HMG box proteins bind to four-way DNA junctions in their open conformation

J.Richard G.Pöhler, David G.Norman,
Janice Bramham, Marco E.Bianchi¹ and
David M.J.Lilley²

CRC Nucleic Acid Structure Research Group, Department of Biochemistry, The University of Dundee, Dundee DD1 4HN, UK and
¹Department of Genetics and Microbiology, University of Milano, via Celoria 26, 20133 Milano, Italy

²Corresponding author
e-mail: dmjlilley@bad.dundee.ac.uk

The HMG box is an 80 amino acid domain found in a variety of eukaryotic chromosomal proteins and transcription factors. Binding to DNA is associated with recognition of structural distortion or manipulation of DNA structure. All the HMG box domains bind to four-way DNA junctions, which must therefore present some feature that is common to the binding targets of this wide variety of proteins. Since the four-way junction can itself adopt a variety of structures depending upon conditions, it is important to determine in which form it exists in complexes with HMG boxes. We find that a single HMG box domain is bound exclusively to the open square form of the junction and that conditions that stabilize the stacked X structure significantly lower affinity for the HMG box. We suggest that the HMG domain binds to one arm of the junction in the minor groove at the point of strand exchange and we present a model for the structure of the complex.

Keywords: DNA–protein interaction/HMG1/Holliday junction/SRY/UBF

Introduction

The HMG box is a relatively small eukaryotic DNA binding motif that is found in a large number of proteins of varying function (reviewed in Bianchi, 1995; Read *et al.*, 1995), from architectural components of chromatin to transcription factors. These modules exhibit significant structural selectivity in binding to DNA and it is the purpose of the present study to understand the basis of their recognition of distorted DNA structure.

HMG boxes are typically 70–80 amino acids in size, with little in the way of sequence conservation; rather, the amino acid type is conserved. They are usually found as domains of larger proteins, frequently in more than a single copy. HMG boxes can be classified into three subtypes according to their DNA binding preferences. The first class contains the non-histone chromosomal proteins, such as HMG1 and HMG2 (Goodwin *et al.*, 1973). Proteins of this class do not bind to a particular duplex DNA sequence, but bind to bent or distorted structures, such as DNA containing four-way DNA junctions (Bianchi *et al.*, 1989), *cis*-platinum adducts (Pil and Lippard, 1992) or

base bulges (Payet and Travers, 1997). These proteins also induce considerable structural distortion, indicated by enhanced rates of circularization, for example (Paull *et al.*, 1993). The second class includes the nucleolar and mitochondrial transcription factors, such as UBF, the upstream binding factor in RNA polymerase I transcription (Jantzen *et al.*, 1990), and mtTF1, the human mitochondrial transcription factor (Parisi *et al.*, 1993). Such proteins produce a discrete footprint on their target promoters, but have no discernible recognition sequence. However, this class of HMG box proteins exhibits structure-selective binding to four-way junctions (Kuhn *et al.*, 1994). The last class of HMG box proteins contains sequence-specific transcription factors, such as the testis determining factor SRY (Sinclair *et al.*, 1990) and the lymphoid enhancer binding factor LEF-1 (Travis *et al.*, 1991; van de Wetering *et al.*, 1991). These proteins bind duplex DNA targets with high sequence specificity. Nevertheless, they will also bind to DNA junctions of any base sequence (Ferrari *et al.*, 1992). Thus HMG box proteins of all classes exhibit structure-selective binding of four-way DNA junctions.

The solution structures of several HMG boxes have been solved by NMR (Read *et al.*, 1993; Weir *et al.*, 1993; Jones *et al.*, 1994; Hardman *et al.*, 1995; Love *et al.*, 1995; Werner *et al.*, 1995), including both structure-specific and sequence-specific domains. The overall structure of the HMG box is remarkably similar in all cases, being predominantly α -helical, with the helices arranged in an L shape. In the complexes of LEF-1 and SRY with their respective DNA targets it can be seen that the HMG domain introduces a significant bend in the DNA, in agreement with previous electrophoretic circular permutation experiments (Ferrari *et al.*, 1992; Giese *et al.*, 1992). This bending of the DNA has also been demonstrated for the structure-specific HMG boxes (Pil and Lippard, 1992). Thus HMG box proteins bend DNA as well as binding to already distorted DNA substrates.

HMG box proteins of all classes have been shown to bind with high affinity to the four-way DNA junction (Bianchi *et al.*, 1989; Ferrari *et al.*, 1992; Bracale *et al.*, 1994; Kuhn *et al.*, 1994; Pontiggia *et al.*, 1994; Read *et al.*, 1994; Grasser, 1995; Peters *et al.*, 1995; Teo *et al.*, 1995; Payet and Travers, 1997); indeed the junction may be regarded as a universal binding substrate for HMG box proteins. Quite why this is so has been unclear. It would seem highly improbable that the four-way DNA junction would be a natural substrate for a protein such as LEF-1; rather, it is more likely that some feature of the structure of the junction is similar to that required for binding of the protein to its duplex target. Yet even if non-biological in direct terms, the nature of this interaction could yield insight into the normal binding mode of the protein. It is therefore of considerable importance to establish what structural aspects of the junction are important in this

process and thus shed light on the recognition process in general. It has been shown recently that proteins whose primary role involves interaction with the four-way junction, such as the DNA junction resolving enzymes (reviewed in White *et al.*, 1997), distort the structure of the junction upon binding. It is therefore important to establish the form of the junction that is being recognized by HMG box proteins, in order to dissect the process of molecular recognition involved.

We have investigated the global structure of the complex formed between the four-way DNA junction and a number of HMG boxes, representing the three subtypes of HMG box. We show that each type of HMG box opens the structure of the four-way junction into an extended square conformation and we present a model for how this form of the junction is recognized by the protein.

Results

Binding of HMG boxes to a four-way DNA junction

We have analysed the binding to a four-way DNA junction of three HMG boxes as representative members of the subtypes of HMG box. These are the A box of rat HMG1, the N-terminal HMG box of UBF and the HMG box of human SRY. All have been previously demonstrated to bind to the four-way DNA junction (Bianchi *et al.*, 1989; Ferrari *et al.*, 1992; Kuhn *et al.*, 1994). Binding to a radioactively ^{32}P -labelled four-way DNA junction was studied as a function of protein concentration in a gel electrophoretic retardation assay in the presence of 90 mM Tris–borate, pH 8.3, 1 mM EDTA. Analysis of the binding of rat HMG1 box A is presented in Figure 1A. A discrete retarded complex is formed at low protein concentrations. Further retarded species become apparent at higher protein concentrations; however, these bands are diffuse and are not indicative of discrete species. The extent of complex formation was quantified by phosphorimaging and the data fitted to a model for the binding of a single monomeric (see below) HMG box (Figure 1B), from which a dissociation constant (K_D) of 200 nM was calculated. Formation of the first complex is highly structure selective; this complex is not displaced by a 1000-fold excess of duplex DNA (Figure 1C).

Analysis of binding of the UBF and SRY HMG boxes is presented in Figure 1D and E. In both cases formation of discrete complexes is apparent. The dissociation constants for UBF and SRY HMG boxes binding to four-way junctions were calculated from these data as above. Together with HMG1 box A, these cover a range of affinities in the order UBF (K_D 1.5 μM) < HMG1 box A (K_D 200 nM) < SRY (K_D 47 nM). This corresponds to a variation in binding free energy of 2 kcal/mol between UBF and SRY. The affinity of SRY for binding to junctions was closely similar to that for its cognate duplex. We carried out an electrophoretic retardation analysis using a 27 nt duplex containing the consensus SRY binding sequence 5'-TAACAATG-3' under the same conditions as the experiments with the junction. We observed the formation of a discrete complex, analysis of which gave a value for the dissociation constant of K_D = 54 nM.

Expression of rat HMG1 box A as a fusion with maltose binding protein

HMG boxes are relatively small, typically comprising ~80 amino acids with a molecular weight of ~10 kDa. For this

reason they generate a relatively small electrophoretic retardation on binding to four-way DNA junctions. This can pose problems for analysis of the global shape of the complex by comparative gel electrophoresis (below), where the mass of junction required for analysis is ~75 kDa. Moreover, an extended version of HMG1 box A was also required for analysis of quaternary structure (see below). We therefore increased the effective mass of protein by expressing rat HMG1 box A as an N-terminal fusion with maltose binding protein (MBP), which has a mass of ~40 kDa. The 55 kDa fusion protein could be purified by affinity chromatography on an amylose resin and the purity analysed by polyacrylamide gel electrophoresis in the presence of SDS (Figure 2A). The fusion protein contained an IEGR site between the MBP and HMG1 box A sections, thus enabling the component proteins to be separated by digestion with protease factor Xa.

Binding of MBP–HMG1 box A to a four-way DNA junction

The binding of MBP–HMG1 box A to a four-way DNA junction was analysed by electrophoretic retardation analysis. As with the unfused protein, a discrete complex was clearly formed (Figure 2B). Clearly the substantial mass of additional protein at the N-terminus does not prevent binding to the four-way junction, as was anticipated, given the modular character of HMG boxes. However, quantitative analysis of binding showed that affinity for the fused protein was reduced 10-fold (K_D = 2 μM), indicating that MBP may interfere with binding to the four-way junction by ~1.3 kcal/mol. Nevertheless, competition experiments using unlabelled duplex DNA showed that the fused HMG box retained its specificity for the branched DNA species (data not shown).

The HMG box binds to a four-way junction as a monomer

Many proteins that have evolved to bind DNA junctions do so in oligomeric form. For example, all the junction resolving enzymes bind four-way junctions as dimers (Pöhler *et al.*, 1996; White and Lilley, 1996; Parkinson and Lilley, 1997). In contrast, NMR structures of LEF-1 (Love *et al.*, 1995) and SRY (Werner *et al.*, 1995) complexes with DNA reveal that these proteins bind their duplex targets in monomeric form. We therefore analysed the quaternary structure of a HMG box bound to a DNA junction. In this method we use a fusion of the protein of interest where a protease-sensitive linker permits enzymatic conversion into the non-fusion protein (White and Lilley, 1996).

We subjected the MBP–HMG1 box A fusion to increasing digestion with factor Xa protease, leading to formation of an increasing proportion of non-fusion HMG1 box A protein. Binding to a DNA junction of the partially digested protein as an n -fold oligomer should lead to the formation of $n + 1$ different complexes and, because of the different protein masses, these can generally be resolved by gel electrophoresis. For example, the dimeric resolving enzymes ($n = 2$) lead to three different complexes, as the homodimeric fusion, the homodimeric non-fusion and the heterodimeric fusion + non-fusion species. Application of this experiment to MBP–HMG1 box A leads to forma-

Fig. 1. Binding of HMG box proteins to a four-way DNA junction. **(A)** Binding a single HMG1 box A to a four-way junction. A fixed concentration (1 nM) of radioactively $5'$ - 32 P-labelled DNA junction 3 (Duckett *et al.*, 1988), assembled from four oligonucleotides each of 30 nt, was incubated with increasing concentrations of HMG1 box A in the presence of 5 mM EDTA. The complexes were analysed by electrophoresis in an 8% polyacrylamide gel in the presence of 1 mM EDTA and autoradiography. The complexes migrate as species that are retarded relative to the free DNA junction (indicated on the left). Note the initial formation of a well-defined first complex (indicated on the right), followed by further retarded more diffuse species at high protein concentrations. Lane 1, junction without added protein; lanes 2–10, junction incubated with HMG1 box A at concentrations of 28, 56, 112, 224, 448 and 896 nM and 1.79, 3.58 and 7.17 μ M respectively. **(B)** Isotherm for binding of HMG1 box A to junction 3 in the absence of added magnesium ions. The extent of HMG1 box A binding to junction 3 as a function of total protein concentration was estimated by gel electrophoresis as in (A). The fraction of DNA junction bound to protein was calculated for each protein concentration and plotted against protein molarity on a logarithmic scale. The data were fitted to a model for the binding process (see Materials and Methods), from which the binding affinity was calculated. **(C)** Competition with double-stranded DNA of related sequence. Radioactively labelled junction 3 (1 nM), assembled from four oligonucleotides each of 30 nt, was incubated with a fixed concentration (750 nM) of HMG1 box A in the presence of increasing concentrations of non-radioactive competitor duplex DNA. The competitor DNA was an 80 bp duplex corresponding to strand x from junction 3 bound to its complement. The complexes were analysed by electrophoresis in an 8% polyacrylamide gel in the presence of 1 mM EDTA and autoradiography. Lane 1, junction without added protein; lanes 2–5, junction incubated with HMG1 box A in the presence of a 1, 10, 100 and 1000-fold excess of duplex competitor DNA respectively. **(D)** Binding a single HMG box from *Xenopus* UBF to a four-way junction. A fixed concentration (1 nM) of radioactively $5'$ - 32 P-labelled DNA junction 3 was incubated with increasing concentrations of UBF HMG box in the presence of 5 mM EDTA. The complexes were analysed by electrophoresis in an 8% polyacrylamide gel in the presence of 1 mM EDTA and autoradiography. Lanes 1–10, junction incubated with UBF HMG box at concentrations of 8, 16, 32, 64, 128, 256 and 512 nM and 1.02, 2.05 and 4.10 μ M respectively. **(E)** Binding a single HMG box from human SRY to a four-way junction. A fixed concentration (1 nM) of radioactively $5'$ - 32 P-labelled DNA junction 3 was incubated with increasing concentrations of SRY HMG box in the presence of 5 mM EDTA. The complexes were analysed by electrophoresis in an 8% polyacrylamide gel in the presence of 1 mM EDTA and autoradiography. Lane 1, junction without added protein; lanes 2–8, junction incubated with SRY HMG box at concentrations of 5, 10, 20, 39, 78, 156 and 313 nM respectively.

tion of just two junction–protein complexes (Figure 3), indicative of binding as a monomer ($n = 1$). Equivalent results were obtained for the MBP–SRY HMG box (data not shown), showing that this HMG box also binds to the four-way DNA junction in monomeric form.

Binding of the HMG box to the four-way DNA junction is inhibited in the presence of magnesium ions

We have previously noted that it was not possible to observe formation of electrophoretically retarded com-

Fig. 2. Generation and binding properties of an MBP-HMG1 box A fusion polypeptide. (A) Denaturing gel electrophoretic analysis of MBP-HMG1 box A and the individual HMG1 box A. The proteins were analysed by polyacrylamide gel electrophoresis in SDS, alongside two mixtures of proteins (lanes 1 and 2) to serve as molecular weight standards (molecular weights indicated either side). Lane 3, MBP-HMG1 box A fusion polypeptide; lane 4, HMG1 box A. (B) Binding of MBP-HMG1 box A to a four-way junction. A fixed concentration (1 nM) of radioactively 5'-³²P-labelled DNA junction 3 was incubated with increasing concentrations of MBP-HMG1 box A in the presence of 5 mM EDTA. The complexes were analysed by electrophoresis in a 6% polyacrylamide gel in the presence of 1 mM EDTA and autoradiography. Lane 1, junction without added protein; lanes 2–10, junction incubated with MBP-HMG1 box A at concentrations of 118, 236, 472 and 944 nM and 1.89, 3.78, 7.55, 15.10 and 30.20 µM respectively.

plexes of DNA junctions and HMG box proteins when the electrophoresis buffer contained significant concentrations of magnesium ions (D.R.Duckett and D.M.J.Lilley, unpublished observations). We therefore explored the effect of inclusion of magnesium ions on affinity of binding.

Electrophoretic retardation analysis of formation of the complex between a radioactively ³²P-labelled four-way DNA junction and MBP-HMG1 box A was performed in the presence of increasing concentrations of magnesium ions. The calculated association constants are shown in Figure 4, where it is clear that affinity steadily decreases as the magnesium ion concentration increases in the range 0–350 µM. In the presence of 500 µM or higher magnesium ions we could detect no complex formation using the electrophoretic assay. Destabilization of binding was not a general effect of raised ionic strength, since sodium ions had no effect on binding. Thus the effect appears to be specific to magnesium ions.

Fig. 3. Quaternary structure of HMG1 box A in the major complex with a four-way DNA junction. MBP-HMG1 box A was subjected to increasing extent of digestion by factor Xa protease to generate partial cleavage of the MBP fusion peptides. The digested protein (5 µM final concentration) was then incubated with junction 3 (1 nM) and the complexes analysed by electrophoresis in 8% polyacrylamide in the presence of 1 mM EDTA and autoradiography. Only two retarded complexes are observed, corresponding to binding of fusion and non-fusion HMG1 box in monomeric form (indicated right). Lane 1, junction without added protein; lane 2, junction incubated with undigested MBP-HMG1 box A; lanes 3–7, junction incubated with MBP-HMG1 box A that had been subjected to preincubation with 800:1, 400:1, 200:1, 100:1 and 50:1 (w/w) ratios of MBP-HMG1 box A to factor Xa.

Fig. 4. Affinity of MBP-HMG1 box A for a four-way junction as a function of magnesium ion concentration. Association constants (K_A) for MBP-HMG1 box A binding to junction 3 were measured by electrophoretic retardation as a function of magnesium ion concentration. Note the pronounced loss of affinity with increasing ion concentration.

HMG boxes bind to an extended global conformation of the four-way DNA junction

To understand recognition of DNA junctions by HMG box proteins and thus the ramifications of recognition by HMG box proteins generally, the most important question to ask concerns the structure adopted by the junction in the bound complex. Comparative gel electrophoresis can be used to reveal the global structure of the four-way DNA junction in solution, either alone (Duckett *et al.*, 1988) or in combination with proteins (Duckett *et al.*, 1995; Parsons *et al.*, 1995; Pöhler *et al.*, 1996; White and

Fig. 5. Comparative gel electrophoretic analysis of the global structure of the complexes of four-way junctions and HMG box proteins. Junction 3 was assembled in the six possible species with two long and two short arms. These were each incubated with protein and analysed by electrophoresis in polyacrylamide. The junction comprises four arms named sequentially B, H, R and X as illustrated. The long-arm, short-arm species are named according to the long arms, e.g. the BH species has long B and H arms and short R and X arms. (A) Global structure of the complex with MBP-HMG1 box A. Binding was carried out under conditions where both free junction and first complex were formed using radioactively $5'$ - ^{32}P -labelled junction species. These were electrophoresed in 6% polyacrylamide in the presence of 1 mM EDTA and the bands visualized by autoradiography. The faster free DNA migrates in a slow, fast, slow, slow, fast, slow pattern of mobilities, corresponding to a square configuration of arms (far right). This leads to the expected pattern of species (right). The retarded complexes also migrate in a similar slow, fast, slow, slow, fast, slow pattern of mobilities, suggesting that there is a square conformation of arms in the complex. Lanes 1–6, BH, BR, BX, HR, HX and RX species respectively. (B) Global structure of the complex with HMG1 box A. Unlabelled junction 3 long-arm, short-arm species were incubated with HMG1 box A and electrophoresed in 9% polyacrylamide. After transfer to nitrocellulose the filter was incubated with an antiserum against HMG1 box A and a photographic image of the protein distribution was generated using a secondary antibody and an enzyme-linked assay. The single HMG1 box A exhibits a slow, fast, slow, slow, fast, slow pattern of mobilities, indicating a square configuration of arms. Lanes 1–6, BH, BR, BX, HR, HX and RX species respectively. (C) Global structure of the complexes with UBF and SRY HMG boxes. UBF and SRY HMG boxes were incubated with radioactively $5'$ - ^{32}P -labelled junction 3 long-arm, short-arm species. These were electrophoresed in 10% polyacrylamide in the presence of 1 mM EDTA and the bands visualized by autoradiography. The complexes with both proteins migrate in slow, fast, slow, slow, fast, slow patterns of mobilities, indicative of a square conformation of arms in the complexes. Lanes 1–6, SRY HMG box binding to BH, BR, BX, HR, HX and RX species respectively; lanes 7–12, UBF HMG box binding to BH, BR, BX, HR, HX and RX species respectively.

Lilley, 1997). In this method all possible combinations of junctions with two long and two short arms are derived from a four-way junction with arms each of 40 bp in length. The electrophoretic mobility in polyacrylamide of these six two-long, two-short arm species can be compared and analysed on the basis of the theoretical relationship (Lumpkin and Zimm, 1982) between electrophoretic mobility and the angle between the two long arms. This approach led to the proposal of an antiparallel stacked X structure for the four-way DNA junction in free solution (Duckett *et al.*, 1988), which was confirmed by subsequent analysis using other methods (Cooper and Hagerman, 1989; Murchie *et al.*, 1989; Clegg *et al.*, 1992).

MBP-HMG1 box A was incubated with the six possible species of junction 3 containing two long and two short arms and the complexes analysed by polyacrylamide gel electrophoresis in the absence of added magnesium ions

(Figure 5A). Under these conditions the protein-free junction adopts an open square (not necessarily planar) configuration of arms characterized by the slow, fast, slow, slow, fast, slow pattern; this is formed by all four-way DNA junctions under these conditions (Duckett *et al.*, 1988). The retarded protein-junction complexes also display a clear slow, fast, slow, slow, fast, slow pattern, indicating that the arms of the junction are directed towards the four corners of a square in the complex.

The experiment was repeated with unfused HMG1 box A. However, some confusion could arise because of the relatively small retardation induced by binding of the smaller protein, with the faster species of the complex having electrophoretic mobilities comparable with the slower free DNA species. To remove any ambiguity in assignment of species we used a procedure that was designed to reveal the junction-protein complexes

Fig. 6. Accessibility of central bases in a HMG box junction complex analysed by chemical probing. Potassium permanganate attacks the 5,6 bond of thymine bases rendered accessible by opening, generating a diol that leads to lability of the phosphodiester backbone to base cleavage. The H strand of junction 3 contains a thymine base at the point of strand exchange (open type). The junction was assembled from four oligonucleotides each of 30 nt, of which the H strand was radioactively 5'-³²P-labelled. The junction was subjected to modification by 2 mM KMnO₄ as a function of ionic conditions and the concentration of protein. The DNA was cleaved at the modified thymine by treatment with hot piperidine and analysed by denaturing gel electrophoresis and autoradiography. Note that the thymine located at the point of strand exchange is uniquely modified in the free junction in the absence of added divalent ions (arrowed right), but reactivity is suppressed by addition of magnesium ions. On addition of MBP-HMG1 box A fusion protein the reactivity is restored, despite the presence of 100 μM magnesium ions, indicative of an opening of the structure by the protein. Lane 1, sequence marker generated by subjecting 5'-³²P-labelled H strand to hydrazine modification; lanes 2 and 3, modification of free H strand 5'-³²P-labelled junction 3 with permanganate in the presence of 5 mM EDTA and 100 μM magnesium ions respectively; lanes 4–6, modification of H strand 5'-³²P-labelled junction 3 preincubated with 2, 4 and 8 μM MBP-HMG1 box A respectively in the presence of 100 μM magnesium ions.

uniquely. We incubated HMG1 box A with the six long-short species of junction 3 without radioactive labelling. After electrophoresis the complexes were transferred to a nitrocellulose filter which was then incubated with an antiserum raised against HMG1 box A. A photographic image of the protein distribution was generated by means of a secondary antibody and an enzyme-linked assay. From this it is clear that the complex with the single HMG1 box A exhibits a slow, fast, slow, slow, fast, slow pattern of mobilities (Figure 5B), indicating a square configuration of arms.

We also analysed the global structure of complexes with the UBF and SRY HMG boxes. This was performed using conventional radioactively ³²P-labelled junction 3, selecting conditions giving complete complex formation (Figure 5C). Note that junction 3 does not contain a binding site for SRY in any of its arms. Complexes with

both proteins exhibited electrophoretic mobility patterns described by slow, fast, slow, slow, fast, slow. Thus these HMG boxes also appear to bind to the four-way junction in its extended square conformation. Differences in absolute mobility between the complexes with UBF and SRY suggest that there are some conformational differences between them, but in broad terms the global shapes must be quite similar.

Chemical probing evidence for an open centre to the junction in the complex

Comparative gel electrophoretic analysis of the global shape of the complex between the HMG box and a four-way DNA junction indicates that the junction adopts a square configuration of arms. This implies that the helical arms of the junction do not undergo coaxial stacking in the presence of a bound HMG box and suggests that the centre could be relatively open to the solvent. In the free junction (Duckett *et al.*, 1988) and in complexes with other proteins that generate an open configuration of arms (Bennett and West, 1995; White and Lilley, 1997) this leads to thymine bases at the point of strand exchange becoming reactive to electrophilic addition by osmium tetroxide or permanganate. The latter is preferred for analysis of DNA–protein complexes due to the lower reactivity with the protein (Sasse-Dwight and Gralla, 1989).

We analysed the reactivity of thymine bases at the centre of a junction as a function of binding of HMG1 box A (Figure 6). For this purpose we used junction 3 comprising four strands each of 30 bp, with radioactive 5'-³²P-labelling on the H strand. This strand was chosen as it contains a thymine base located immediately 3' of the point of strand exchange. The junction was reacted with 2 mM potassium permanganate under various conditions and the DNA cleaved at the modified thymine bases by reaction with piperidine. The products were then analysed by sequencing gel electrophoresis and autoradiography. In the absence of added divalent metal ions this thymine base in the free junction was reactive to permanganate as expected, giving a radioactive band corresponding to modification at the thymine present at the junction. Upon addition of 100 μM MgCl₂ this reactivity was suppressed due to protection of the thymine bases by coaxial helical stacking in the stacked X structure. When the junction was incubated with increasing concentrations of MBP-HMG1 box A in the presence of 100 μM magnesium ions reactivity of the thymine bases became apparent. Thus binding of the HMG box led to exposure of the thymine base at the point of strand exchange, consistent with formation of an open central region.

Discussion

The four-way DNA junction may be regarded as the universal binding substrate for HMG box proteins, from the non-sequence-selective proteins such as HMG1 to the sequence-specific transcription factors such as SRY. In all cases binding to the junction is essentially structure selective; even those HMG boxes that bind to duplex targets in a sequence-specific manner will bind to four-way junctions of any sequence. In order to assess the significance of this observation it is necessary to establish

Fig. 7. Modelling the complex of a single HMG box bound to a four-way DNA junction. (A) The LEF-1 HMG box (orange ribbon) bound to one arm of an essentially square planar DNA four-way junction (grey ribbon and bonds). Met10 (magenta) is shown partially inserted between the last two base pairs of one arm of the junction. α -helices 1 and 2 of the box are positioned over the centre of the junction. (B and C) Space filling representation showing minor (B) and major (C) groove sides of an open square planar four-way DNA junction (white) with the LEF-1 HMG box as a solvent-accessible surface (orange) bound on one arm with α -helices filling the space at the centre of the junction on the minor groove side. The major groove side is shown to be extensively solvent exposed.

the nature of the interaction. Since the four-way DNA junction is structurally polymorphic (Duckett *et al.*, 1988), it is vital to establish the form in which it exists in the complex with HMG box proteins.

From our experiments we note the following properties of binding of HMG boxes to four-way junctions: (i) a single major discrete complex is formed, comprising one molecule of the HMG box bound to one junction, any further protein binding apparently being significantly less specific; (ii) binding to the four-way junction is not abolished by a major N-terminal extension of the protein; (iii) the global structure of the junction complexed with a single HMG box molecule is described by a square configuration of arms, closely similar to the structure of the free junction in the absence of divalent cations; (iv) central thymine bases in the complex are reactive to attack by permanganate, indicating that the central region is

open; (v) affinity of binding of the HMG box to junctions is inversely proportional to the concentration of magnesium ions present.

Thus HMG box proteins form a specific complex with the square open form of the four-way junction and conditions that stabilize the stacked X structure (i.e. addition of magnesium ions) destabilize the complex. In order to understand why the four-way junction serves as a specific binding substrate for HMG box proteins we therefore need to consider the features of the extended square structure that might selectively bind the protein. Moreover, since the major complex comprises a single molecule of HMG box protein bound to the junction, can we understand why binding of one molecule might tend to occlude binding of further HMG boxes with the same selectivity and affinity?

NMR studies of complexes of SRY (Werner *et al.*,

1995) and LEF-1 (Love *et al.*, 1995) bound to their duplex targets have revealed significant distortion of the DNA. In both cases the protein is bound in the minor groove, which undergoes consequent widening and unwinding to accommodate the HMG box. In the four-way junction the minor groove must effectively be widened at the point of strand exchange (von Kitzing *et al.*, 1990) and this region might therefore be predisposed to serve as a binding site for the HMG box. Opening of the minor groove at the centre of the square four-way junction resembles opening of the minor groove seen in the published NMR structures and so suggests an orientation for the HMG box in the complex. The experimental observation that a bulky fusion on the N-terminus of the HMG box did not appreciably alter the mode of binding is further evidence that the region furthest away from the bulky fusion protein, helices 1 and 2, must be in intimate contact with the junction, leaving the N terminus pointing away along the receding duplex.

We used this as a starting point for a computer molecular modelling exercise and generated a model four-way junction in which one arm took the base sequence of the LEF-1 consensus binding site. The binding site was then translated progressively towards the junction and the HMG box was placed on the DNA as dictated by the NMR structure. This was repeated until we reached the limit caused by steric interaction between the LEF-1 HMG box and the DNA of the opposing arm of the junction. The most centrally located binding site was found to be one in which Met10 was intercalated between the last two base pairs at the junction centre and positively charged residues on helices 1 and 2 of the HMG box made favourable electrostatic interactions with the phosphodiester backbone around the junction centre (Figure 7A). Because the model retains the base roll at the intercalation site of Met10 the protein-bound arm retains a small bend, however, the overall shape of the junction remains square, in agreement with the comparative gel electrophoresis results. This position is not conserved and is an alanine in HMG1 box A; the degree of bending that might be induced is not an important feature of this model. The position of α helices 1 and 2 at the centre of the junction accounts for the observed 1:1 binding stoichiometry, which results from occlusion of the other three symmetrical positions on the junction. The HMG box essentially fills the open centre of the junction from the minor groove side, but makes only a small incursion through the centre (Figure 7B and C). Bases at the point of strand exchange are still exposed to the solvent and should be susceptible to attack by permanganate. In contrast, although modelling indicates that the stacked X structure of the four-way junction also has a widened minor groove at the position of strand exchange (von Kitzing *et al.*, 1990), steric hindrance caused by the interleaving of phosphodiester backbones would be expected to prevent HMG box binding.

The HMG box proteins exhibit impressive selectivity for binding to four-way DNA junctions, but is this exploited in the cell? We have shown that the protein binds in the open conformation of the junction and have proposed a structural basis for this. This suggests that recognition of DNA junctions is an accidental consequence of a coincidental similarity in DNA structure to the normal

binding site. Moreover, because the structure recognized is rendered inaccessible in the form of the junction adopted in the presence of magnesium ions, binding to junctions is severely reduced under these conditions and is probably not significant under physiological conditions. In contrast, binding of proteins with a known role in the manipulation of DNA junctions, such as the junction resolving enzymes, bind with full affinity under these conditions (White *et al.*, 1997).

A demonstration of selective binding of a given protein to DNA junctions need not necessarily imply a biological role for the interaction and further experiments are required in order to test this possibility. Nevertheless, this interaction can provide insight into the mechanism of recognition of distorted DNA structure by these proteins in general.

Materials and methods

Expression and purification of HMG box proteins

HMG box A of rat HMG1 was expressed from plasmid pT7-HMG1A (Bianchi *et al.*, 1992) in *Escherichia coli* BL21 (DE3). Cells were grown in LB medium containing 100 μ g/ml ampicillin at 37°C to an A_{660} of 0.6. Cultures were induced with 1 mM isopropyl β -D-thiogalactoside (IPTG) and growth continued for 2 h. Cells were harvested by centrifugation at 4°C and the cell pellet resuspended in 5 ml 50 mM Tris-HCl, pH 7.8, 250 mM NaCl, 1 mM EDTA (sonication buffer). Cells were lysed by sonication (four 15 s bursts on ice) and the lysate cleared by centrifugation at 20 000 g at 4°C for 15 min. Ammonium sulfate was added to a cleared cell lysate to a final concentration of 45% saturation and centrifuged at 14 000 r.p.m. for 15 min at 4°C. Further ammonium sulfate was added to a final concentration of 70% saturation and centrifuged as before. The pellet was resuspended in 5 ml 50 mM phosphate buffer, pH 6.0, and dialysed against the same buffer for 3 h at 4°C. The dialysate was applied to an S-Sepharose column operating in an HPLC apparatus (Perseptive Biosystems) and HMG1 box A eluted using a linear gradient of 0–1 M NaCl. The peak fraction was concentrated and the buffer exchanged using a microconcentrator (Amicon). The protein was stored at –20°C in 50 mM phosphate buffer, pH 7.5, 1 mM DTT, 50% (v/v) glycerol.

Box A of rat HMG1 was also cloned as a fusion with MBP (New England Biolabs). The DNA encoding the HMG1 box was amplified by PCR from the expressing plasmid pT7-7HMG1, using PCR primers that included unique restriction sites to facilitate subsequent subcloning into plasmid pMALc2 (NEB) at the *EcoRI* and *HindIII* sites. Constructs were grown in *E. coli* strain DH5 α F' in LB medium containing 25 μ g/ml kanamycin and 100 μ g/ml carbenicillin at 37°C to an A_{660} of 0.6. Protein expression was induced with IPTG to a final concentration of 0.5 mM and growth continued for a further 2 h. Cells were harvested by centrifugation at 4°C and lysed by sonication as above. The cleared lysate was applied to amylose resin (NEB) and extensively washed with sonication buffer. The protein was eluted with 10 mM maltose in sonication buffer. Peak fractions were pooled and concentrated using a 30 kDa cut-off Centrprep (Amicon). The protein was stored in 50 mM Tris-HCl, pH 7.8, 250 mM NaCl, 30% (v/v) glycerol at –20°C.

The HMG box of human SRY was expressed from pT7-hSRYbox in *E. coli* BL21 (DE3) as previously described (Ferrari *et al.*, 1992). Cells were grown at 37°C in 150 ml LB medium containing 0.2% glucose and 100 μ g/ml ampicillin. Cultures were induced with 0.5 mM IPTG after reaching an A_{660} of 0.6 and growth continued for 90 min. Cells were harvested by centrifugation at 4°C and the cell pellet resuspended in 5 ml 50 mM Tris-HCl, pH 8.0, 0.5 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride, 20 mM EDTA. Cells were lysed by sonication (four 20 s bursts on ice), NaCl was added to 150 mM and DEAE-Sepharose resin (Pharmacia) was mixed in to 50 mg/ml. The equilibrated slurry was centrifuged at 20 000 g at 4°C for 15 min. The supernatant was mixed on ice with solid ammonium sulfate to a final concentration of 2.8 M and centrifuged again. The supernatant was applied to a phenyl-Sepharose column (Pharmacia) and SRY HMG box protein was eluted with a decreasing concentration of ammonium sulfate. Peak fractions were pooled and concentrated using a 10 kDa cut-off Centrprep (Amicon). The protein was stored in 10 mM Na phosphate,

pH 7.5, 100 mM NaCl, 0.5 mM DTT, 0.1 mM EDTA, 5% (v/v) glycerol at -20°C .

HMG box 1 from *Xenopus* UBF was cloned as a hexahistidine fusion separated from the main protein by a thrombin cleavage site, using vector pET15b. The construct was grown in *E. coli* strain BL21 (DE3) pLysS in LB medium containing 60 $\mu\text{g}/\text{ml}$ ampicillin at 37°C to an A_{600} of 0.5. Expression was induced with 0.1 mM IPTG for 4 h. Cells were harvested by centrifugation and ruptured by freeze-thaw and sonication. Soluble components were purified by metal affinity chromatography using imidazole elution. Following removal of the His₆ fusion by thrombin cleavage, the UBF HMG box protein was finally purified by reverse phase HPLC.

Determination of protein concentration

Protein concentrations were determined using the Bradford method (Bradford, 1976), standardized using bovine serum albumin (BSA).

Electrophoretic retardation analysis of HMG box-DNA junction interactions

Samples of purified HMG box proteins were incubated with radioactively $5'$ - ^{32}P -labelled four-way DNA junctions in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT, 100 $\mu\text{g}/\text{ml}$ BSA (binding buffer) containing either 5 mM EDTA or MgCl_2 at various concentrations in a 10 μl total volume for 15 min at 20°C , prior to addition of a 1/5th volume of 35% (v/v) Ficoll type 400, 0.25% (w/v) xylene cyanol, 0.25% (w/v) bromophenol blue. Samples were applied to 5 or 8% polyacrylamide gels (29:1 acrylamide:bisacrylamide) and electrophoresed in 90 mM Tris-borate, pH 8.3, in the presence of either EDTA or MgCl_2 as stated. Electrophoresis buffers were recirculated continuously at 1 l/h. After electrophoresis gels were dried on Whatman 3MM paper and exposed to X-ray film (Konica) for documentation or storage phosphor screens for quantification.

Measurement of binding affinity

Radioactively $5'$ - ^{32}P -labelled four-way DNA junction (1 nM) was incubated with a range of known concentrations of protein in binding buffer and bound complexes and free junction separated by gel electrophoresis as described above. After exposure of dried gels to storage phosphor screens data were analysed as fraction DNA bound (f_b) versus protein concentration and were fitted by non-linear regression analysis to the equation

$$f_b = \frac{\{(1 + K_A P_T + K_A D_T) - \sqrt{[(1 + K_A P_T + K_A D_T)^2 - (4 D_T K_A^2 P_T)]}\}}{(2 D_T K_A)}$$

where P_T is the total protein concentration, D_T is the total DNA concentration and K_A is the association constant. When studying the effect of magnesium ions on protein binding a discrete complex was not always observed, due to dissociation of the protein from the junction within the lifetime of the experiment. In such cases the smear above the level of the free DNA was counted as bound DNA. The dissociation constant (K_D) is the reciprocal of K_A .

Proteolytic digestion of MBP-HMG1 box A fusion protein

The linker region between the MBP polypeptide and the HMG box contains a site for protease factor Xa. Cleavage with factor Xa (NEB) results in liberation of the HMG box from its fusion partner. Partial digestions of the protein were performed at 25°C in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl for 3 h. Aliquots were removed and incubated with radioactively $5'$ - ^{32}P -labelled four-way DNA junctions in binding buffer, followed by electrophoresis as described above.

Potassium permanganate probing

Radioactively $5'$ - ^{32}P -labelled four-way DNA junction 3, uniquely labelled on the X strand, was incubated at room temperature in the presence or absence of MBP-HMG1 box A protein in 20 mM Tris-HCl, pH 8, 50 mM NaCl, 50 $\mu\text{g}/\text{ml}$ calf thymus DNA and either 5 mM EDTA or 100 μM MgCl_2 in a total reaction volume of 25 μl . The reaction was initiated by addition of 5 μl freshly prepared stock 10 mM KMnO_4 and stopped after 1 min by addition of 2 μl β -mercaptoethanol. After ethanol precipitation DNA samples were resuspended in 100 μl 1 M piperidine at 95°C for 30 min. Piperidine was removed by extensive lyophilization. The dried samples were resuspended in formamide and analysed by denaturing gel electrophoresis in 15% (w/v) polyacrylamide containing 7 M urea. After electrophoresis gels were exposed to X-ray film as before.

Oligonucleotide synthesis

Oligonucleotides were synthesized using β -cyanoethyl phosphoramidite chemistry (Beaucage and Caruthers, 1981; Sinha *et al.*, 1984) implemented on a 394 DNA/RNA synthesizer (Applied Biosystems). Fully deprotected oligonucleotides were purified by gel electrophoresis in 12 or 15% (w/v) polyacrylamide containing 7 M urea, DNA eluted from excised bands and recovered by ethanol precipitation. Oligonucleotides were radioactively labelled at their $5'$ -termini using [γ - ^{32}P]ATP and T4 polynucleotide kinase (Amersham).

Assembly of four-way DNA junctions

Stoichiometric quantities of four strands were annealed by incubation in 50 mM Tris-HCl, pH 7.6, 10 mM MgCl_2 , 5 mM DTT, 0.1 mM spermidine and 0.1 mM EDTA for 2 min at 85°C , followed by cooling to room temperature overnight. Junctions were purified by electrophoresis in 5 or 8% (w/v) polyacrylamide gels. DNA was recovered from excised bands by electroelution and ethanol precipitation.

Junction 3. All of the junction binding experiments described here were performed on junction 3 (Duckett *et al.*, 1988). For analysis of binding a version assembled from four strands, each of 30 nt (generating a junction with four arms each of 15 bp), was used. The comparative gel electrophoretic analysis was based on a version of junction 3 with four arms of 40 bp in length (Duckett *et al.*, 1988). Six forms of radioactively $5'$ - ^{32}P -labelled junction 3 comprising two arms of 40 bp and two of 15 bp were derived from this junction.

SRY recognition sequences. Duplexes containing a consensus sequence for binding of SRY (bold type) were generated using the 27 nt strands SRY-top ($5'$ -CCCTGCAGG**TAACAATG**TCGGCTCGGG- $3'$) and SRY-bottom ($5'$ -CCCGCGCCG**ACATTGTT**ACCTGCAGGG- $3'$).

Sequencing of DNA

A sequence marker was generated by pyrimidine-specific cleavage using the hydrazine reaction followed by cleavage with piperidine (Maxam and Gilbert, 1980).

Acknowledgements

We thank Dr Brian McStay for supplying a plasmid containing the UBF gene, our colleagues Drs Marie-Jo Giraud-Panis and Malcolm White for discussions and the Cancer Research Campaign for financial support. M.E.B. and D.M.J.L. were members of the EU Network on HMG Proteins, for which support is gratefully acknowledged.

References

- Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxy-polynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859–1862.
- Bennett, R.J. and West, S.C. (1995) Structural analysis of the RuvC–Holliday junction complex reveals an unfolded junction. *J. Mol. Biol.*, **252**, 213–226.
- Bianchi, M.E. (1995) The HMG-box domain. In Lilley, D.M.J. (ed.), *DNA-Protein: Structural Interactions*. IRL Press at Oxford University Press, Oxford, UK, pp. 177–200.
- Bianchi, M.E., Beltrame, M. and Paonessa, G. (1989) Specific recognition of cruciform DNA by nuclear protein HMG1. *Science*, **243**, 1056–1059.
- Bianchi, M.E., Falciola, L., Ferrari, S. and Lilley, D.M.J. (1992) HMG1 protein contains a DNA binding site composed of two similar segments, which have counterparts in other eukaryotic regulatory proteins. *EMBO J.*, **11**, 1055–1063.
- Bracale, M., Galli, M.G., Savini, C. and Bianchi, M.E. (1994) Specific interaction of plant HMG-like proteins with cruciform DNA. *J. Exp. Bot.*, **45**, 1493–1496.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein dye binding. *Anal. Biochem.*, **72**, 248–254.
- Clegg, R.M., Murchie, A.I.H., Zechel, A., Carlberg, C., Diekmann, S. and Lilley, D.M.J. (1992) Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry*, **31**, 4846–4856.
- Cooper, J.P. and Hagerman, P.J. (1989) Geometry of a branched DNA structure in solution. *Proc. Natl Acad. Sci. USA*, **86**, 7336–7340.
- Duckett, D.R., Murchie, A.I.H., Diekmann, S., von Kitzing, E., Kemper, B.

- and Lilley, D.M.J. (1988) The structure of the Holliday junction and its resolution. *Cell*, **55**, 79–89.
- Duckett, D.R., Giraud-Panis, M.-E. and Lilley, D.M.J. (1995) Binding of the junction-resolving enzyme bacteriophage T7 endonuclease I to DNA: separation of binding and catalysis by mutation. *J. Mol. Biol.*, **246**, 95–107.
- Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.*, **11**, 4497–4506.
- Giese, K., Cox, J. and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein complexes. *Cell*, **69**, 185–195.
- Goodwin, G.H., Sanders, C. and Johns, E.W. (1973) A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. *Eur. J. Biochem.*, **38**, 14–19.
- Grasser, K.D. (1995) Plant chromosomal high mobility group (HMG) proteins. *Plant J.*, **140**, 185–192.
- Hardman, C.H., Broadhurst, R.W., Raine, A.R.C., Grasser, K.D., Thomas, J.O. and Laue, E.D. (1995) Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry*, **34**, 16596–16607.
- Jantzen, H.-M., Admon, A., Bell, S.P. and Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature*, **344**, 830–836.
- Jones, D.N., Searles, M.A., Shaw, G.L., Churchill, M.E., Ner, S.S., Keeler, J., Travers, A.A. and Neuhaus, D. (1994) The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*. *Structure*, **2**, 609–627.
- Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M.E. and Grummt, I. (1994) Functional differences of two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity. *EMBO J.*, **13**, 416–424.
- Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791–795.
- Lumpkin, O.J. and Zimm, B.H. (1982) Mobility of DNA in gel electrophoresis. *Biopolymers*, **21**, 2315–2316.
- Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–560.
- Murchie, A.I.H., Clegg, R.M., von Kitzing, E., Duckett, D.R., Diekmann, S. and Lilley, D.M.J. (1989) Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. *Nature*, **341**, 763–766.
- Parisi, M.A., Xu, B.J. and Clayton, D.A. (1993) A human mitochondrial transcriptional activator can functionally replace a yeast mitochondrial HMG-Box protein both *in vivo* and *in vitro*. *Mol. Cell. Biol.*, **13**, 1951–1961.
- Parkinson, M.J. and Lilley, D.M.J. (1997) The junction-resolving enzyme T7 endonuclease I: quaternary structure and interaction with DNA. *J. Mol. Biol.*, **270**, 169–178.
- Parsons, C.A., Stasiak, A., Bennett, R.J. and West, S.C. (1995) Structure of a multisubunit complex that promotes DNA branch migration. *Nature*, **374**, 375–378.
- Paull, T.T., Haykinson, M.J. and Johnson, R.C. (1993) The nonspecific DNA-binding and DNA-bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.*, **7**, 1521–1534.
- Payet, D. and Travers, A. (1997) The acidic tail of the high mobility group protein HMG-D modulates the structural selectivity of DNA binding. *J. Mol. Biol.*, **266**, 66–75.
- Peters, R., King, C.Y., Ukiyama, E., Falsafi, S., Donahoe, P.K. and Weiss, M.A. (1995) An SRY mutation causing human sex reversal resolves a general mechanism of structure-specific DNA recognition: application to the four-way DNA junction. *Biochemistry*, **34**, 4569–4576.
- Pil, P.M. and Lippard, S.J. (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin. *Science*, **256**, 234–237.
- Pöhler, J.R.G., Giraud-Panis, M.-J.E. and Lilley, D.M.J. (1996) T4 endonuclease VII selects and alters the structure of the four-way DNA junction; binding of a resolution-defective mutant enzyme. *J. Mol. Biol.*, **260**, 678–696.
- Pontiggia, A., Rimini, R., Harley, V.R., Goodfellow, P.N., Lovell-Badge, R. and Bianchi, M.E. (1994) Sex-reversing mutations affect the architecture of SRY–DNA complexes. *EMBO J.*, **13**, 6115–6124.
- Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C. and Norman, D.G. (1993) Solution structure of a DNA-binding domain from HMG1. *Nucleic Acids Res.*, **21**, 3427–3436.
- Read, C.M., Cary, P.D., Preston, N.S., Lnenicek-Allen, M. and Crane-Robinson, C. (1994) The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. *EMBO J.*, **13**, 5639–5646.
- Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C., Carrillo, M.O.M. and Norman, D.G. (1995) The structure of the HMG box and its interaction with DNA. In Eckstein, F. and Lilley, D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Germany, Vol. 9, pp. 222–250.
- Sasse-Dwight, S. and Gralla, J.D. (1989) KMnO₄ as a probe for *lac* promoter DNA melting and mechanism *in vivo*. *J. Biol. Chem.*, **264**, 8074–8081.
- Sinclair, A. *et al.* (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240–244.
- Sinha, N.D., Biernat, J., McManus, J. and Koster, H. (1984) Polymer support oligonucleotide synthesis XVIII: use of β -cyanoethyl-*N,N*-dialkylamino/*N*-morpholino phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acids Res.*, **12**, 4539–4557.
- Teo, S.H., Grasser, K.D., Hardman, C.H., Broadhurst, R.W., Laue, E.D. and Thomas, J.O. (1995) Two mutations in the HMG-box with very different structural consequences provide insights into the nature of binding to four-way junction DNA. *EMBO J.*, **14**, 3844–3853.
- Travis, A., Amsterdam, A., Berlinger, C. and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein, with an HMG domain, regulates T-cell receptor α enhancer function. *Genes Dev.*, **5**, 880–894.
- van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.*, **10**, 123–132.
- von Kitzing, E., Lilley, D.M.J. and Diekmann, S. (1990) The stereochemistry of a four-way DNA junction: a theoretical study. *Nucleic Acids Res.*, **18**, 2671–2683.
- Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R.C., Laue, E.D. and Thomas, J.O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.*, **12**, 1311–1319.
- Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY–DNA complex. *Cell*, **81**, 705–714.
- White, M.F. and Lilley, D.M.J. (1996) The structure-selectivity and sequence-preference of the junction-resolving enzyme CCE1 of *Saccharomyces cerevisiae*. *J. Mol. Biol.*, **257**, 330–341.
- White, M.F. and Lilley, D.M.J. (1997) The resolving enzyme CCE1 of yeast opens the structure of the four-way DNA junction. *J. Mol. Biol.*, **266**, 122–134.
- White, M.F., Giraud-Panis, M.-J.E., Pöhler, J.R.G. and Lilley, D.M.J. (1997) Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Mol. Biol.*, **269**, 647–664.

Received October 21, 1997; revised November 12, 1997;
accepted November 13, 1997