to $OD_{600nm}=1.0$). IPTG was added to 0.1 mM and cells grown for a further 3 hr. Cells were centrifuged down and resuspended at 1g wet cell pellet per 4ml of 150 mM NaCl, 20 mM sodium phosphate buffer (pH 7.3), 10 mM β-mercaptoethanol (βME), 1 mM EDTA, 1 mM PMSF, 1 mM benzamidine. Cells were lysed by sonication and debris spun down at 10,000×g for 10 min. at 4°C. The supernatant was added batchwise to glutathione-agarose beads (Sigma) and rolled for 30 min. at 4°C. The beads were then washed 3× with 150 mM NaCl, 20 mM sodium phosphate buffer (pH 7.3), 10 mM βME and 2× with 150 mM NaCl, 50 mM Tris HCl (pH 8.0), 10 mM βME.

Cleavage of the fusion protein whilst still attached to the glutathione-agarose beads was achieved by the addition of $MgCl_2$ and $CaCl_2$ to 2 mM each and human plasma thrombin (75 units/litre of culture). HMG1 box 2 was eluted from a column built from the digest, using 150 mM NaCl, 50 mM Tris HCl (pH 8.0), 10 mM βME.

HMG1 box 2 was purified by passage through a 50 ml DEAE-Sephadex column in 150 mM NaCl, 10 mM $NaHCO_3$, 10 mM βME and by gel-filtration on a 180 ml Sephadex G50 column in the same buffer. Protein was desalted on a 50 ml Sephadex G15 column in 2 mM potassium phosphate buffer (pH 6.0). Ten litres of bacterial culture yielded ~50 mg of purified HMG1 box 2.

For the production of uniformly labelled ^{15}N HMG1 box 2 fragment, pchHMG1/5 (in BL21) was cultured in ampicillin containing minimal M9 medium containing 9.7 mM $^{15}NH_4Cl$ as sole nitrogen source. Cells were grown to $OD_{600nm}=1.0$ prior to IPTG induction and then grown for a further 4 hr. before harvesting. Protein was purified as described. Ten litres of bacterial culture yielded ~25 mg of ^{15}N-labelled HMG1 box 2.

Gel retardation assay

Junction **z**, radiolabelled with ^{32}P at one 5'-end was prepared according to (15). At a concentration of 20 nM the junction was mixed with either 100, 200 or 1000 nM HMG1 box 2, in a buffer containing 10% Ficoll 400, 10mM Hepes (pH 7.9), 10 mM NaCl, 5mM KCl, 1 mM $MgCl_2$, 1 mM spermidine, and optionally 0.5 mM dithiothreitol (DTT) (10 μl final volume). After incubation on ice for 10 min., samples were analysed in a 7% polyacrylamide gel containing 0.5× TBE buffer. The dried gel was autoradiographed at −80°C with an intensifying screen.

Sample preparation and NMR measurments

2D ^1H NMR spectra were recorded at 600 MHz and 3D ^{15}N-^1H spectra at 500 MHz with sample concentrations of 4.8 mM in 45 mM potassium phosphate (pH 5.46) at 297 K. Solutions were made up in either 100% D_2O or 90% H_2O/10% D_2O and the pH adjusted (uncorrected glass-electrode readings shown). 2D ^1H nuclear Overhauser enhancement (NOESY) spectra (25, 26) were collected in a phase-sensitive manner by the time proportional phase increment method. ^1H homonuclear Hartmann-Hahn (HOHAHA) spectra (27, 28) were collected in reverse mode with transfer of magnetization by the WALTZ17 mixing sequence (29) and a mixing time of 50 ms. In the case of spectra recorded in 90% H_2O/10% D_2O, a 'jump return' sequence (38) was used to supress the water signal. In spectra recorded in 100% D_2O, irradiation of the residual water signal was carried out during the relaxation delay or mixing time, and low power irradiation during t_1. Receiver phase was adjusted to limit base-line distortion (31). Deconvolution of the free induction decay, prior to Fourier transformation in F_2, using a Gaussian function (32) was used to reduce t_2 ridges in the fully transformed spectrum. 3D ^1H nuclear Overhauser enhancement ^{15}N-^1H heteronuclear multiple quantum coherence (NOESY-HMQC) and ^1H homonuclear Hartmann-Hahn ^{15}N-^1H HMQC (HOHAHA-HMQC) experiments (33, 34) were acquired as 128×32×512 complex points using spectral widths of 5102.04, 963.38 and 6024.09 Hz for the F_1, F_2 and F_3 dimensions, respectively. A mixing time of 200 ms was employed in the NOESY-HMQC experiment. Slowly exchanging amide protons were identified by 2D ^{15}N-^1H heteronuclear single-quantum coherence (HSQC) NMR experiments. The dried sample was redissolved in cold D_2O at pH 5.91, 285 K and spectra obtained at intervals after dissolution. To confirm the assignments of the peaks under these conditions additional spectra were obtained at 288, 292 and 297 K. $^3J_{NH\alpha}$ spin-spin coupling constants were determined by line-shape fitting to traces of peaks from a 2D ^{15}N-^1H HMQC-J experiment (35). All data were processed using the FELIX II software package (Hare Research Inc.).

RESULTS AND DISCUSSION

Expression of HMG1 box 2

Box 2 of hamster HMG1 (23) was expressed using the pGEX system in *E.coli* to produce a fusion protein with gluathione S-transferase, from which the HMG box was isolated by proteolysis with thrombin (see Materials and Methods). The expressed fragment represents the 79 residues from N92 to A170 (numbering system of human HMG1, (36); all mammalian HMG1s have almost identical sequences. With respect to the 71 residue 'minimum HMG box' (see Figure 1), the expressed fragment contained one less residue at the N-terminus and nine more residues at the C-terminus. The isolated fragment also contained two additional N-terminal residues (Gly-Ser) from the pGEX-2T expression vector and is thus 81 residues in total length.

Sample characterisation

The purified HMG1 box 2 fragment was homogeneous as judged by SDS-PAGE electrophoresis and reverse phase HPLC (data not shown). Analysis of the CD spectrum indicated a secondary structure content of ~56% α-helix and no β-sheet, as expected from measurements on the native protein and its proteolytic fragment LF (37). The molecular weight was determined by low-speed sedimentation equilibrium since it has been reported that individual boxes from HMG1 form homodimers (9). At a concentration of 1 mg/ml a molecular mass of 11,200 D was obtained from a $\ln C/r^2$ plot. This value is only slightly greater than the expected molecular mass and demonstrates that the expressed box does not associate strongly. We therefore expected to be able to solve the structure of the box as a monomer.

Electrospray mass spectrometry of samples following NMR indicated a single species of molecular mass 9044.23 ± 1.52 D, a value 74.63 D larger than the mass calculated from the amino acid composition (8969.60 D). Treatment of this sample with 2 mM DTT followed by further mass spectrometry showed

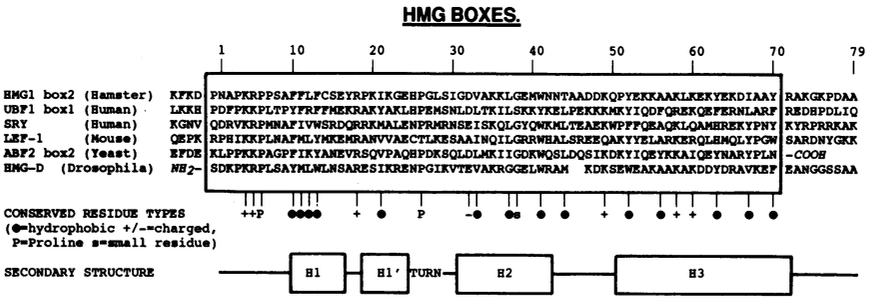

Figure 1. Comparison of the sequence studied here (top line, residues 1 to 79) with 5 other selected HMG boxes. The approximate boundaries of the HMG box have been previously defined using deletion mutants in DNA binding assays (7, 9) and from comprehensive sequence homology compilations (57). The enclosed sequences represent the 71 residue 'minimum HMG box'; the N-terminus being taken as the first residue of HMG-D and the C-terminus as the last residue of ABF2 box 2. The conserved residue types are assessed qualitatively from more complete compilations of HMG box sequences (57 and C. C-R., unpublished). The secondary structure shown is that determined in the present work. UBF1 is one of two human genes coding for Upstream Binding Factor, a protein that contains six HMG boxes in tandem flanked by an N-terminal dimerisation domain and an acidic C-terminal segment (2, 47). SRY (Sex Region of the Y chromosome) is the human gene product responsible for male sex development (3) and binds *in vitro* within the promoters of male specific genes (58). LEF-1 (Lymphoid Enhancer Binding Factor 1), is a mouse protein that binds to the enhancer of the T-cell receptor α gene (5). These two proteins both contain a single HMG box. The yeast ARS (Autonomously Regulating Sequence) binding factor 2 protein, ABF2 (7) contains two HMG boxes and interacts with the ARS1 sequence of mitochondrial DNA. HMG-D from drosophila contains a single HMG box and an acidic C-terminal segment (48, 57).

a single species of 8969 D. The sample used for NMR therefore contained a molecule of βME attached to the single cysteine at position 14. As a further check 20 cycles of Edman degradation confirmed the N-terminal amino acid sequence of the fragment.

A gel retardation assay was used to test the DNA binding of the HMG1 box 2 fragment, both as the βME adduct and in the fully reduced form. We used the 4-way DNA junction **z** (15) in which the four strands (of 30 nucleotides) contain sequences not related to those recognised by the sequence-specific HMG boxes. Junction **z**, at a concentration of 20 nM, was mixed with several concentrations of HMG1 box 2 and analysed by polyacrylamide gel electrophoresis. At the higher protein to DNA ratios a band shift corresponding to the box binding to junction **z** was observed for both forms (Figure 2). This shift is similar to that described for the binding of HMG1 and its fragments to other 4-way junctions (8, 9, 15). The expressed box 2 fragment thus represents a DNA-binding domain, in which the presence of the βME adduct reduces the affinity for junction DNA by a small amount.

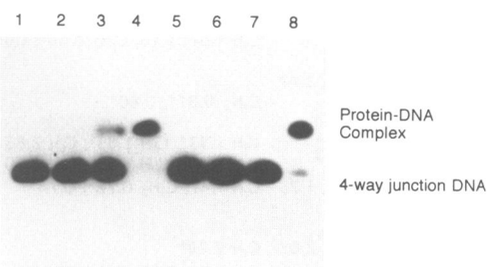

Figure 2. Gel retardation assay of the binding of HMG1 box 2 to 4-way junction DNA at 20 nM. Lanes 1–4 contain 0, 100, 200, 1000 nM of added DTT-reduced box and lanes 5–8 the same concentrations of the box with the βME adduct.

NMR assignment of HMG1 box 2

Box 2 of hamster HMG1 was produced by bacterial expression in normal and ^{15}N labelled medium for 2D and 3D NMR spectroscopy. The sequential assignment of amino acid spin systems was made by comparison of strips from the amide region of 3D NOESY-HMQC and HOHAHA-HMQC NMR spectra (34, 38). Side-chain assignments were obtained using 2D HOHAHA and NOESY spectra of unlabelled samples both in H_2O and D_2O, in conjunction with the 3D-NMR spectra. Complete spin-system identification of amino acid side-chains was obtained for 67 of the 81 residues of the protein. Table 1 lists the proton resonance assignments obtained.

Sequential assignment was achieved by identifying stretches of spin systems connected by ^{15}NH-^{15}NH sequential nuclear Overhauser effects (NOEs). Assignment of these segments to specific residues in HMG1 box 2 was firstly achieved for the longest segments which contained amides for which spin system information was available (principally the glycine, alanine and aromatic residues). Breaks in the sequential assignment due to proline residues were identified from sequential $C_\alpha H$-^{15}NH and $C_\delta H$-^{15}NH NOE connectivities. Three breaks occurred in the sequential assignment due to a pair of amides having similar chemical shifts (between residues 30 and 31, 35 and 36 and 74 and 75). These were resolved from their $C_\alpha H$-^{15}NH sequential connectivities. In just one case, that between residues A72 and K73, both the ^{15}NH resonances and $C_\alpha H$-^{15}NH NOEs were almost identical in chemical shift. Sequential assignment in this case was determined from side chain information. We were thus able to account for all ^{15}NH peaks in the 3D NMR spectra and obtain a complete sequential assignment of all residues in the box.

Location of secondary structure elements

Analysis of the short range NOE connectivities $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$, that are indicative of α-helical secondary structure (39), provided direct evidence for four α-helical segments located between residues F10 to E16 (helix 1), P19 to E24 (helix 1'), G31 to N42 (helix 2) and P51 to R71 (helix 3), see Figure 3A. The measured $^3J_{NH\alpha}$ spin-spin coupling constants, the location of slowly exchanging peptide NH protons (Figure 3A) and the chemical shifts of $C_\alpha H$ protons (40) were largely consistent with this secondary structure.

The C-terminal segment between residues 72 and 79 was clearly in an extended conformation since strong sequential $d_{\alpha N}$

Table 1. Proton resonance assignments of HMG1 Box 2 at 297K and pH 5.46

Residue	NH	$C_\alpha H$	$C_\beta H$	Others
Asn 1	8.54	4.67	2.70, 2.63	NH_2 7.51, 6.86
Ala 2	8.01	3.47	0.93*	
Pro 3		4.23	2.25, 1.75	$C_\gamma H$ 1.65, 1.55; $C_\delta H$ 3.15, 2.91
Lys 4	8.37	4.17	1.62, 1.55	$C_\gamma H$ 1.34,1.37; $C_\delta H$ -; $C_\varepsilon H$ 2.87
Arg 5	8.21	3.29	1.40, 1.36	$C_\gamma H$ 1.43,1.28; $C_\delta H$ 2.95*;$N_\varepsilon H$ -
Pro 6		4.53	1.98, 1.92	$C_\gamma H$ 1.76*; $C_\delta H$ 2.84, 2.62
Pro 7		4.25	1.92, 1.84	$C_\gamma H$ 1.83*; $C_\delta H$ 3.55, 3.46
Ser 8	7.63	4.33	4.30, 3.78	
Ala 9	8.85	3.88	1.49*	
Phe 10	8.28	2.57	2.26*	$C_\delta H$ 6.26*; $C_\varepsilon H$ 7.01*; $C_\zeta H$ 7.01
Phe 11	7.87	3.52	3.09*	$C_\delta H$ 6.97*; $C_\varepsilon H$ 7.22*; $C_\zeta H$ 7.20
Leu 12	8.02	3.94	2.37, 1.65	$C_\gamma H$ 1.87; $C_\delta H_3$ 0.94*, 0.86*
Phe 13	7.70	3.61	2.37*	$C_\delta H$ 6.64*; $C_\varepsilon H$ 7.13*; $C_\zeta H$ 6.67
Cys 14	8.87	3.52	2.43, 1.82	βME CHs 3.36, 3.32; 2.25, 2.19
Ser 15	7.97	3.73	3.60*	
Glu 16	7.14	3.88	1.93, 1.83	$C_\gamma H$ 1.58*
Tyr 17	7.73	4.03	2.50*	$C_\delta H$ 6.45*; $C_\varepsilon H$ 6.67*
Arg 18	8.79	3.85	1.80*	$C_\gamma H$ 1.37*; $C_\delta H$ 3.04, 2.95; $N_\varepsilon H$ -
Pro 19		4.14	2.15, 1.70	$C_\gamma H$ 1.85*; $C_\delta H$ 3.41, 3.37
Lys 20	6.88	4.00	1.86*	$C_\gamma H$ 1.37*; $C_\delta H$ -; $C_\varepsilon H$ -
Ile 21	7.92	3.93	1.88	$C_\gamma H$ 1.50, 1.35; $C_\gamma H_3$ 0.85*; $C_\delta H_3$ 0.77*
Lys 22	8.18	3.90	1.73, 1.62	$C_\gamma H$ 1.35*; $C_\delta H$ 1.46*; $C_\varepsilon H$ 2.87*
Gly 23	7.59	3.73*		
Glu 24	7.41	3.96	1.80, 1.57	$C_\gamma H$ 2.24, 1.92
His 25	7.69	4.93	3.06, 2.99	$C_{\delta 2} H$ 7.17; $C_{\varepsilon 1} H$ 8.39
Pro 26		4.46	2.27, 1.87	$C_\gamma H$ 1.89, 1.87; $C_\delta H$ 3.48, 3.26
Gly 27	8.65	3.92, 3.70		
Leu 28	7.31	4.35	1.64, 1.46	$C_\gamma H$ 1.75; $C_\delta H_3$ 0.67*, 0.67*
Ser 29	9.15	4.34	4.19, 4.00	
Ile 30	8.61	3.73	1.74	$C_\gamma H$ 1.52, 1.16; $C_\gamma H_3$ 0.86*; $C_\delta H_3$ 0.79*
Gly 31	8.56	3.79*		
Asp 32	7.86	4.46	2.79, 2.47	
Val 33	8.44	3.35	2.13	$C_\gamma H_3$ 0.87*, 0.80*
Ala 34	7.78	3.99	1.44*	
Lys 35	7.93	3.99	1.80*	$C_\gamma H$ 1.33*; $C_\delta H$ 1.63*; $C_\varepsilon H$ 2.85
Lys 36	7.85	4.08	1.80*	$C_\gamma H$ 1.40*; $C_\delta H$ 1.53*; $C_\varepsilon H$ 2.70
Leu 37	8.48	4.27	2.11, 1.73	$C_\gamma H$ 1.47; $C_\delta H_3$ 0.78*, 0.74*
Gly 38	8.25	3.83, 3.74		
Glu 39	7.85	3.98	2.12, 2.03	$C_\gamma H$ 2.28*
Met 40	8.46	3.96	2.31*	$C_\gamma H$ 2.68, 2.48; $C_\varepsilon H$ 1.80*
Trp 41	8.66	3.78	3.09*	$C_{\delta 1} H$ 6.79; $C_{\varepsilon 3} H$ 5.48; $C_{\zeta 3} H$ 6.35; $C_{\eta 2} H$ 6.94; $C_{\zeta 2} H$ 7.38; $N_{\varepsilon 1} H$ 10.01
Asn 42	7.99	4.14	2.74, 2.70	NH_2 7.45, 6.70
Asn 43	7.41	4.60	2.88, 2.54	NH_2 7.45, 6.84
Thr 44	7.14	3.96	3.61	$C_\gamma H_3$ 1.04*
Ala 45	9.04	4.01	1.41*	
Ala 46	8.69	3.65	1.31*	
Asp 47	8.83	4.28	2.56, 2.40	
Asp 48	7.19	4.69	2.72, 2.63	
Lys 49	7.65	4.28	1.62, 1.56	$C_\gamma H$ 0.65,-0.40; $C_\delta H$ 0.84, 0.78; $C_\varepsilon H$ 1.30*
Gln 50	7.32	4.25	2.21, 2.14	$C_\gamma H$ 2.54, 2.41; $N_\delta H_2$ 7.50, 6.80
Pro 51		4.12	2.17, 1.23	$C_\gamma H$ 1.95,1.55 ; $C_\delta H$ 3.67, 3.61
Tyr 52	7.09	4.12	3.55, 3.40	$C_\delta H$ 7.36*; $C_\varepsilon H$ 7.10*
Glu 53	8.05	4.15	2.20, 2.08	$C_\gamma H$ 2.40, 2.26
Lys 54	9.00	4.04	1.79, 1.75	$C_\gamma H$ 1.51*; $C_\delta H$ 1.55*; $C_\varepsilon H$ 2.85*
Lys 55	7.57	4.00	1.82*	$C_\gamma H$ 1.35*; $C_\delta H$ -; $C_\varepsilon H$ -
Ala 56	8.37	3.88	1.52*	
Ala 57	8.22	4.16	1.50*	
Lys 58	7.82	4.12	1.95, 1.88	$C_\gamma H$ 1.51*; $C_\delta H$ -; $C_\varepsilon H$ 2.88*
Leu 59	8.12	4.12	1.88, 1.38	$C_\gamma H$ 1.85; $C_\delta H_3$ 0.86*, 0.83*
Lys 60	8.28	3.97	2.03*	$C_\gamma H$ 1.41*; $C_\delta H$ 1.68*; $C_\varepsilon H$ -
Glu 61	7.97	3.98	2.13, 2.07	$C_\gamma H$ 2.37*
Lys 62	7.61	3.98	1.88*	$C_\gamma H$ 1.36*; $C_\delta H$ 1.58*; $C_\varepsilon H$ -
Tyr 63	7.97	4.30	3.14, 3.07	$C_\delta H$ 6.96*; $C_\varepsilon H$ 6.50*
Glu 64	8.39	3.67	2.06, 1.93	$C_\gamma H$ 2.58, 2.34
Lys 65	7.57	4.00	1.88*	$C_\gamma H$ 1.38*; $C_\delta H$ 1.57*; $C_\varepsilon H$ -
Asp 66	8.66	4.33	2.70, 2.37	
Ile 67	8.93	3.88	1.33	$C_\gamma H$ 1.00*; $C_\gamma H_3$ 0.63*; $C_\delta H_3$ 0.62*
Ala 68	7.25	4.03	1.40*	
Ala 69	7.62	4.10	1.39*	
Tyr 70	7.94	4.12	3.03*	$C_\delta H$ 6.98*; $C_\varepsilon H$ 6.67*
Arg 71	8.11	3.89	1.74*	$C_\gamma H$ 1.36*; $C_\delta H$ 1.58*; $N_\varepsilon H$ -
Ala 72	7.56	4.10	1.33*	
Lys 73	7.60	4.11	1.72, 1.65	$C_\gamma H$ 1.34*; $C_\delta H$ 1.50*; $C_\varepsilon H$ 2.81*
Gly 74	7.96	3.76*		
Lys 75	7.87	4.48	1.70, 1.57	$C_\gamma H$ 1.30,0.71; $C_\delta H$ 1.05*; $C_\varepsilon H$ 4.13*
Pro 76		4.30	2.16*	$C_\gamma H$ 1.87*; $C_\delta H$ 3.67, 3.53
Asp 77	8.29	4.42	2.58, 2.51	
Ala 78	7.97	4.19	1.27*	
Ala 79	7.78	3.99	1.22*	

(*) indicates protons with degenerate chemical shifts.
(−) indicates a resonance that was not observed or could not be assigned.

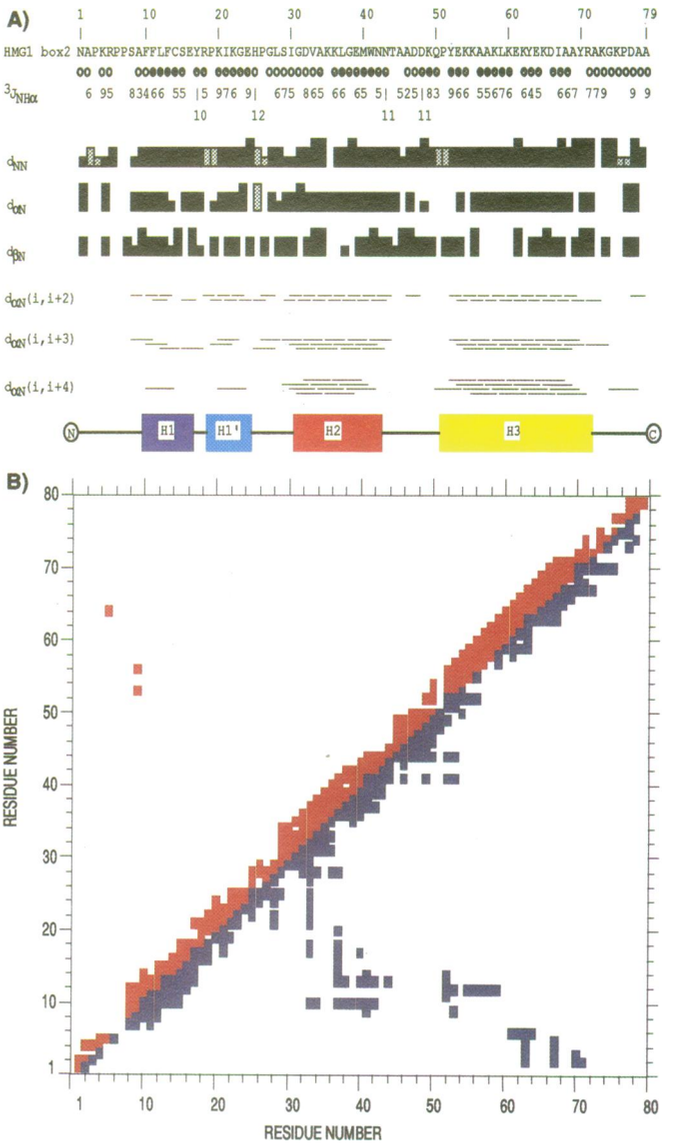

Figure 3. A) Secondary structure deduced from sequential and medium range NOEs. The top line shows the amino acid sequence of HMG1 box 2. Open, hatched and filled circles below the sequence represent fast, medium, and slowly exchanging amide protons as identified in 2D ^{15}N-^1H HMQC experiments. $^3J_{NH\alpha}$ coupling constants (to the nearest Hz) are shown below the appropriate residue in the HMG1 box 2 sequence. Solid bars indicate the sequential d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$ connectivities, where the height of the bar reflects the NOE intensity (3 classes); for connectivities involving a proline residue, the $C_\delta H$ protons have been used in place of the NH proton and are shown as hatched bars. Short and medium range NOE connectivities $d_{\alpha N}(i, i+2)$, $d_{\alpha N}(i, i+3)$ and $d_{\alpha N}(i, i+4)$ are represented by lines between the two residues. Also shown is the location of the four helices identified in the present work. **B)** Summary of observed sequential, medium and long range NOE contacts for each residue. Each filled square represents at least one distance restraint between backbone protons (above diagonal, red) or between backbone-sidechain and sidechain-sidechain protons (below diagonal, blue).

and d_{NN} connectivities were observed but medium range connectivities were mostly absent. There was also evidence of flexibility in this region as judged from narrow line-widths and rapid exchange of peptide NH protons.

Determination of tertiary structure

Detailed inspection of the 2D and 3D NOESY spectra enabled 1109 inter-residue contacts, of which 346 were long range, to

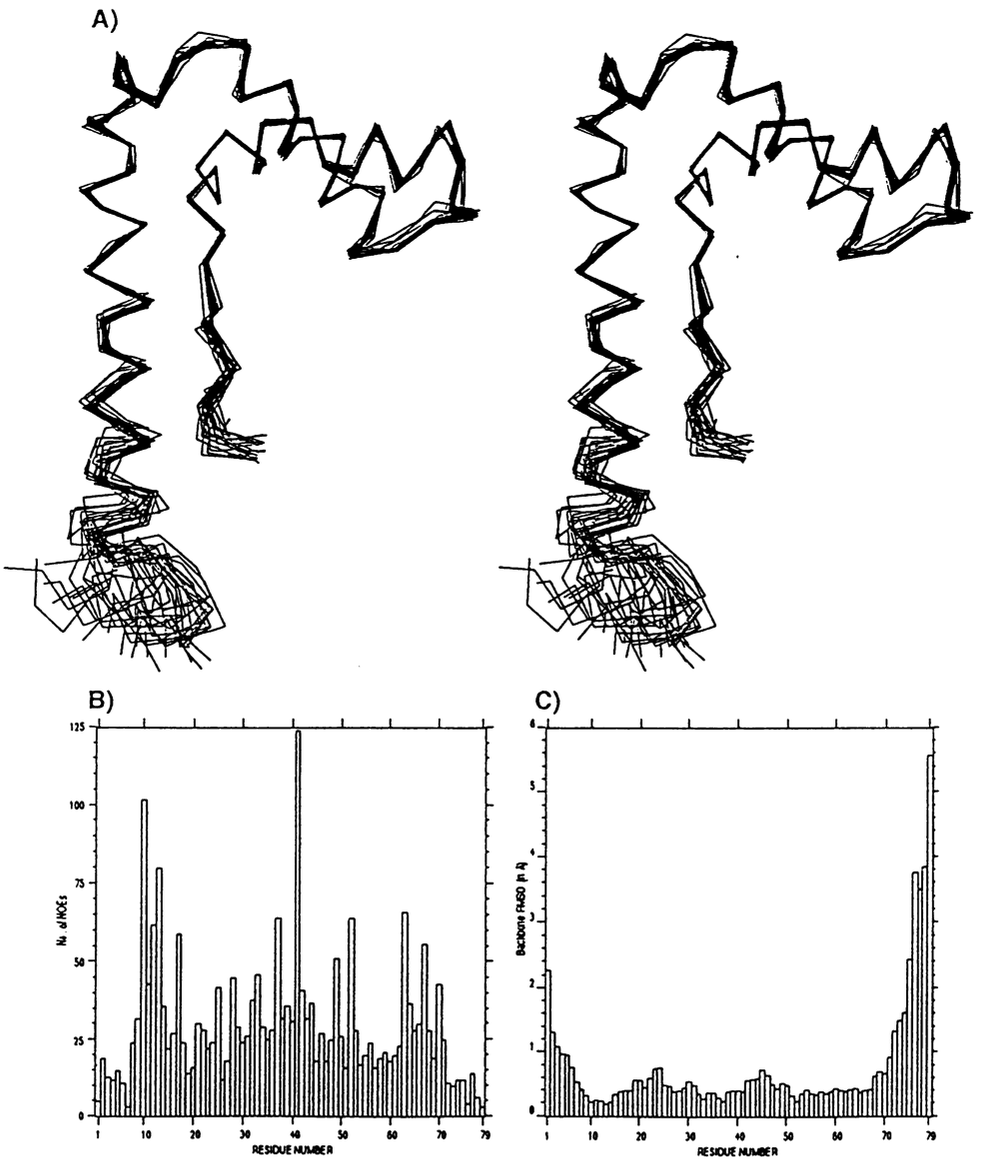

Figure 4. A) Superposition of the backbone (between residues 1 and 72) of the 30 final structures of HMG1 box 2, shown in stereoview. **B)** Plot of the number of NOE distance restraints used in the structure determination against residue. **C)** Plot of the backbone r.m.s. deviations (in Å) versus residue for the family of structures shown in **A)**. The final energy values for these 30 structures were: F_{total}= 2157.6 (SD= 62.98) kJ mole^{-1}; F_{NOE}= 417.9 (SD= 13.67) kJ mole^{-1}; F_{VDW}= −1344.5 (SD= 31.06) kJ mole^{-1}; $F_{dihedral}$= 1726.5 (SD= 47.93) kJ mole^{-1}; F_{angle}= 1228.7 (SD= 20.39) kJ mole^{-1}; $F_{improper}$= 33.1 (SD= 3.83) kJ mole^{-1}; $F_{Cdihedral}$= 0.0 (SD= 0.00) kJ mole^{-1}. RMS deviations from experimental distance restraints were 0.089 Å (all); 0.082 Å (intra-residue); 0.087 Å (short range); 0.097 Å (long range) and 0.00014 Å (H-bonds).

be identified from the cross-peaks. A diagonal plot summarising these sequential, medium and long range contacts is shown in Figure 3B. Almost all of the long range contacts are concentrated into three areas and represented contacts: 1) between the N-terminal segment and C-terminal half of helix 3; 2) between helix 1 and the N-terminus of helix 3, and 3) between helices 1 and 1′ to helix 2. This showed that the N-terminal segment up to residue A9 lies close to helix 3 and runs alongside it in an anti-parallel manner. Likewise, helix 1 lies in close proximity to both helices 2 and 3. Helix 1′ contacts only helix 2. Remarkably though, the only NOE contact between helices 2 and 3 was that of the aromatic side chain of residue W41, located near the C-terminal end of helix 2, with residues Y52 and E53 at the N-terminal end of helix 3 (see Figure 3B). Thus helices 2 and 3 are not close to each other, but both are in close proximity to helix 1.

The intensities of the NOE cross-peaks were assigned to three classes: strong <2.75 Å, medium <3.75 Å, and weak <5.25 Å, to yield a set of distance restraints. Since the NOESY spectra were obtained with a mixing time of 200 ms, very weak NOE cross-peaks were assumed to result from spin diffusion between protons ≤6.0 Å apart and were included as an additional distance class. Three dimensional structures were then calculated from these experimental restraints with the program XPLOR v3.0 (41). Initial structures were generated using a dynamic simulated annealing protocol YASAP (42) starting from randomly generated backbone conformations with extended side-chains. This produced a backbone structure containing α-helices only. An iterative approach was then employed in which more and more distance restraints were incorporated to define and improve the tertiary structure. The final structures were calculated from 1228 NOE distance restraints of which 119 were intra-residue, 763

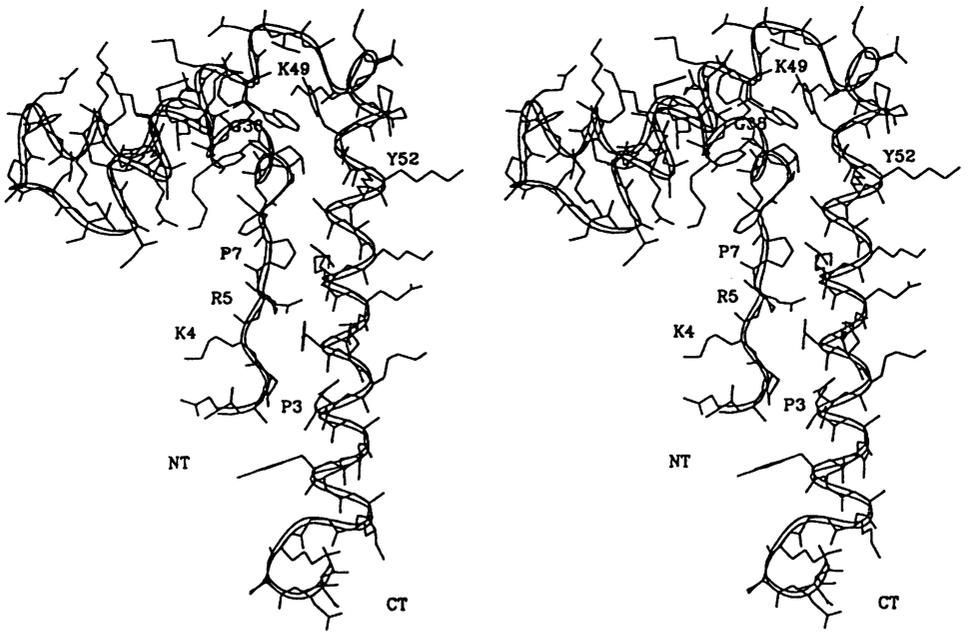

Figure 5. Stereo-view line plot showing sidechain and mainchain conformations of the coordinate averaged structure of HMG1 box 2. The numbers represent residues which when mutated in SRY and LEF-1 lead to loss of DNA binding. The βME adduct can be seen extending into the solvent space between the wings of the fold. NT= N-terminus, CT= C-terminus. The final energy values for the averaged structure were: F_{total}= 2239.9 kJ mole^{-1}; F_{NOE}= 381.29 kJ mole^{-1}; F_{VDW}= −1297.2 kJ mole^{-1}; $F_{dihedral}$= 1813.8 kJ mole^{-1}; F_{angle}= 1215.8 kJ mole^{-1}; $F_{improper}$= 31.8 kJ mole^{-1}; $F_{Cdihedral}$= 0.0 kJ mole^{-1}. RMS deviations from experimental distance restraints were 0.086 Å (all); 0.079 Å (intra-residue); 0.084 Å (short range); 0.094 Å (long range) and 0.00 Å (H-bonds).

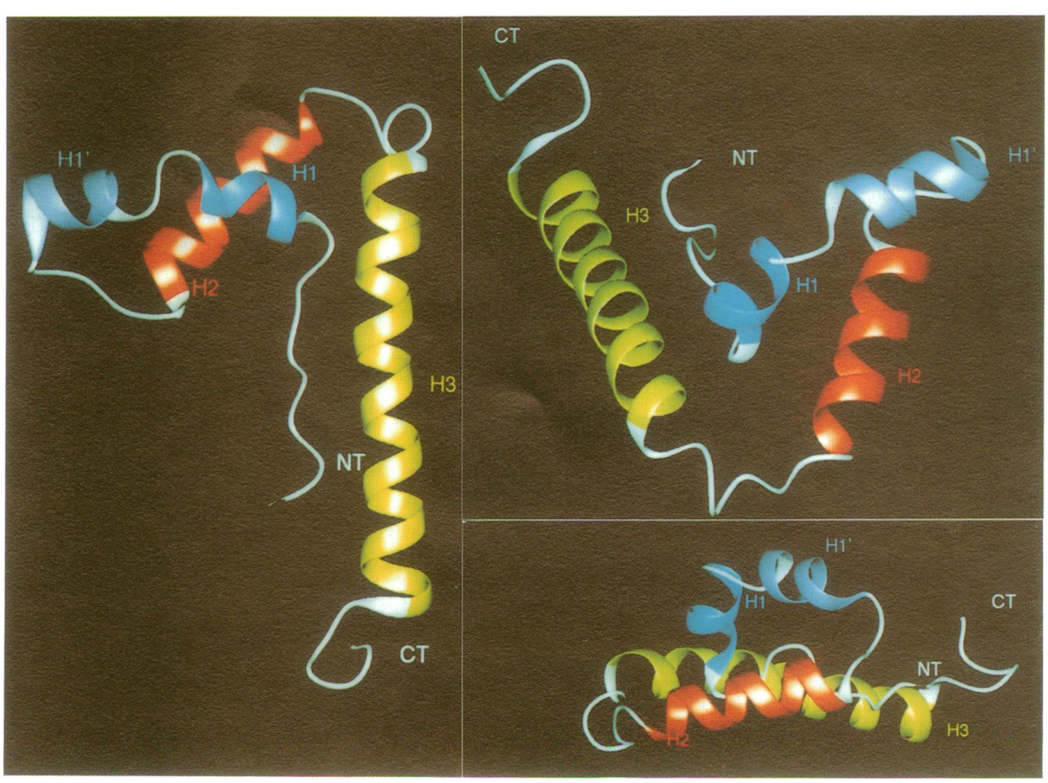

Figure 6. Three ribbon cartoon representations of the coordinate averaged structure of HMG1 box 2 showing the relative orientation of the four helices (H1, H1′, H2 and H3). NT= N-terminus, CT= C-terminus.

were short range (i−j <4) and 346 were long range (i−j ≥ 4) contacts. In addition, 64 backbone hydrogen bond restraints (for 32 residues) and 76 ϕ,ψ dihedral angle restraints for the α-helical regions were included.

Evaluation of the structures

Simulated annealing of 43 extended helical conformations yielded structures of which 30 were selected on the basis of their F_{total} and F_{NOE} energies. All 30 structures had an F_{NOE} <460 kJ mole^{-1}. Most structures showed no distance restraint violations greater than 0.6 Å and in all structures no violations of dihedral angle restraints were observed. A superposition of the backbone atoms for these 30 structures is presented in Figure 4A. Plots of the number of NOE distance restraints and the average r.m.s. deviation of backbone atoms against residue number are shown in Figures 4B and 4C. An averaged structure was then obtained by calculating the mean coordinate positions of these 30 structures, followed by restrained energy minimisation. Figure 5 shows this averaged structure with the side chains depicted and Figure 6 shows three ribbon representations emphasising the overall secondary and tertiary structure organisation. For the averaged structure there were no distance restraint violations greater than 0.60 Å and no dihedral angle restraint violations. The r.m.s. deviation of the 30 structures from this averaged structure between residues 3 and 74 is 0.625 Å for backbone atoms and 1.07 Å for all heavy atoms. Between residues 6 and 62 the r.m.s. deviation of the 30 structures from this averaged structure is 0.37 Å for backbone atoms and 0.88 Å for all heavy atoms. These data indicate that the structures are of medium resolution.

Conformation of the HMG box

The tertiary structure of this HMG box (Figures 5 and 6) shows it to be an 'all α-helix' fold with four helices, in the form of a V-shaped arrowhead with one rather flat face. The two wings of the arrowhead comprise helix 3 (residues P51-R71) along one edge and helices 2 (residues G31 to N42) and 1′ (residues P19 to E24) on the other edge. The overall angle at the apex is ~70°. The helices 2 and 3 together with the extended N-terminal region lie approximately in a plane, forming a rather flat surface to one side of the domain, with helices 1 and 1′ protruding from the opposite side. Helices 2 and 3 are connected by 8 residues, the first 3 of which (N43 to A45) together with the last residue of helix 2 (N42) form a reverse turn. Residues A46 to Q50 are in an approximately helical conformation with its axis directed at ~50° to helix 3. Proline 51 forms the N-terminal residue of helix 3 so that the section beyond A46 could be regarded as a single kinked helix.

The first 9 residues of the box are in an extended conformation lying anti-parallel to helix 3, such that the N-terminus of the box and the C-terminus of helix 3 lie close together. We term this structural element the 'terminal unit'. The conformation and sequence of this part of the fragment bear a strong resemblance to a section of avian pancreatic polypeptide, aPP (43).

Helix 1 (residues F10−E16) is positioned towards the apex and crosses beneath helix 2 such that its N-terminus lies close to helix 3 and its C-terminus protrudes beyond helix 2. Residue R18 is in an extended conformation and leads into helix 1′ with proline 19 as its N-terminal residue. Helices 1 and 1′ could be regarded as a single kinked helix. Helix 1′ (residues P19 to E24) is followed by a reverse turn (H25 to L28) such that helix 1′ lies approximately anti-parallel to helix 2.

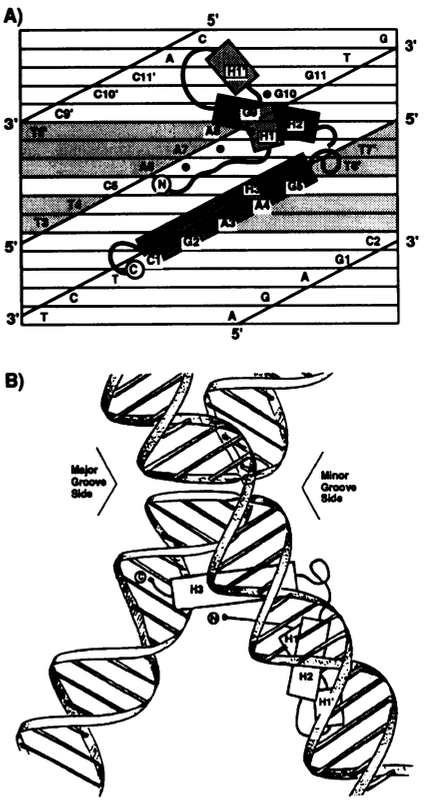

Figure 7. A) 2D helical surface representation of the specific DNA contacts made by LEF-1 (12) and TCF1 (14). Superimposed on the helical surface is the outline structure of the HMG box in the preferred orientation. No attempt has been made to display the proposed distortion of the lower part of the major groove (see text). Filled circles represent the bases for which methylation inhibited protein binding (N7 of guanine in the major groove and N3 of adenine in the minor groove) and the shaded circle at G10 indicates partial inhibition of protein binding. AT base pairs that could be replaced by IC with no effect on protein binding are shown shaded. The horizontal distance on the 2D surface corresponds to the circumference at the outermost diameter of the DNA duplex (22 Å). **B)** 4-way junction DNA (59) with an HMG box inserted from the major groove side into the acute angle between adjacent arms.

Anatomy of the core

The apex of the fold contains the hydrophobic core around which the three helices are arranged. Helix 1 is sandwiched between helices 2 and 3 such that residues at the N-terminal end of helix 1 contact predominantly residues in helix 3, whilst residues nearer the C-terminal end of helix 1 predominantly contact helix 2. The six contiguous hydrophobic residues F10 to C14 in helix 1, together with A9 play a key role in holding together the two wings of the fold, i.e. the 'terminal unit' and the helix 1/helix 2 section.

The methyl group of residue A9 is in close contact with the methyl of residue A56 and with the β- and γ-methylenes of residue E53. These two residues lie on the inner face of helix 3 and the contact with A9 provides a link between the start of helix 1 and helix 3. The A9 methyl group is also in contact with the indole ring of W41 (in helix 2), so that residue A9 bridges between helix 3 and helix 2. The sidechains of F10 and L37 are in close contact and represent a hydrophobic interaction between helix 1 and helix 2. The F11 sidechain points out into the solvent space between the wings of the fold, although its β-methylene group is in contact with the β-methylene of P7, providing a link between helix 1 and the N-terminal residues. The sidechain of

L12 points into the hydrophobic core and is in van der Waals contact with the sidechains of both A56 and Y52 in helix 3. The F13 sidechain however, is in close contact with the sidechains of both M40 and W41 in helix 2. Thus the adjacent residues L12 and F13 in helix 1 interact with the opposite wings of the fold. The sidechain of C14 points out into the solvent, away from the core, rather like that of F11.

At the centre of the hydrophobic core is the sidechain of W41 (in helix 2) that makes contacts across the apex to the N-terminal end of helix 3: in particular the indole ring contacts the β-methylene of Y52 and the γ-methylene of E53. The loop at the very apex of the fold (N43 to Q50) is hydrophilic, however the methylene sidechain of the highly conserved basic residue K49 is internally located and extends across the face of the W41 indole ring. The methyl group of T44 is also in contact with the γ- and ϵ-methylenes of K49. These contacts help to anchor the loop onto the hydrophobic core.

Relation to other HMG boxes

The derived tertiary structure of HMG1 box 2 provides an explanation for the majority of sequence identities and homologies between different members of the HMG family (see Figure 1) as being those important for maintaining the integrity of the fold. The hydrophobic cluster at the intersection of helices 1 and 2 is composed of the rings of F10, F13, W41 (all of which are invariably aromatic), plus the side chain of L37. The hydrophobic face of helix 3 is made up of residues Y52, A56, L59, Y63, I67 and Y70, three of which (L59, Y63 and I67) make hydrophobic contacts to the three proline residues (P7, P6 and P3, respectively) of the extended N-terminal region. The conserved proline 26 is at position 2 of the turn that reverses the chain direction between helices 1′ and 2. The small conserved residue at position 38 (typically glycine) in helix 2 allows for close packing of the aromatic ring of F10 from helix 1. A high degree of conservation for these structurally important residues makes it clear that other closely homologous members of the HMG box family must adopt essentially the same fold as found for HMG1 box 2.

Differences in the length of HMG box sequences lie predominantly in the region of helix 1′. For example, box 1 of mammalian HMG1s (44) and trout HMGT (21) have two extra residues, maize HMG1 (27) has one extra, whilst yeast ste11 (46) has four fewer residues. Several boxes of hUBF1 are also shorter in this region (2, 47). Conformational variation in this part of the box is therefore to be expected. In addition the proline at position 19 is not conserved in the HMG box family and this suggests that in some boxes helix 1 might extend further than in the present structure such that helix 1′ is no longer a separate helix. The region between helices 2 and 3 is very constant in length despite its considerable sequence variation (see Figure 1). Only in drosophila HMG-D is there a two-residue deletion (48), suggesting that the reverse turn immediately following helix 2 is absent in this case.

The amino acid sequence of several HMG boxes such as SRY (3), LEF-1 (5) and ABF2 box 2 (7) that bind to defined DNA sequences all contain a proline at position 68 (see Figure 1). In HMG1 box 2 however, residue A68 is located within helix 3 but the presence of a proline would be expected to break or kink the helix at this position. A different conformation in the most C-terminal region of the HMG boxes from SRY, LEF-1 and ABF2 is thus probable. Since the extended N-terminus lies adjacent to and contacts helix 3 in the present structure, the conformation of the most N-terminal region (up to residue 3) may also differ in HMG boxes with a proline at position 68.

Interpretation of mutagenesis data

The structure provides a rationale for understanding the results of mutagenesis in SRY (49, 50) and LEF-1 (12). The mutations Y52S in LEF-1 and G38R in SRY both result in loss of DNA binding. Amino acids 52 and 38 are highly conserved internal residues (see Figure 5) and loss of DNA binding in these two cases is thus probably a consequence of gross structural perturbation. Residue 49 is a highly conserved lysine and the mutation K49I in SRY also results in loss of DNA binding. K49 has its side chain packed along the face of the indole ring of W41, with its ϵ-amino group positioned to hydrogen-bond with the carbonyl group of W41, thereby placing a positive charge at a suitable location to interact with the dipole of helix 2 (see Figure 5). This mutation is also likely to result in a distortion of tertiary structure.

The mutations V3L and M7I in SRY, that result in sex reversal, and the double mutant K4E,K5E in LEF-1, all fail to bind DNA (12, 49, 50). These mutations are located in the N-terminal segment which is associated with the C-terminal region of helix 3 and seem likely to be mutants that directly affect DNA binding rather than perturbing the fold of the HMG box.

CONCLUSIONS

The structure determined for HMG1 box 2 indicates an all α-helical fold in the form of a V-shaped arrowhead. Helices run along two edges of the arrowhead and one face of the arrowhead is rather flat. One consequence of this architecture is the gap between the two wings of the arrowhead at its base (the C_α atoms of residues R5 and I30 being some 12 Å apart). The fold is not found in the current protein structure database.

Whilst HMG1 box 2 is a helical domain that binds to DNA, it is not related to the helix-turn-helix (HTH) proteins (51). Helices 2 and 3 of HMG1 box 2 though similar in size and orientation to the same helices of HTH proteins are however further apart and separated by helix 1, rather than directly interacting as in the HTH proteins. Also, helices 1/1′ and 2 of HMG1 box 2 do not have the same relative orientation as the HTH motif. The structure determined for this HMG box is the first example of a new fold to which DNA binds.

The similarity of the aPP structure (43) to the extended N-terminal region and helix 3 implies that together they could form an integral structural unit—the terminal unit. Tryptic cleavage of HMG1 results in a folded product comprising residues 12 to 67 from box 1 (52). These correspond to residues P7 to K60 in box 2 and this observation suggests that the HMG box consists of 2 quasi-independent structural units: the terminal unit and the remainder. The location within the first 7 residues of amino acids that appear critical in mediating DNA binding implies that the terminal unit is directly involved. The importance of the most N-terminal residues in DNA binding is also suggested by the sequence homology between residues 1 and 7 of HMG boxes and the three basic regions of proteins HMGI/Y/I(C) (47) that have been shown *in vitro* to mediate DNA binding in the minor groove (53). The generality of a motif of this type has been recognised (54) and termed the GRP repeat. Additionally, residue K6 in box 1 of HMG1 is a site of post-translational acetylation

(55), a modification that might modulate DNA binding. Although all this evidence is indirect, it supports the primacy of the terminal unit in the interaction of HMG boxes with DNA.

DNA binding of HMG boxes

It is apparent that HMG boxes, whether structure- or sequence-specific, must adopt essentially the same fold. Thus, in our modelling we looked for a common structural feature in the DNAs they bind. We noted that there is a resemblance between a single structure-specific box bound in the acute angle of two duplexes in a 4 way junction, to a single sequence-specific box binding to linear DNA (12), to give a strongly bent duplex with the box on the inside arc. We have further assumed no change in the HMG box fold on binding DNA since the numerous hydrophobic contacts in the core that determine theorientation of the wings would be disrupted.

Methylation and diethyl pyrocarbonate interfernce footprinting together with IC base pair substitution experiments indicate that LEF-1 and TCF1 principally contact the minor groove, located on one side of the duplex between base pairs 3 and 8 (12, 14 and summarised in Figure 7A). In the major groove, contact is only noted for base G9 and to a lesser extent G10. In translating these observations to the 4-way junction we positioned the box so that the N-terminal residues strongly implicated in binding to DNA via the minor groove, indeed made minor groove contact. As well, some contact in one part of a major groove was allowed, but contact to other major grooves was largely avoided. If the arrowhead of HMG1 box 2 is inserted into the acute angle of the 4-way junction from the major groove side a reasonable fit can be obtained (Figure 7B). The N-terminal residues lie along the minor groove of one arm of the junction (the RH arm in Figure 7B), the N-terminal end of helix 2 approaches the adjacent major groove of the same arm, and helix 3 runs along the phosphodiester chain of this arm but also stretches across to the phosphodiester chain of the other (LH) arm, running in the same 5′ to 3′ direction.

The symmetry of the junction means that the equivalent site in the opposite half is also accessed from the major groove side. With a box occupying both sites, the distance between the C-terminus of the first box and the N-terminus of the second is about 40 Å, which could be bridged by the 12 amino acids that separate boxes 1 and 2 of HMG1.

Figure 7A suggests how the sequence-specific HMG boxes might bind to a single DNA duplex in an analogous fashion. In this arrangement helix 3 lies along a single phosphodiester chain, the N-terminal residues align along the minor groove. Helix 2 is partially inserted in the 'upper' part of the major groove and its position resembles that of the recognition helix in bacterial HTH proteins (51). If the distortion of the DNA induced by the binding of these HMG boxes is closely related to that in the 4-way junction, then the minor groove and the 'upper' part of the major groove remain largely unchanged. However a strong kink brings the 'lower' part of the major groove in Figure 7A towards the viewer, with the result that helix 3 bridges between two phosphodiester chains.

It must be emphasised that we discuss only outline models, principally because the conformation of 4-way junction DNA is not known with great precision and might change somewhat on HMG box binding, nor is it known in what way LEF-1 etc. bend the DNA to which they bind. These uncertainties make it not worth attempting quantitative modelling. We nevertheless think that the outline models presented are plausable and represent testable hypotheses.

Note added

After completing this structure determination we learned of the work of Weir et al. (56) who have presented a structure based on 2D NMR, of the same domain of HMG1 (with no adduct at Cys14). The sequence of their HMG box is longer by 4 residues (FKDP) at the N-terminus, but shorter by 6 residues (GKPDAA) at the C-terminus. Weir et al. did not observe sequential NOEs in their N-terminal FKDP sequence, which they attributed to flexibility. There is a common sequence of 73 residues. The parameters of the two final models are as follows (using the present numbering system):

	RMS Deviations		Van der Waals
	Backbone Atoms	All Heavy Atoms	Energy
Weir et al.	0.69Å	0.94Å (residues 6–57)	−1120±46kJ/mol
Present work	0.37Å	0.88Å (residues 6–62)	−1342±29kJ/mol

Certain local differences between the models are apparent: for example, helix 1 is interrupted and helix 3 begins 5 residues later in the present model. In addition, the angle between the wings of the fold is ~70° in the present structure, but ~80° in that of Weir et al. The cause of these differences is at present unclear.

ACKNOWLEDGEMENTS

We are grateful to G.H.Goodwin and I.D.Campbell for encouragement and support throughout this project and to A.G.Murzin, O.B.Ptitsyn and T.Moss for fruitful discussions. We acknowledge G.H.Dixon for the gift of the hamster cDNA clone pCH1, P.Morgan for sedimentation equilibrium measurements, A.Carne for amino acid sequencing and C.H.Turner for preliminary NMR spectra. P.C.D. and D.G.N. are members of the Oxford Centre for Molecular Sciences supported by the SERC and the MRC. P.C.D. is a Royal Society supported University Research Fellow. The project was supported in Portsmouth by the Wellcome Trust.

REFERENCES

1. Goodwin, G. H. and Johns, E. W. (1973) *Eur. J. Biochem.*, **40**, 215–219.
2. Jantzen, H.-M., Admon, A., Bell, S. P. and Tjian, R. (1990) *Nature*, **344**, 830–836.
3. Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A.-M., Lovell-Badge, R. and Goodfellow, P. N. (1990) *Nature*, **346**, 240–244.
4. van der Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) *EMBO J.*, 10, 123–132.
5. Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes & Dev.*, **5**, 880–894.
6. Laudet, V., Stehelin, D. and Clevers, H. (1993) *Nucleic Acids Res.*, **21**, 2493–2501.
7. Diffley, J. F. X. and Stillman, B. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7864–7868.
8. Bianchi, M. E., Beltrame, M. and Paonessa, G. (1989) *Science*, **243**, 1056–1059.
9. Bianchi, M. E., Falciola, L., Ferrari, S. and Lilley, D. M. J. (1992) *EMBO J.*, 11, 1055–1063.
10. Pil, P. M. and Lippard, S. J. (1992) *Science*, **256**, 234–237.
11. Bruhn, S. L., Pil, P. M., Eissigmann, J. M., Hansman, D. E. and Lippard, S. J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2307–2311.
12. Giese, K., Amsterdam, A. and Grosschedl, R. (1991) *Genes & Dev.*, **5**, 2567–2578.
13. Giese, K., Cox, J. and Grosschedl, R. (1992) *Cell*, **69**, 185–195.
14. van der Wetering, M. and Clevers, H. (1992) *EMBO J.*, **11**, 3039–3044.

15. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1992) *EMBO J.*, **11**, 4497–4506.
16. Lilley, D. M. J. (1992) *Nature*, **357**, 282–283.
17. Reeck, G. R., Isackson, P. J. and Teller, D. C. (1982) *Nature*, **300**, 76–78.
18. Tremethick, D. J. and Molloy, P. L. (1986) *J. Biol. Chem.*, **261**, 6986–6992.
19. Watt, F. and Molloy, P. L. (1988) *Nucleic Acids Res.*, **16**, 1471–1486.
20. Singh, J. and Dixon, G. H. (1990) *Biochemistry*, **29**, 6295–6302.
21. Pentecost, B. T., Wright, J. M. and Dixon, G. H. (1985) *Nucleic Acids Res.*,**13**, 4871–4888.
22. Wright, J. M. and Dixon, G. H. (1988) *Biochemistry*, 27, 576–581.
23. Lee, K. L. D., Pentecost, B. T., D'Anna, J. A., Tobey, R. A., Gurley, L.R., and Dixon, G. H. (1987) *Nucleic Acids Res.*, **15**, 5051–5068.
24. Smith, D. B. and Johnson, K. S. (1988) *Gene*, **67**, 31–40.
25. Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Phys. Chem.*, 71, 4546–4553.
26. Kumar, A., Ernst, R. R. and Wuthrich, K. (1981) *Biochem Biophys. Res. Commun.*, **95**, 1–6.
27. Braunschweiler, L. and Ernst, R. R. (1983) *J. Magn. Reson.*, **53**, 521–528.
28. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc.*, **107**, 2820–2821.
29. Bax, A., Sklenar, V., Clore. G. M. and Gronenborn, A. M. (1987) *J. Am. Chem. Soc.*, 109, 6511–6513.
30. Plateau, P. and Gueron, M. (1982) *J. Am. Chem. Soc.*, 104, 7310–7311.
31. Marion. D. and Bax. A. (1988) *J. Magn. Reson.*, 80, 528–533.
32. Driscoll. P. C., Clore. G. M., Beress, L. and Gronenborn, A. M. (1989) *Biochemistry*, 28, 2178–2187.
33. Messerle, B. A., Wider, G., Otting, G., Weber, C. and Wuthrich, K. (1989) *J. Magn. Reson.*, **85**, 608–613.
34. Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T. and Gronenborn, A. M. (1990) *Biochemistry*, 29, 3542–3556.
35. Kay, L. E. and Bax, A. (1990) *J. Magn. Reson.*, **86**, 110–126.
36. Wen, L., Huang, J.-K., Johnson, B. H. and Reeck, G. R. (1989) *Nucleic Acids Res.*, **17**, 1197–1214.
37. Cary, P. D., Turner, C. H., Leung, I., Mayes, E. and Crane-Robinson, C. (1984) *Eur. J. Biochem.*, **143**, 323–330.
38. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. and Clore, G. M. (1990) *Biochemistry*, **28**, 6150–6156.
39. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
40. Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) *Biochemistry*, **31**, 1647–1651.
41. Brunger, A. T., Clore, G. M., Gronenborn, A. M. and Karplus, M. (1987) *Protein Engineering*, 1, 399–406.
42. Nilges, M., Gronenborn, A. M., Brunger, A. T. and Clore, G. M. (1988) *Protein Engineering*, 2, 27–38.
43. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. and Wu, C.-W. (1981) *Proc. Natl. Acad. Sci. USA*, 78, 4175–4179.
44. Paonessa, G., Frank, R. and Cortese, R. (1987). *Nucleic Acids Res.*, **15**, 9077.
45. Grasser, K. D. and Feix, G. (1991) *Nucleic Acids Res.*, **19**, 2573–2577.
46. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. and Yamamoto, M. (1991) *Genes & Dev.*, **5**, 1990–1999.
47. Bachvarov, D. and Moss, T. (1991) *Nucleic Acids Res.*, **19**, 2331–2335.
48. Wagner, C. R., Hamana, K. and Elgin, S. C. R. (1992) *Mol. Cell. Biol.*, 12, 1915–1923.
49. Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebl, M. and Alexander-Bridges, M. (1991) *Nature*, **354**, 317–320.
50. Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. N. (1992) *Science*, **255**, 453–456.
51. Pabo, C. O. and Sauer, R. T. (1992) *Ann. Rev. Biochem.*, **61**, 1053–1095.
52. Cary, P. D., Turner, C. H., Mayes, E. and Crane-Robinson, C. (1983) *Eur. J. Biochem.*, 131, 367–374.
53. Reeves, R. and Nissen, M. S. (1990) *J. Biol. Chem.*, **265**, 8573–8582.
54. Churchill, M. E. A. and Travers, A. A. (1991) *Trends Biochem. Sci.*, **16**, 92–97.
55. Sterner, R., Vidali, G. and Allfrey, V. G. (1979) *J. Biol. Chem.*, **254**, 11577–11583.
56. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D. and Thomas, J. O. (1993) *EMBO J.*, 12, 1311–1319.
57. Ner, S. S. (1992) *Current Biology*, 2, 208–210.
58. Haqq, C. M., King, C.-Y., Donahoe, P. K. and Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1097–1101.
59. Bhattacharyya, A., Murchie, A. I. H., von Kitzing, E., Diekmann, S., Kemper, B. and Lilley, D. M. J. (1991) *J. Mol. Biol.*, **221**, 1191–1207.