The EMBO Journal vol.14 no.6 pp.1264–1275, 1995

HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG

Mair E.A.Churchill[2], David N.M.Jones[1,2], Theodore Glaser, Heidi Hefner, M.Alexandra Searles[1] and Andrew A.Travers[1,2]

Department of Cell and Structural Biology and the Biophysics Program, University of Illinois, Urbana, IL 61801, USA and [1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[2]Corresponding authors

Communicated by A.A.Travers

The high mobility group (HMG) protein HMG-D from *Drosophila melanogaster* is a highly abundant chromosomal protein that is closely related to the vertebrate HMG domain proteins HMG1 and HMG2. In general, chromosomal HMG domain proteins lack sequence specificity. However, using both NMR spectroscopy and standard biochemical techniques we show that binding of HMG-D to a single DNA site is sequence selective. The preferred duplex DNA binding site comprises at least 5 bp and contains the deformable dinucleotide TG embedded in A/T-rich sequences. The TG motif constitutes a common core element in the binding sites of the well-characterized sequence-specific HMG domain proteins. We show that a conserved aromatic residue in helix 1 of the HMG domain may be involved in recognition of this core sequence. In common with other HMG domain proteins HMG-D binds preferentially to DNA sites that are stably bent and underwound, therefore HMG-D can be considered an architecture-specific protein. Finally, we show that HMG-D bends DNA and may confer a superhelical DNA conformation at a natural DNA binding site in the *Drosophila fushi tarazu* scaffold-associated region.

Key words: *cis*-DDP/DNA bending/DNA specificity/footprinting/HMG

Introduction

Highly abundant chromosomal proteins must be able to bind many different DNA sequences with moderate to high affinity, in distinct contrast to other proteins, such as transcription factors, which bind DNA in a site- or sequence-specific fashion. However, the mechanism by which chromosomal proteins recognize DNA in general is poorly understood (von Hippel and Berg, 1986; von Hippel, 1994). A large family of such proteins, typified by HMG1 and HMG2, contain a common DNA binding motif known as the HMG domain (Jantzen *et al.*, 1990; Ner, 1992; Landsman and Bustin, 1993; Grosschedl *et al.*, 1994). This domain is also found in less abundant sequence-specific DNA binding proteins, such as the transcription factors LEF-1 and UBF and the mammalian testis-determining factor SRY (Jantzen *et al.*, 1990; Nasrin *et al.*, 1991; Ferrari *et al.*, 1992; Giese *et al.*, 1992). Members of both classes of HMG domain bind preferentially to distorted DNA sites such as synthetic four-way DNA junctions (Bianchi *et al.*, 1992) and those formed by adducts between DNA and the anticancer drug *cis*-diamminedichloroplatinum (*cis*-DDP). Consequently, they have been termed architecture-specific proteins (Bruhn *et al.*, 1992, 1993; Pil and Lippard, 1992; Grosschedl *et al.*, 1994). In addition, they can induce bends in linear duplex DNA (Giese *et al.*, 1992; Paull *et al.*, 1993; Pil *et al.*, 1993) and can constrain negative supercoils in plasmid DNA (Sheflin and Spaulding, 1989; Megraw and Chae, 1993; Sheflin *et al.*, 1993; Bazett-Jones *et al.*, 1994; Ner *et al.*, 1994; Stros *et al.*, 1994). Furthermore, they can promote the formation of higher order nucleoprotein complexes (Crothers, 1993; Paull *et al.*, 1993; Grosschedl *et al.*, 1994; Travers *et al.*, 1994), a property also characteristic of a bacterial chromosomal protein HU (Drlica and Rouviere-Yaniv, 1987; Haykinson and Johnson, 1993).

Although the primary role of the prototypical HMG1/2 proteins *in vivo* is not known (Landsman and Bustin, 1993), the homologous protein from *Drosophila melanogaster*, HMG-D (Ner, 1992; Wagner *et al.*, 1992), is associated with mitotic chromosomes during the earliest stages of embryogenesis, at a time when histone H1 is absent. As H1 accumulates and the relative concentration of HMG-D decreases, there is a correlated decrease in the size of the chromatids. It has been proposed that HMG-D may stabilize a less condensed state of chromatin than that containing H1. This less condensed state could facilitate the rapid cycles of cell division that occur prior to blastoderm formation (Ner and Travers, 1994).

HMG-D is closely related to the vertebrate HMG1 and 2 proteins but, in contrast, it contains only a single copy of the prototypical DNA binding domain of 74–80 residues (Landsman and Bustin, 1991, 1993), a feature usually associated with the sequence-specific HMG domain proteins such as LEF-1 and SRY (Giese *et al.*, 1991; Nasrin *et al.*, 1991; Alexander-Bridges *et al.*, 1992; Ner, 1992). The HMG domain is followed by a 'tail' region that has basic sequences similar to the C-terminal domain of histone H1 and a short C-terminal acidic stretch similar to those seen in HMG1/2. The NMR structure of the HMG domain from HMG-D is very similar to the previously determined structure of the B-domain of HMG-1, which revealed the characteristic L-shaped fold formed by three α-helices (Weir *et al.*, 1993; Read *et al.*, 1993; Jones *et al.*, 1994). Precisely how the HMG domain binds to DNA is not known, as no structure of an appropriate protein–DNA complex is available. However, footprinting and mutagenesis experiments suggest that HMG domain proteins bind to the minor groove (Giese *et al.*, 1991; van

de Wetering *et al.*, 1993). In addition, NMR experiments show that at least one residue may partially intercalate into the DNA, via the minor grove (King and Weiss, 1993).

Here we describe DNA binding studies of HMG-D. These show that HMG-D is an architecture-specific protein and prefers to bind to deformable DNA sites. Binding site selection methods, DNase I footprinting experiments and competitive gel electrophoretic mobility shift assays (EMSA) revealed that HMG-D prefers to bind to A/T-rich DNA containing the dinucleotide sequence TG. The TG site is a readily deformable DNA base step and consistent with this we show that HMG-D binds preferentially to pre-bent DNA sites and induces substantial bends in DNA. NMR studies of a complex of HMG-D with DNA containing this preferred sequence show that the region of DNA affected by binding extends over 5–6 bp and includes the TG dinucleotide site and that a highly conserved aromatic residue may interact with this TTG site.

Results

HMG-D prefers to bind to DNA containing 'TG' sites

HMG-D was isolated from a screen of a *Drosophila* λgt11 expression library (Ner *et al.*, 1993) under conditions that were favorable for the selection of a sequence-specific protein; a specific fragment of the *Drosophila fushi tarazu* (*ftz*) scaffold-associated regions (SAR) (Amati *et al.*, 1990; Harrison and Travers, 1988) was used to probe this library in the presence of a vast excess of carrier DNA (Ner *et al.*, 1993). However, in gel mobility shift assays HMG-D gives rise to a ladder of bands characteristic of non-sequence-selective DNA binding proteins such as *Escherichia coli* HU, FIS and *Eco*RV (in the absence of magnesium) (Bonnefoy and Rouviere-Yaniv, 1992; Lazarus and Travers, 1993; Vipond and Halford, 1993). In order to resolve this apparent contradiction, we decided to investigate whether HMG-D shows any preference for binding to particular DNA sequences.

A binding site selection protocol was used to search for preferred binding sites for HMG-D (Pollock and Treisman, 1990). A specific sequence or even short binding site was not selected using this procedure, as determined by scanning 32 of the selected sequences either by eye or by using the Motif Exploration program (MEP; Staden, 1989). However, analysis of the selected sequences using the Dinucleotide Frequency program (DNF; Satchwell *et al.*, 1986) revealed a striking preference for certain patterns of dinucleotides (Table I). The DNA sequences selected by HMG-D contain a significantly higher (nearly 3-fold) proportion of TG, TT, GT and GG motifs in the randomized region than the complementary dinucleotide sequences on a single strand. It is possible that this non-random distribution in the 'top strand' may arise from bias in the original oligonucleotide synthesis of the first strand, but this has not previously been observed. Interestingly, three of the most common dinucleotides found in the binding sites of the sequence-specific HMG domain proteins are TT, TG and GT (Landsman and Bustin, 1993; Grosschedl *et al.*, 1994) (Table II). In a similar study using xUBF, no preferred binding sites were observed (Copenhaver *et al.*, 1994), but DNF analysis reveals a slight preference for TT, TG and GT steps.

Table I. Dinucleotide frequencies seen in selected inserts

Dinucleotide	Frequency[a] (%)	Dinucleotide	Frequency[a] (%)
GT	9.8	AC	3.2
TG	9.1	CA	3.8
GG	8.6	CC	3.2
TT	7.4	AA	4.8
AG	7.1	CT	6.0
GA	6.9	TC	7.0
GC	5.3	CG	6.0
AT	5.8	TA	5.6

[a]Top strand of inserts: CG−$(N)_n$−GA and average dinucleotide frequency should be 6.25%.

Table II. Binding sites for HMG-box domains

HMG domain protein	Binding site sequence
LEF-1, TCF-1	CCT**TTG**AAGCT
SRY	CCA**TTG**TTCT
MatMc	CT**TTG**TT
sox-4	CT**TTG**TT
sox-5	CT**TTG**TT or
	A**TTG**TT
IRE-ABP	CCT**TTG**AA
Rox1	YYA**TTG**TTCT
Ste11	TTCT**TTG**TTY
DP1	TC**TG**AA
HMG-D	TA**TTG**
Consensus	YYA**TTG**AA
	T TT

The TG dinucleotide is the most conserved sequence among the well-characterized sequence-specific HMG domain DNA binding sites (Landsman and Bustin, 1993; Bazett-Jones *et al.*, 1994). Therefore, a series of band shift/competition experiments were performed to confirm whether HMG-D exhibits a preference for binding to DNA containing a TG dinucleotide embedded in an A/T-rich sequence. In the absence of competitor DNA, HMG-D binds to a long DNA probe (>60 bp) giving a ladder of bands, indicating the formation of higher order protein−DNA complexes (Figure 1B) with an apparent affinity of ~50–200 nM (data not shown). However, in the presence of a 200-fold excess of a successful, specific competitor DNA (Figure 1A) having an affinity of ~1 μM (data not shown), the relative amount of higher order complexes formed decreases (Figure 1B). At HMG-D concentrations of 100 and 250 nM (Figure 1B) the sequences that compete better in this assay are those oligonucleotides which contain the sequence TTG (oligomers TTG and TATTG). In comparison, the A and ATA DNA fragments compete approximately half as well, suggesting that HMG-D indeed favors binding to the TG site over the AA, AT or TA sequences found in an equivalent position in the other competitor oligonucleotides.

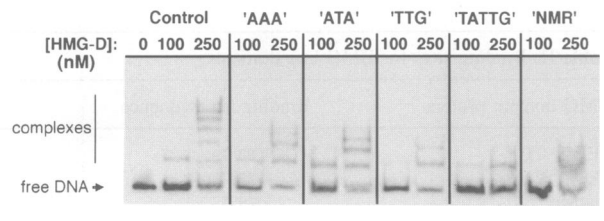
HMG-D contacts the DNA at a TATTG site

In order to study the mode of binding of HMG-D, we used this preferred DNA sequence for NMR experiments. The minimum size of the oligonucleotide required for

A

Name	Variable sequence (21 bp)
AAA	CGAAAAAAAAAAAAAAAAACG
ATA	CGAAAAAAAATAAAAAAAACG
TTG	CGAAAAAAATTGAAAAAAACG
TATTG	CGCAAAAATATTGAAAAACGC
	Variable length (TATTG)
TG10	GCTATTGGAG
TG12	GCATATTGAACG
TG14	GCGATATTGAACGC
TG15	GCGAATATTGAACGC
TG16A	GCGAATATTGAAACGC
TG16B (NMR)	GCTTATTGAAAAATCG

B



C

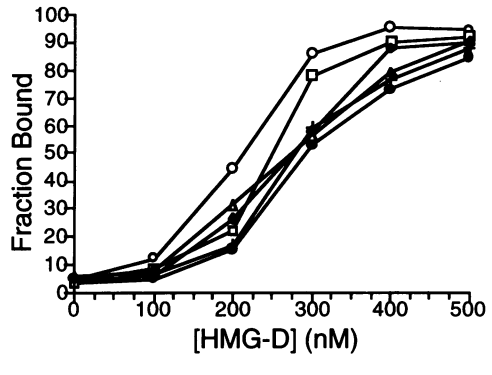


Fig. 1. HMG-D competitive electrophoretic mobility shift assays. (**A**) Table showing sequences and lengths of oligonucleotides used as unlabeled competitor DNA fragments. (**B**) Composite diagram of images from competition experiments with the competitor DNA shown in (A). Labeled 92 bp *ftz* SAR DNA (1 nM) and unlabeled 21 bp duplex DNA molecules and the NMR 16mer fragment (200 nM) were incubated together with a series of concentrations of HMG-D. Only two HMG-D concentrations, 100 and 250 nM, from each experiment are shown. (**C**) Plot of competition binding curves for different length DNA fragments containing the sequence TTG. The saturation binding curves were calculated from experiments as in (B). HMG-D bound to 92 bp *ftz* SAR DNA with no competitor (○) and with 10mer (□), 12mer (△), 14mer (+), 15mer (◆), 16mer (●) competitor oligonucleotides.

HMG-D–DNA complex formation was determined by competition assays of DNA containing the favored TATTG sequence. These showed that the DNA fragment must be at least 12 bp in length to compete effectively with the labeled probe DNA (Figure 1C) and the 16 bp oligomers gave the best competition of higher order complexes (data not shown). These values agree well with a site size of 14 bp determined for LEF-1 and HMG 1 by DNase I footprinting and fluorescence quenching respectively (Butler *et al.*, 1985; Giese *et al.*, 1991). Consequently, the 16 bp oligonucleotide (TG16B) and a 99 residue construct of HMG-D which lacks the acidic tail (see Materials and methods) were used to prepare samples of the complex for NMR studies.

Sequence-specific assignments of the HMG domain of the free protein (residues 2–74) have been reported elsewhere (Jones *et al.*, 1994). In HMG-D-100 only the resonances from the four C-terminal residues of the HMG domain (71–74) shift significantly compared with the 74 residue construct. Sequential connectivities between these four residues and several residues in the basic tail are observed in an HSQC-NOESY spectrum, which suggests that helix 3 may persist for an additional 3–4 residues compared with the isolated domain. The remaining residues in the tail are unstructured, as judged by the narrow line widths, lack of chemical shift dispersion and tertiary NOE connectivities. Sequential assignments of the imino proton resonances of TG16B were obtained from a NOESY spectrum recorded at 10°C (data not shown) and were confirmed from NOE connectivities between the non-labile base protons and the 1′ and 2′/2″ ribose ring protons. The remaining sequence-specific assignments were made using a combination of DQF-COSY, NOESY and TOCSY spectra.

A series of 1-dimensional spectra of the complex acquired with increasing protein:DNA ratios shows selective broadening of several imino protons even at the lowest concentrations of protein used (~0.3 mM). However, in spectra with equimolar or excess protein there is a dramatic broadening of all resonances in the spectrum, even in comparison with spectra of the complex with a protein:DNA ratio of 0.85. This observation is consistent with the control band shift assays of the NMR samples, which show that single DNA–protein complexes form for protein:DNA ratios from 0.3 to 1, whereas higher order complexes appear at ratios >1 (data not shown). Consequently, all further experiments used samples that contained excess DNA (protein:DNA ratio ~0.85).

Figure 2A shows the sequential connectivities between the imino proton and H2 resonances of bp 3–9 in the free DNA. In the complex, the resonances from bp 4–8 are broadened and no sequential connectivities are observed between them (Figure 2B). In addition, the imino proton of T9′ is much broader than those of the remaining base pairs. This broadening most probably results from exchange between free and bound DNA, as almost identical patterns of broadening are observed for other DNA resonances. The resonances from bp 1–3 and 10–16 exhibit only small changes in line width, consistent with formation of a higher molecular weight complex. Sequential connectivities are observed between all these latter base pairs in the complex.

A diagram of the 'footprint' established from the NMR data (Figure 3) shows the location of those base protons and ribose 1′ and 2′/2″ protons which broaden significantly on binding DNA. Little information can be obtained for the 3′, 4′ and 5′/5″ ribose protons, due to overlap with protein resonances. This 'footprint' indicates that HMG-D binds preferentially to a single site located between residues 4 and 9 of TG16B, whilst bp 1–3 and 10–16 are essentially unaffected by binding. Therefore, we conclude from the NMR data that HMG-D shows a high preference for the sequence TTATTG over AAAATCG.

Tyr12 of HMG-D is directly implicated in binding DNA

In the spectra of the complex the protein resonances show characteristically different behavior to the DNA

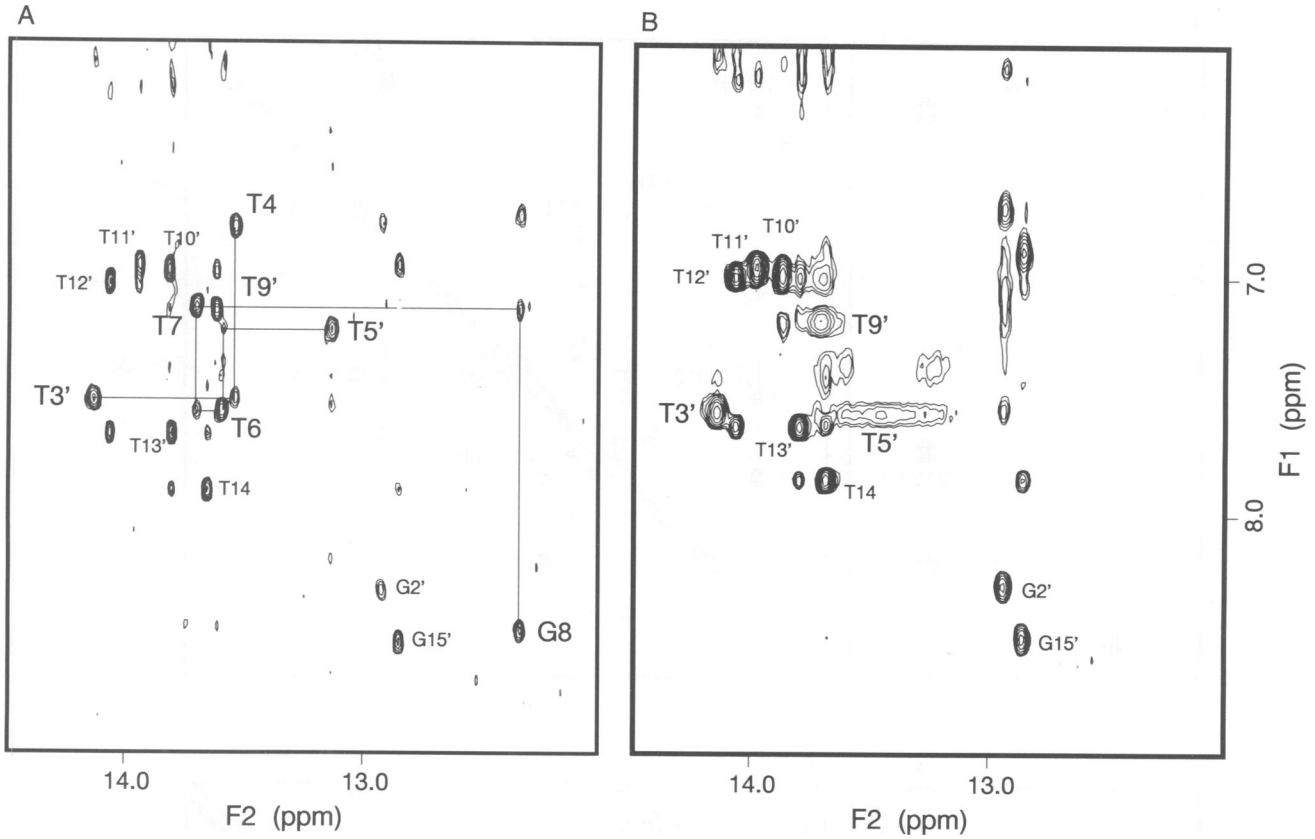


Fig. 2. Regions of NOESY spectra recorded in H_2O showing the sequential connectivities observed between the imino protons and the H2 protons of bp 3–8 of oligonucleotide TG16B (**A**) Free DNA and (**B**) DNA in the complex formed with HMG-D. The intra-residue connectivity is labeled and the sequential inter-residue connectivities are indicated by a black line.

resonances. For example, the indole proton from each tryptophan is doubled (Figure 4B), in contrast to their appearance in the free protein (Figure 4A). This doubling is observed for many other protein resonances and suggests that they are in slow exchange on the NMR time scale. Furthermore, both resonances from the indole proton of Trp 43 are shifted downfield relative to the free protein (Figure 4B). This suggests that the protein is exchanging between two forms, neither of which is directly comparable to the free form. In the presence of excess DNA all the protein is bound and so we conclude that two distinct conformations of the protein may be involved in binding DNA. However, the precise nature of these different conformations is not yet understood.

In a $^1H-^{15}N$ HSQC spectrum of the complex (data not shown) a large number of backbone amide resonances are doubled, including a significant number from residues in the basic tail, which suggests that these residues are also directly involved in binding DNA. However, the complicated nature of the spectra precluded a complete determination of the structure of the protein–DNA complex. Despite this problem, the spectra can be simplified and information about non-exchangeable protons can be obtained by exchanging the complex into D_2O.

In a NOESY spectrum of the free protein recorded in D_2O a large number of NOE connectivities are observed between the aromatic resonances (Figure 4C) and aliphatic resonances in other parts of the spectrum. In a spectrum of the complex, many of these resonances are doubled (Figure 4B and D). Furthermore, no resonances are

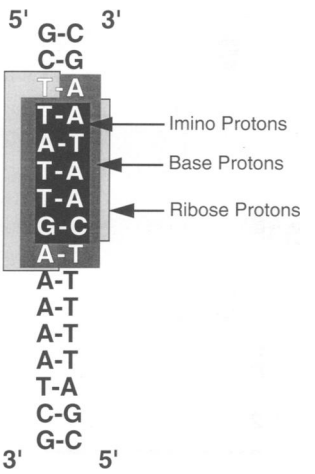


Fig. 3. Schematic representation of TG16B highlighting those nucleotides which show broadening of resonances on binding to HMG-D. The different types of protons are represented by different shaded boxes as indicated.

observed for Tyr12 and all of the connectivities from Trp15 and Trp43 are severely attenuated. In contrast, connectivities from Tyr63 and Phe70 in the C-terminal helix are essentially unaffected by DNA binding. There is no apparent doubling of the resonances from these residues and essentially the whole set of NOE connectivities observed in the free protein are also present in the spectrum of the complex. In these studies we observe that

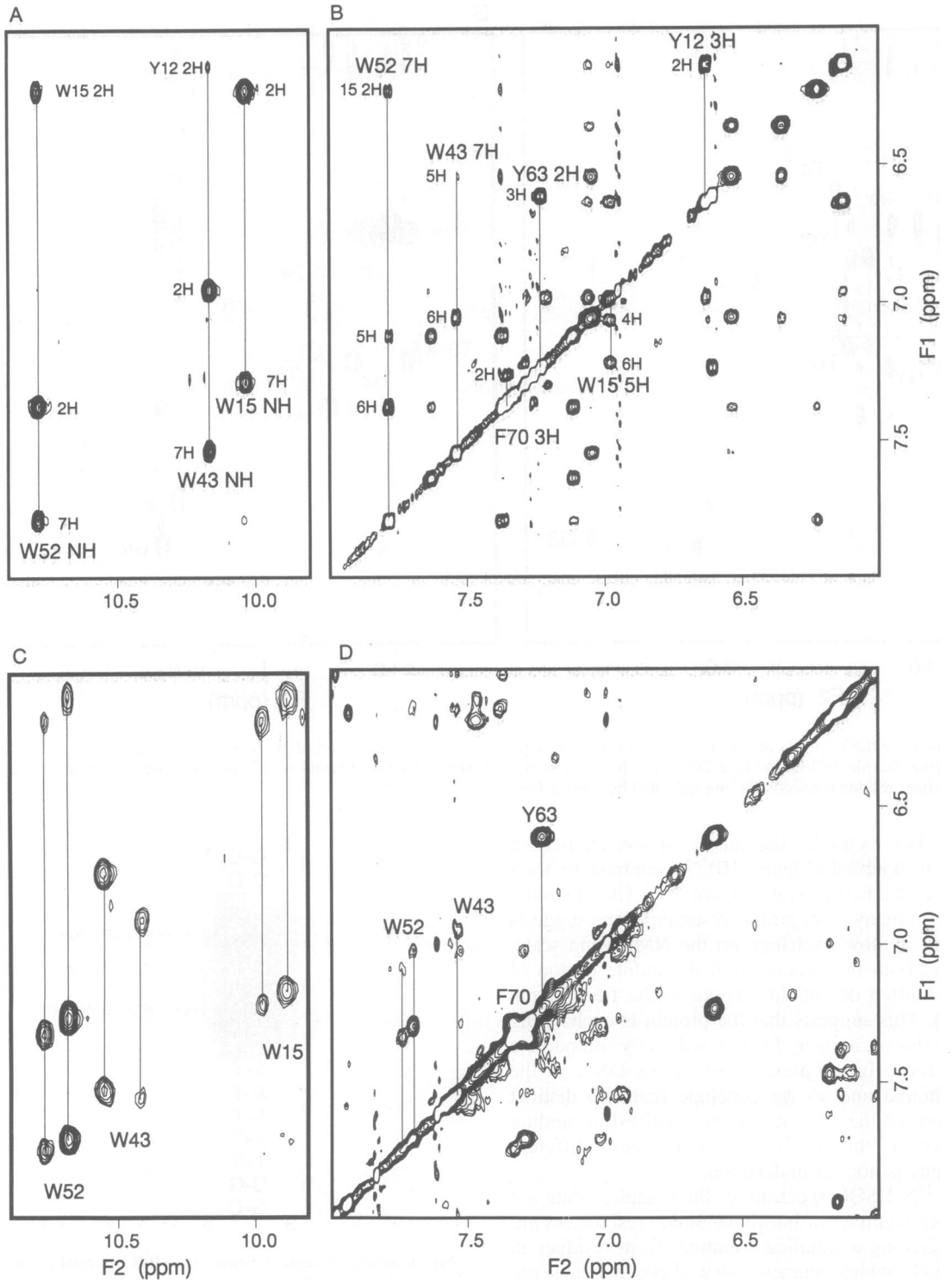


Fig. 4. Regions of NOESY spectra recorded in H_2O showing connectivities between the indole protons of the three tryptophans in (**A**) the free protein and (**B**) the HMG-D–DNA complex and regions of spectra recorded in D_2O showing NOE connectivities between aromatic ring protons in (**C**) free protein and (**D**) HMG-D–DNA complex. Selected protons are labeled (large type) and the patterns of NOE connectivities shown by a line. The identity of the other proton involved is indicated. Note that the resonances from Tyr and Phe residues are non-stereospecifically assigned.

the binding of DNA only significantly affects residues in the main hydrophobic core and in the basic tail regions.

Binding sites of HMG-D in the ftz SAR DNA

Despite NMR evidence for a site-selective, albeit weakly bound, HMG-D–DNA complex and the successful selection for TG-containing sites, we were unable to detect binding sites for HMG-D using standard DNA footprinting methods on the seven 'selected' DNA fragments that we assayed. However, we did see effects of HMG-D binding on the DNase I cleavage pattern with the DNA that was originally used to isolate HMG-D from the *Drosophila* expression library, region 421–442 of the *ftz* SAR (Harrison and Travers, 1988; Amati *et al.*, 1990; Ner *et al.*, 1993).

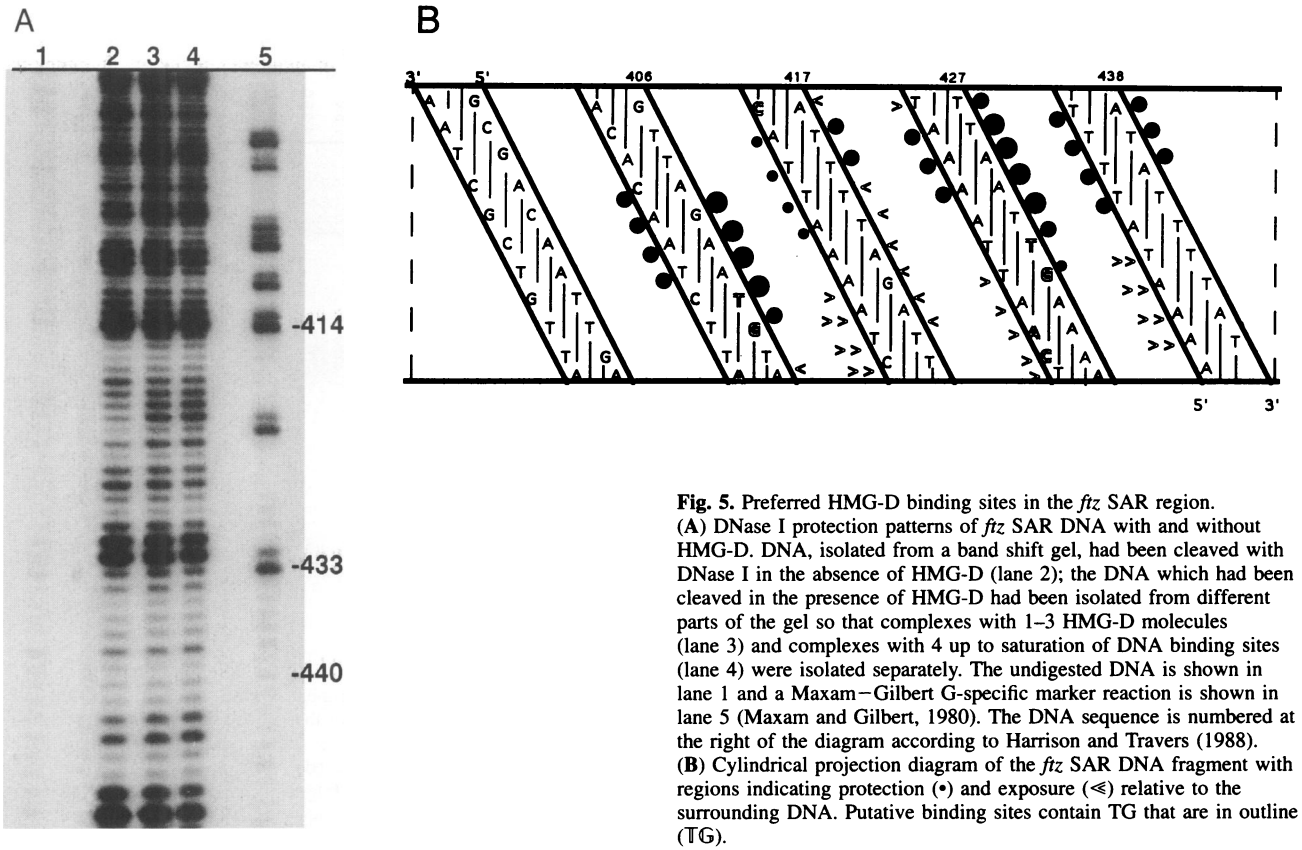


Fig. 5. Preferred HMG-D binding sites in the *ftz* SAR region. (**A**) DNase I protection patterns of *ftz* SAR DNA with and without HMG-D. DNA, isolated from a band shift gel, had been cleaved with DNase I in the absence of HMG-D (lane 2); the DNA which had been cleaved in the presence of HMG-D had been isolated from different parts of the gel so that complexes with 1–3 HMG-D molecules (lane 3) and complexes with 4 up to saturation of DNA binding sites (lane 4) were isolated separately. The undigested DNA is shown in lane 1 and a Maxam–Gilbert G-specific marker reaction is shown in lane 5 (Maxam and Gilbert, 1980). The DNA sequence is numbered at the right of the diagram according to Harrison and Travers (1988). (**B**) Cylindrical projection diagram of the *ftz* SAR DNA fragment with regions indicating protection (•) and exposure (≪) relative to the surrounding DNA. Putative binding sites contain TG that are in outline (TG).

The DNase I footprint of one strand of the 400–455 region of the *ftz* SAR is shown (Figure 5A). This strand exhibits weak, but significant, perturbations compared with free DNA. These footprints and others on both strands were analyzed quantitatively and the patterns mapped on the helical projection diagram in Figure 5B (Churchill and Suzuki, 1989; Smith and Thomas, 1990). This diagram shows the regions on the top strand, for example, centered at 411–413, 418–419, 429–430 and 439 that are slightly protected from DNase I cleavage and 421–424 that are hypersensitive to DNase I cleavage in the presence of HMG-D.

Based on the dimensions and positions of the protected regions and hypersensitive sites, it is reasonable to conclude that two HMG-D binding sites contribute to the footprint. This is also consistent with the sequence and extent of the binding site determined by NMR spectroscopy. These regions are 19 bp apart and each 'site' contains a single TG dinucleotide step located in the 3′ part of the protected region. The TG dinucleotide is on the same strand in both sites and the protection patterns are oriented in the same direction, suggestive of a tandem arrangement of binding. At exactly the mid-point between the two sites is a region of enhanced DNase I cleavage (421–424), which is indicative of DNA distortion (Suck *et al.*, 1988). It is this same region (~422) that appears to be hypersensitive in the DNase I protection patterns previously obtained with stage-dependent *Drosophila* embryo extracts (Harrison and Travers, 1988).

HMG-D prefers to bind to 'bent' DNA and bends DNA

Both the site-specific and chromosomal HMG domain proteins have been shown to bind preferentially to bent or structured DNA, such as synthetic Holliday junctions (Bianchi *et al.*, 1992; Ferrari *et al.*, 1992; Giese *et al.*, 1992; Pontiggia *et al.*, 1993) and DNA containing *cis*-DDP lesions (Bruhn *et al.*, 1992, 1993; Pil *et al.*, 1993). Therefore, we expected that HMG-D may also prefer to bind to structured DNA and used gel mobility shift assays with two 'structured' DNA probes to determine whether HMG-D preferentially recognizes bent DNA. We found that HMG-D had no significant preference for a four-way junction formed by four 17 nucleotide strands (J1) (Churchill *et al.*, 1988) compared with the control 17 bp linear duplexes (data not shown), although we did find that HMG-D has a 2- to 3-fold higher affinity for platinated DNA than for unplatinated DNA in an assay of a 24 nucleotide fragment containing a *cis*-DDP lesion (Figure 6A and B); the estimated dissociation constants are ~200 and ~560 nM respectively. Since DNA containing a single *cis*-DDP lesion is bent by about 30° and untwisted by about 13°, depending on the sequence, HMG-D appears to preferentially bind to bent and underwound DNA (Bellon and Lippard, 1990; Bellon *et al.*, 1991).

Other HMG domain proteins have also been shown to bend and possibly unwind DNA (Sheflin and Spaulding, 1989; Ferrari *et al.*, 1992; Haykinson and Johnson, 1993; Paull *et al.*, 1993; Pil *et al.*, 1993; Sheflin *et al.*, 1993). The circularization rate assay is the most effective method for determining whether a non-sequence-specific protein can bend DNA (Kotlarz *et al.*, 1986; Haykinson and Johnson, 1993; Paull *et al.*, 1993). In this assay the *E.coli* protein HU bends DNA (Haykinson and Johnson, 1993), whereas the drug Hoechst 33258 slows down the circularization rate (M.E.A.Churchill, unpublished data). We compared the circularization rate of the 169 bp *tyrT*

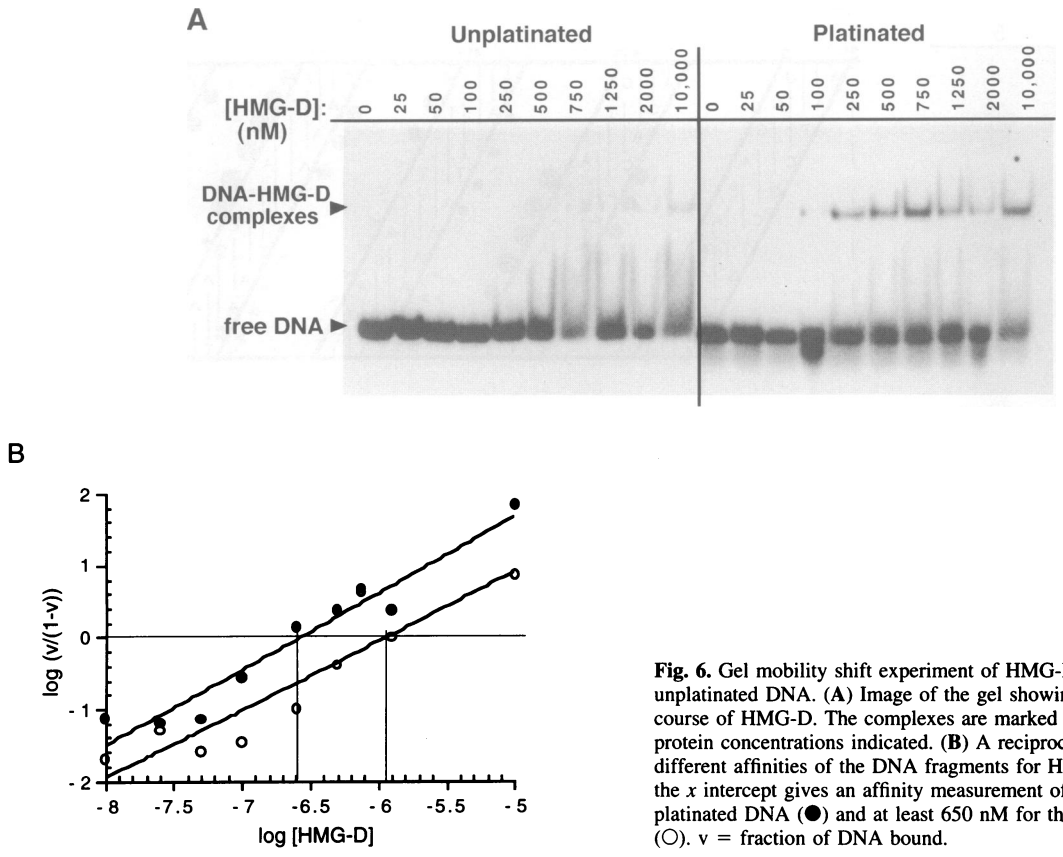


Fig. 6. Gel mobility shift experiment of HMG-D with platinated and unplatinated DNA. (**A**) Image of the gel showing a concentration course of HMG-D. The complexes are marked with arrows and the protein concentrations indicated. (**B**) A reciprocal plot illustrates the different affinities of the DNA fragments for HMG-D. The value of the *x* intercept gives an affinity measurement of ~250 nM for the platinated DNA (●) and at least 650 nM for the unplatinated DNA (○). v = fraction of DNA bound.

RNA promoter DNA restriction fragment (Drew and Travers, 1985) in the presence of HMG-D and control proteins with Hoechst 33258 to determine the relative degrees of bending (Haykinson and Johnson, 1993; Pil *et al.*, 1993). The rate of circularization was compared by examining the accumulation of circular products of the ligation reaction as a function of time by native PAGE (Figure 7A). The plot of the circularization rates (Figure 7B) shows that compared with the control, Hoechst 33258 has a slower rate, whereas HMG-D at 75 nM and HU at 50 nM have faster rates (2-fold) and are comparable under these conditions. Therefore HMG-D can bend DNA, but appears to be less effective than HU.

Discussion

We have shown that HMG-D, in common with other HMG domain proteins, binds preferentially to pre-bent DNA and facilitates DNA bending. Both these properties are characteristic of architecture-specific proteins. More importantly, the results of competition binding studies, NMR spectroscopy and DNase I footprinting indicate that HMG-D exhibits sequence selectivity in binding to duplex DNA and favors sites containing a TG dinucleotide. Finally, binding of HMG-D to at least one arrangement of preferred sites gives rise to a multiprotein–DNA complex that can be detected using DNase I footprinting. This is the first characterization of a preferred DNA binding site of a predominantly architecture-specific chromosomal protein.

Implications of 'TG' sites in DNA recognition

The TG dinucleotide step identified in our study is also found in the DNA binding sites of all well-characterized sequence-specific HMG domain proteins (Alexander-Bridges *et al.*, 1992; Landsman and Bustin, 1993), usually within the sequence TTG (Table II). The extent of the NMR footprint of HMG-D also agrees with the principal contact regions of the sequence-specific HMG domain proteins LEF-1 (Giese *et al.*, 1991), SRY (Nasrin *et al.*, 1991), TCF-1 (van de Wetering and Clevers, 1992) and MatMc (Dooijes *et al.*, 1993), as defined by mutagenesis and interference studies (Giese *et al.*, 1991, 1992), except that the sequence-specific binding sites, as defined by footprinting, extend 2 bp further on the 3′ side of the TTG sequence. We therefore provide experimental evidence for a role of TTG as a core recognition element for HMG domain proteins, with the exception of UBF, where at least one of the HMG domain binding regions in the rRNA promoter lacks such a sequence (Landsman and Bustin, 1993; Leblanc *et al.*, 1993).

We have shown that HMG-D, in common with other HMG domain proteins, facilitates the circularization of linear DNA duplexes and thus bends DNA, consistent with the recognition of a deformable DNA region. Our data further suggest that HMG-D does not bind as well to conformationally rigid sequences, as typified by poly-(dA–dT) sequences. The pyrimidine–purine dinucleotide steps TA and TG are both frequently found in deformable regions of DNA (Travers, 1991) and both exhibit fast rates of imino proton exchange in appropriate sequence contexts (Cheung *et al.*, 1984; Leroy *et al.*, 1988;

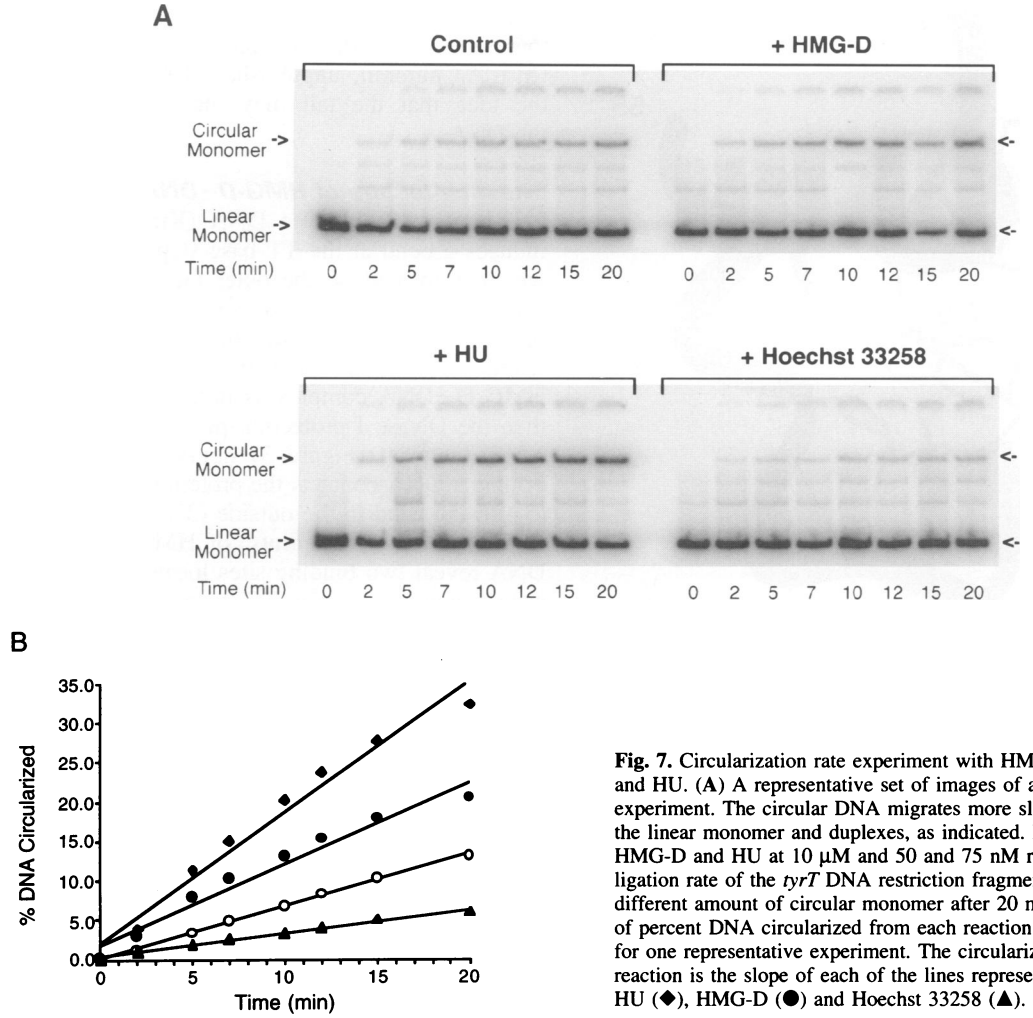


Fig. 7. Circularization rate experiment with HMG-D, Hoechst 33258 and HU. (**A**) A representative set of images of a time course ligation experiment. The circular DNA migrates more slowly on the gel than the linear monomer and duplexes, as indicated. Hoechst 33258, HMG-D and HU at 10 μM and 50 and 75 nM respectively affect the ligation rate of the *tyrT* DNA restriction fragment as seen by the different amount of circular monomer after 20 min ligation. (**B**) A plot of percent DNA circularized from each reaction as a function of time for one representative experiment. The circularization rate for each reaction is the slope of each of the lines representing the control (○), HU (◆), HMG-D (●) and Hoechst 33258 (▲).

Folta-Stogniew and Russu, 1994). The TG step in particular is frequently found at the site of severe bending induced by proteins (Travers and Klug, 1987; Travers, 1991); for example, binding of *E.coli* catabolite gene activator protein (CAP) induces a bend of ~90° in its target DNA (Schultz *et al.*, 1991), which results almost completely from a 40° bend at each of two TG steps 10 bp apart. However, the conservation of the TG step in the known HMG domain binding sites not only reflects a requirement for deformability of the DNA, but also suggests that some recognition feature of the TG dinucleotide allows even the non-specific HMG domains to discriminate between sites containing TG and TA steps.

Protein–DNA interactions

King and Weiss have shown that binding of SRY is, in part, mediated by partial intercalation of an isoleucine side chain at a TT base step within the sequence element GATTG (King and Weiss, 1993). This isoleucine is situated towards the N-terminus of helix 1 and corresponds to residue Met13 in HMG-D (Haaq *et al.*, 1994). In the NMR spectrum of the complex no signals are observed from the ring protons of the neighboring residue Tyr12. This observation is consistent with the involvement of Tyr12 in binding DNA. The estimated dissociation constant for the complex is ~600 nM. If we assume that formation of the complex is diffusion controlled, then we expect that the components of the complex will be exchanging on the micro- to millisecond time scale. At protein–DNA ratios where half the DNA is bound, the imino proton of G8 is dramatically broadened and is shifted downfield by ~250–300 Hz. This is consistent with DNA exchanging between free and bound forms on this time scale. Consequently, exchange broadening of Tyr12 may result from the DNA exchanging on and off at this site. Other residues in this region are not affected to the same extent. The intensities of NOE connectivities from Trp15 and Trp43, which pack against Tyr12 (Figure 8), are attenuated. However, the intensities of connectivities from Trp52, located on the opposite face to Tyr12, are not significantly affected, reinforcing the idea that Tyr12 is intimately involved in binding DNA.

The aromatic residue at position 12 of HMG-D is very highly conserved in the equivalent position in nearly all HMG domains. In the solution structure of HMG-D (Jones *et al.*, 1994) Met13 lies parallel to Tyr12 (Figure 8). Therefore, by analogy with SRY, Met13 may interact with the T^6-T^7 step of the DNA, whilst Tyr12 may interact with a neighboring base step, one of which is the conserved T^7-G^8 step. This implies that the N-terminus of helix 1 recognizes a core element, TTG, of the binding site, whereas sequence-specific recognition of the surrounding DNA may be conferred by a distinct region of the domain comprising the N-terminal residues (Berta *et al.*, 1990;

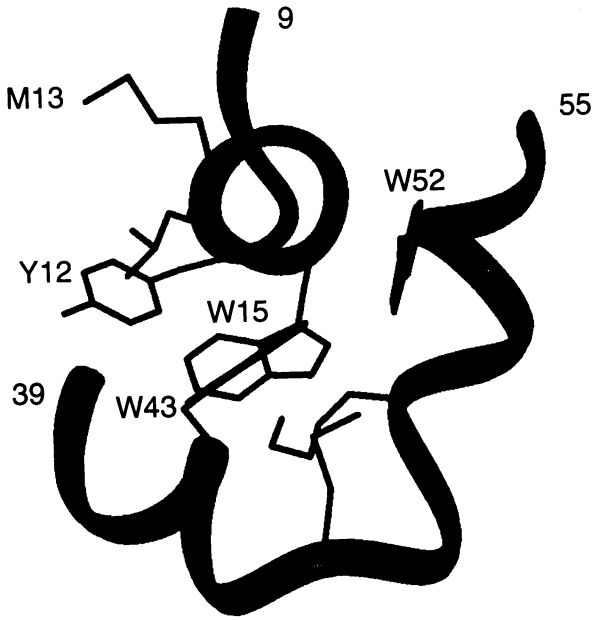


Fig. 8. Schematic representation of the main hydrophobic core of HMG-D showing the relative conformations of the main residues within this region. The diagram is based on the structure with the smallest RMSD from the mean structure as determined by NMR spectroscopy (Jones *et al.*, 1994).

Giese *et al.*, 1991; Nasrin *et al.*, 1991; Read *et al.*, 1994). The conserved aromatic residue at position 12 may interact with the minor groove in different ways: for example, in the complex of DNase I with DNA a tyrosine residue makes van der Waals contacts with a sugar moiety, inducing a widening of the minor groove and a bend of ~25° away from the protein (Suck *et al.*, 1988). In contrast, in the complex of the TATA binding protein (TBP) with DNA, phenylalanine residues partially intercalate between adjacent base pairs, inducing both a bend of ~40° at each site and an associated untwisting (Kim,C.Y. *et al.*, 1993; Kim,J.L. *et al.*, 1993). These three minor groove binding proteins, HMG-D, TBP and DNase I, have completely different DNA binding motifs and distort DNA to differing extents, yet all appear to use aromatic residues and bulky hydrophobic residues to stabilize their interactions with the minor groove of DNA.

The indole resonances of the tryptophan residues in the main hydrophobic core each appear as two signals in slow exchange on the NMR time scale and these signals arise from two distinct conformations of the protein bound to DNA. However, the exact nature and the contribution that each makes to binding is not yet understood. In contrast, the aromatic residues in the secondary hydrophobic core are unaffected by binding DNA. However, mutational analysis has shown that residues in the N-terminal strand are important in binding DNA (Berta *et al.*, 1990; Giese *et al.*, 1991; Nasrin *et al.*, 1991; Read *et al.*, 1995). However, it is possible that the aromatic residues in the secondary core are less sensitive to conformational changes induced on binding DNA than the four closely packed aromatic residues in the main hydrophobic core. Many residues in the basic tail also exhibit conformational changes on binding DNA. These changes are complex and preclude further analysis at this stage. However, they may be associated with either a transition to a more ordered structure or interaction with the DNA. It is known that removal of this tail reduces the affinity for DNA (M.E.A.Churchill, unpublished data) and this reinforces the idea that the tail may make direct contacts with the DNA.

The architecture of HMG-D–DNA complexes

The binding of SRY to DNA (King and Weiss, 1993) induces a bend at the TT base step of the TTG element which is located on the outer face of the DNA. In our studies the (T)TG consensus sequence of the HMG domain proteins is located towards the 3′ end of the binding site. If this sequence has the same rotational orientation in the HMG-D–DNA complex as in the SRY–DNA complex, then the DNase I protection pattern (Figure 5B) and the NMR footprint (Figure 3) imply that in the region of principal DNA contacts the protein tracks from the inside (5′) of the bend to the outside (3′).

The DNase I 'footprints' of HMG-D on the *ftz* SAR DNA reveal two binding sites located 19 bp apart which each contain a single TG step oriented in the same direction. Therefore, the relative helical repeat between the two binding sites is 9.5 bp. A minimum estimate for the intrinsic repeat of the DNA in the complex is 10.5 bp. This difference between the intrinsic and relative repeats implies that successively bound protein molecules could be helically disposed relative to one another. Since the binding of a single HMG domain at a single site is sufficient to induce a bend in the DNA duplex (Ferrari *et al.*, 1992; Giese *et al.*, 1992), we infer that binding of HMG-D to an array of such sites may confer a superhelical path on the duplex axis. In fact, the relative helical repeat for HMG-D is close to the values of 8.5 and 9 bp reported for the *E.coli* protein HU (Broyles and Pettijohn, 1986; Bonnefoy and Rouviere-Yaniv, 1992). This correspondence is fully consistent with the functional equivalence of HU and proteins of the HMG1/2 class (Megraw and Chae, 1993; Paull *et al.*, 1993).

The presence of a DNase I hypersensitive site midway between the two TG sites indicates that the DNA in this region is distorted from its normal structure and suggests the existence of a bend at this position (Suck *et al.*, 1988; Lahm and Suck, 1991). However, in the absence of more direct evidence, we cannot exclude other possibilities. DNase I hypersensitivity is also observed between the binding sites of individual boxes in the xUBF–rRNA–promoter DNA complex (Leblanc *et al.*, 1993). In the xUBF–DNA complex the DNA forms a negative supercoil, which suggests that the observed hypersensitivity is a direct consequence of protein-induced bending.

Previously it has been demonstrated that HMG domain proteins bind preferentially to deformed DNA structures, such as synthetic four-way junctions and *cis*-DDP–DNA adducts. Although HMG-D, in common with similar *Chironomus* proteins (Wisniewski and Schulze, 1994), has no preference for four-way junctions, it does exhibits a 2- to 3-fold preference for binding to *cis*-DDP–DNA adducts over linear DNA. This preference is significantly less than that reported for HMG-1. One difference between our experiments and those of others is that the length of the modified DNA used here is only sufficient to bind a single molecule of HMG-D at the concentrations used, whereas

other workers have used modified oligomers of up to 100 bp (e.g. Pil and Lippard, 1992), a length which would allow multiple, and possibly cooperative, binding of HMG-1. *cis*-DPP adducts induce a kink of 32–34° in the DNA and an associated untwisting of 13° (Bellon and Lippard, 1990; Bellon *et al.*, 1991). However, HMG-D has a substantially higher affinity (K_d = 7 nM) for a synthetic DNA containing an enforced bend of ~30° and underwound by ~30°, where the bend is distributed over three base steps (S.A.Wolfe, A.E.Ferentz, V.Grantcharova, M.E.A.Churchill and G.L.Verdine, submitted). Therefore, for HMG-D the structure of DNA, rather than a specific sequence, is the overriding factor in determining the affinity for a binding site.

We have shown that HMG-D can bind underwound DNA and with a particular arrangement of binding sites may also stabilize a negative writhe. These properties are characteristic of negatively supercoiled DNA. Indeed, HMG-D, in common with other HMG domain proteins and HU, can constrain negative supercoils both *in vivo* and *in vitro* (S.S.Ner, unpublished data). By stabilizing such structures, the protein could facilitate the wrapping of DNA in higher order nucleoprotein complexes (Travers *et al.*, 1994), which would be consistent with the association of HMG-D with condensed chromatin in the early stages of *Drosophila* embryogenesis.

Materials and methods

Protein expression and purification

The gene for full-length HMG-D (112 amino acids) was sub-cloned into pET13a. Use of this vector (Gerchman *et al.*, 1994) was found to improve expression levels. In addition, a construct of the gene encoding the first 100 residues (HMG-D-100), from Met1 to Lys100 inclusive, was cloned between the *Nde*I and *Bam*HI sites of pET13a using methods described previously (Jones *et al.*, 1994).

Expression and purification of HMG-D and HMG-D-100 were essentially as described for HMG-D-74 (Jones *et al.*, 1994), except that for HMG-D-100, a SP Sepharose Fast Flow (Pharmacia) column was used in place of Mono S (Pharmacia). For DNA binding experiments the proteins were stored in buffer of 50 mM HEPES-Na, pH 8.0, 1 mM EDTA, 50 mM NaCl, with added glycerol or bovine serum albumin (BSA), depending on the method of storage required. The over-expressed and purified proteins lack the N-terminal (initiator) methionine, as determined by N-terminal sequence analysis.

Binding site selection

The binding site selection method of Pollock and Treisman was used to obtain DNA sequences that bind preferentially to HMG-D (Pollock and Treisman, 1990). A polyclonal antiserum against the HMG-D protein (a gift from S.Ner) (Ner and Travers, 1994) was used to immunoprecipitate complexes of HMG-D (~22 nM) with DNA fragments containing a 26 bp stretch of randomized DNA sequence flanked by fixed ends for PCR priming and subsequent sub-cloning (Pollock and Treisman, 1990). The immunoprecipitated material was then amplified using the polymerase chain reaction (PCR) and purified by polyacrylamide gel electrophoresis. An excess of poly(dI–dC) (1 μg/reaction) was used as a competitor DNA in the binding reactions to compensate for the A/T sequence preferences of the AK basic tail and to ensure selection of preferred DNA fragments (Churchill and Travers, 1991). This process was repeated three times and we noted that the amount of DNA precipitated during each cycle increased steadily. After four cycles of amplification, the selected fragments were electrophoresed in the presence of HMG-D on a 4% polyacrylamide gel (40:1 acrylamide:bis-acrylamide) for 90 min at 150 V in 0.5× TBE (Tris–borate–EDTA); inserts that retarded the mobility of the DNA were isolated, gel-purified using the crush and soak procedure (Maxam and Gilbert, 1980), subcloned into the *Bam*HI–*Eco*RI site of a Bluescript SK$^-$ vector using standard procedures (Sambrook *et al.*, 1989) and sequenced using standard methods. These sequences were of the form: **CAGGTCAGTTCAGC**-GGATCC*TGT*–CG(N)$_{22-27}$GA–GGCGAATTC**AG***TGCTG***C**AGC. We note that CA and the TG sequences are already present in the insert and are represented by bold or italic lettering respectively.

Competitive electrophoretic mobility shift assays

The probe DNA for band shift/competitive electrophoretic mobility shift assays (EMSA) contains the region 400–455 of the *Drosophila fushi tarazu* (*ftz*) upstream element/scaffold-associated region (USE/SAR) (Harrison and Travers, 1988; Amati *et al.*, 1990). Plasmid p19Sal12, which contains this fragment, was constructed by sub-cloning a 62 bp *Sal*I restriction fragment from pBSSNP12 (a gift of S.Ner; containing a tandem repeat of the 62 bp sequence to which *Sal*I linkers had been added before being subcloned into the *Sal*I site of a Bluescript vector) into the *Sal*I site of pUC19 using standard subcloning protocols. The 62 and 92 bp fragments were excised from the p19Sal12 plasmid using *Sal*I and *Bam*HI/*Hin*dIII digests respectively. These fragments were then labeled at the 5′ end using [γ-^{32}P]ATP (3000 Ci/mmol; Amersham) and T4 polynucleotide kinase. Competition oligonucleotides were synthesized and purified using standard methods. The oligonucleotide duplexes were then annealed at a 1:1 ratio and further purified from single-stranded DNA, unless noted otherwise.

Reactions, for band shift competition assays to determine HMG-D sequence preferences, of 10 μl contained: ~1.5 fmol labeled 92 bp DNA fragment (*Bam*HI/*Hin*dIII cut from p19Sal12), binding buffer (20 mM HEPES, 50 mM KCl, 2 mM $MgCl_2$) with 100 μg/ml BSA, 2 ng poly(dG–dC) and 'sequence' competitor DNA at [200 nM]$_f$. To this reaction was added HMG-D (diluted in 50 mM HEPES, 0.1 M KCl, 1 mM EDTA, 100 μg/ml BSA, 50% glycerol, pH 7.5) to give a series of final concentrations of 50, 100, 150, 200 and 250 nM and incubated with the DNA at 22°C for 30 min. The reactions were loaded onto a 6% native polyacrylamide gel (30:1 acrylamide:bis-acrylamide) and electrophoresed for 1–2 h at 150 V in 0.33× TBE. The gel was dried for 30 min and exposed to X-ray film.

The competition experiments to determine the minimal length of the HMG-D binding site were performed in 5 μl reactions that contained the following: ~1 fmol labeled 62 bp DNA fragment (*Sal*I cut from p19Sal12), because HMG-D affinity for this fragment was lower than the 92 bp probe (100–200 nM; data not shown), 100 μg/ml BSA, 10 ng poly(dG–dC), [1 μM]$_f$ of 'length' competitor DNA in binding buffer. The HMG-D (diluted as above) was added to give 100, 200, 300, 400 and 500 nM final concentrations, which were incubated with the DNA at 22°C for 30 min. The band shift assays and electrophoresis were carried out as described above.

In order to determine the fraction of DNA bound, the gels were exposed to a phosphorimaging screen for between 10 and 24 h and subsequently scanned using the Molecular Dynamics phosphorimaging system. The Molecular Dynamics ImageQuant program was used to read the images and integrate the exposure intensity for each of the bands of interest. These data were then transferred to an Excel (Microsoft) spreadsheet and then to the data analysis and graphing program Kaleidagraph (Abelbeck Software), where the bound and free fractions were calculated and plotted for each competitor DNA at each protein concentration.

Preparation of NMR samples

The single-stranded oligonucleotide components of TG16B were synthesized, purified and quantitated using standard methods. The strands were annealed at a 1:1 molar ratio at 1 mM concentration in 50 mM Tris–HCl, pH 7.9, 1 mM EDTA by heating at 80°C for 5 min, then at 67°C for 30 min, followed by slow cooling to room temperature. The duplex DNA was dialyzed at 4°C against 20 mM $NaPO_4$, pH 5.0, 0.02% NaN_3. NMR samples, with a molar ratio of DNA:protein of 1.0:0.8, were prepared by mixing together 1.0 ml of 0.55 mM TG16B in buffer (20 mM $NaPO_4$, pH 5.0, 0.02% NaN_3) with 1.0 ml of 0.69 mM [^{15}N]HMG-100-D in the same buffer, allowing the mixture to stand for 30 min at room temperature before concentration to 500 μl using a Centricon 3 microconcentrator. D_2O was added to 10% (v/v).

NMR measurements

NMR spectra were recorded on a Bruker AMX500 or a Bruker DMX600 spectrometer equipped with a 5 mm triple resonance (^{1}H/^{15}N/^{13}C) probe. All spectra were acquired in phase sensitive mode and frequency discrimination in the F1 dimension was achieved using time-proportional phase incrementation (Bodenhausen *et al.*, 1980; Marion and Wüthrich, 1983). Chemical shifts are quoted at 20°C unless otherwise stated. ^{1}H chemical shifts are given relative to internal sodium 2,2,3,3-[^{2}H$_4$]-3-trimethylsilylpropionate. Solvent suppression in NOESY spectra recorded

in H_2O was achieved by using a jump-return sequence for the final pulse (Plateau and Guéron, 1982). In all other spectra, suppression of the residual HDO signal was achieved by on resonance phase-coherent presaturation (Zuiderweg *et al.*, 1986).

For DQF-COSY, NOESY and TOCSY spectra of protein in H_2O solutions 2048 (complex) ×512 (real) data points were acquired in the time domain with 64 transients per t_1 increment over a spectral width of 8064 Hz in both dimensions. A spectral width of 5000 Hz in both F1 and F2 was used for spectra acquired using D_2O solutions. Z-Filtered TOCSY spectra (Rance, 1987; Nakaseko *et al.*, 1992) were acquired with mixing times ranging from 42 to 60 ms using the DIPSI-2 mixing sequence (Shaka *et al.*, 1988).

HSQC spectra (Bodenhausen and Ruben, 1980) were acquired with 2048 (complex) ×512 (real) time domain data points and 64 transients per t_1 increment. The spectral widths were 8064 Hz and 2000 Hz in F2 and F1 respectively. ^{15}N decoupling during acquisition was achieved using GARP (Shaka *et al.*, 1985) with a decoupler field strength of $\gamma B_2/2\pi$ = 1.4 kHz.

DNase I protection assay

The DNase I protection pattern of HMG-D on the *ftz* SAR sequence was obtained using the plasmid pBSSNP12 (a gift of S.Ner) which contains a tandem repeat of the *ftz* SAR *SalI* fragment described above. The plasmid was restricted using *Hin*dIII and end-labeled either on the 5′ end (bottom strand) or 3′ end (top strand) using the phosphatase/kinase reaction described above or by end-filling using radioactive [α-^{32}P]dATP (3000 Ci/mmol) and reverse transcriptase (Sambrook *et al.*, 1989). A 157 bp fragment of this DNA was separated from the plasmid by subsequent cleavage using *Ava*I and gel purification from an 8% native polyacrylamide gel (20:1 acrylamide:bis-acrylamide, 1× TBE, 90 min at 150 V).

The footprinting reactions were composed of ~15 fmol labeled 157 bp DNA fragment, binding buffer (20 mM HEPES, 50 mM KCl, 2 mM $MgCl_2$), 100 μg/ml BSA, 10 ng poly(dG−dC), and HMG-D (diluted as above) at either 0, 100 or 500 nM final concentration and were allowed to incubate at 22°C for 15 min. To the reaction vial without HMG-D, 0.4 U DNase I was added and the reaction was quenched after 2 min by adding EDTA to a concentration of 20 mM. The reaction vials containing HMG-D were incubated for 3 min with 0.8 U DNase I before quenching. The reactions were loaded directly onto a 6% native polyacrylamide gel (30:1 acrylamide:bis-acrylamide) and electrophoresed for 2 h at 150 V in 0.33× TBE. The gel was exposed to film and the shifted DNA was excised and purified by electro-elution and concentration, followed by phenol extraction, ether extraction and ethanol precipitation. The DNA cleavage products were redissolved in loading buffer containing 90% formamide, 0.1 M NaOH, 1 mM EDTA, 1% each bromophenol blue (BPB) and xylene cyanol (XC) dyes, heated to 90°C for 1 min and loaded onto a DNA sequencing gel (7% 20:1 acrylamide:bis-acrylamide, denaturing, 8 M urea) and electrophoresed for 2.5 h at 25 W in 1× TBE. The gel was dried for 30 min and exposed to X-ray film.

Autoradiographs were scanned using a custom built densitometer which produced digitized images of the gel (Churchill and Suzuki, 1989). Accurate peak integrals were obtained by analysis of the peak profiles from 1-dimensional scans of the image, using matrix decomposition and simple integration methods implemented in the GELTRAK program (Churchill *et al.*, 1990; Smith and Thomas, 1990). The data were transferred to a spreadsheet where the Kaleidagraph program (Abelbeck) was used to perform the log probability subtraction analysis (Lutter, 1978) and to represent the data graphically (Churchill and Suzuki, 1989).

HMG-D–cis-DDP–DNA binding assays

The single-stranded (ss) DNA, *cis*-DDP-modified DNA, unplatinated and complementary ssDNA were a gift from Dr Rafael Giraldo. The sequence 5′-TCTACACCT**GG**TCACCTCTC-3′ (20 bp) had either been untreated or treated with *cis*-DDP to form an adduct at the guanine positions in boldface (Bruhn *et al.*, 1992). Ten picomoles each of the ss *cis*-DDP–DNA and unplatinated DNA were separately radiolabeled at the 5′ end as described above and purified by phenol extraction, ether extraction and a final SEP-PAK (Waters) purification. The *cis*-DDP–DNA and unplatinated DNA in TE buffer (10 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 8.0) were each annealed with 10 pmol of the unlabeled complementary strand oligonucleotide by heating to 90°C for 1 min and allowing the duplexes to cool slowly (note that there may be an excess of unlabeled ss DNA in these samples, but it will be the same amount for both the unplatinated and the platinated samples).

The 10 μl reactions contained labeled DNA of ~1 nM concentration in binding buffer (20 mM HEPES, 50 mM KCl, 2 mM $MgCl_2$, 100 μg/ml BSA) to which HMG-D (in dilution buffer, as above) was added to give final protein concentrations of 0, 25, 50, 100, 250, 500, 750, 1250, 2000 and 10 000 nM. The samples were incubated at 0°C for 15 min to allow the binding reaction to reach equilibrium and loaded onto an 8% (30:1) polyacrylamide gel and electrophoresed in 33× TBE buffer at 125 V and 4°C for 2 h. The gel was dried and quantitated as described above. The intrinsic affinity of HMG-D for a single site on DNA, the dissociation constant (K_d), can be calculated using the equation: $K_d = ([D_f][P_f])/[DP]$, where $[P_f] = [\text{HMG-D}_{free}]$, $[D_f] = [\text{DNA}_{free}]$, and $[DP] = [(\text{DNA·HMG-D})_{bound}]$. Therefore if $[D_f] = [DP]$ and $[P_{Total}] >> [D_f]$, then $[P_{Total} \approx P_{free}]$ and $K_d \approx [P_{Total}]$. If these conditions are met (which happens at 50% saturation or the mid-point of the saturation binding curve), then the affinity of HMG-D for the DNA is approximated by $[\text{HMG-D}_{Total}]$.

Determination of circularization rate

The 169 bp *tyrT* DNA fragment was excised from the plasmid pΔ98, derived from pΔ98 (Drew and Travers, 1985), using *Eco*RI; it was 5′ end-labeled with [γ-^{32}P]ATP and gel-purified using methods described above. An equal amount, <1 fmol, of radiolabeled *tyrT* DNA was added to each of four reaction vials containing binding buffer (20 mM HEPES, 50 mM KCl, 2 mM $MgCl_2$), 10 mM ATP, 1 mM dithiothreitol, 1 mM EDTA and 100 μg/ml BSA and allowed to equilibrate for 10 min with either water (as a control), HMG-D (75 nM final concentration), HU (50 nM final concentration) or Hoescht 33258 (1 μM final concentration). T4 DNA ligase (10 U/ml) was added to each reaction vial and incubated at 22°C. Aliquots from each reaction that were removed at time intervals of 0, 2, 5, 7, 10, 12, 15 and 20 min were quenched using phenol extraction to halt the ligase reaction instantaneously. The DNA was extracted with ether, ethanol precipitated, dried and redissolved in gel loading buffer (50% glycerol, 1% XC, 1% BPB, 1% binding buffer) and loaded onto a 6% (20:1, 1× TBE buffer) polyacrylamide gel. After electrophoresis at 125 V at room temperature for 2 h, the gel was dried and exposed to X-ray film. Control experiments (Drew and Travers, 1985) were performed to verify the position of the monomer circle on the gel (data not shown). To quantitate the amount of circular and linear DNA in each reaction, the gel was exposed to a phosphorimaging screen and quantitated using the Molecular Dynamics phosphorimaging system and other programs as described above.

Acknowledgements

We thank Dr Sarbjit Ner for his generous gifts of the HMG-D polyclonal antisera and *ftz* SAR plasmid, Dr Rafael Giraldo for providing us with the *cis*-DDP-treated oligonucleotides and Terry Smith and Jan Fogg for preparation of the other oligonucleotides. We are grateful to Drs Sarb Ner and David Neuhaus for many helpful discussions. We also thank Aaron Klug for his support and enthusiasm. M.E.A.C. gratefully acknowledges support from the University of Illinois for H.H. and T.G.

References

Alexander-Bridges,M., Ercolani,L., Kong,X.F. and Nasrin,N. (1992) *J. Cell. Biochem.*, **48**, 129–135.

Amati,B., Pick,L., Laroche,T. and Gasser,S.M. (1990) *EMBO J.*, **9**, 4007–4016.

Bazett-Jones,D.P., Leblanc,B., Herfort,M. and Moss,T. (1994) *Science*, **264**, 1134–1137.

Bellon,S.F. and Lippard,S.J. (1990) *J. Biophys. Chem.*, **35**, 179–188.

Bellon,S.F., Coleman,J.H. and Lippard,S.J. (1991) *Biochemistry*, **30**, 8026–8035.

Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) *Nature*, **348**, 448–450.

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

Bodenhausen,G. and Ruben,D. (1980) *Chem. Phys. Lett.*, **69**, 185–187.

Bodenhausen,G., Vold,R.I. and Vold,R.R. (1980) *J. Magn. Resonance*, **37**, 93–106.

Bonnefoy,E. and Rouviere-Yaniv,J. (1992) *EMBO J.*, **11**, 4489–4496.

Broyles,S.S. and Pettijohn,D.E. (1986) *J. Mol. Biol.*, **187**, 47–60.

Bruhn,S.L., Pil,P.M., Essigmann,J.M., Housman,D.E. and Lippard,S.J. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2307–2311.

Bruhn,S.L., Housman,D.E. and Lippard,S.J. (1993) *Nucleic Acids Res.*, **21**, 1643–1646.

Butler,A.P., Mardian,J.K.W. and Olins,D.E. (1985) *J. Biol. Chem.*, **260**, 10613–10620.
Cheung,S., Ardnt,K. and Lu,P. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3665–3669.
Churchill,M.E.A. and Suzuki,M. (1989) *EMBO J.*, **8**, 4189–4195.
Churchill,M.E.A. and Travers, A.A. (1991) *Trends Biochem. Sci.*, **16**, 92–97.
Churchill,M.E.A., Tullius,T.D., Kallenbach,N.R. and Seeman,N.C. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4653–4656.
Churchill,M.E.A., Tullius,T.D. and Klug,A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5528–5532.
Copenhaver,G.P., Putnam,C.D., Denton,M.L. and Pikaard,C.S. (1994) *Nucleic Acids Res.*, **22**, 2651–2657.
Crothers,D.M. (1993) *Curr. Biol.*, **3**, 675–676.
Dooijes,M., van de Wetering,M., Knippels,L. and Clevers,H. (1993) *J. Biol. Chem.*, **268**, 24813–24817.
Drew,H.R. and Travers,A.A. (1985) *J. Mol. Biol.*, **186**, 773–790.
Drlica,K. and Rouviere-Yaniv,J. (1987) *Microbiol. Rev.*, **51**, 301–319.
Ferrari,S., Harley,V.R., Pontiggia,A., Goodfellow,P.N., Lovell-Badge,R. and Bianchi,M.E. (1992) *EMBO J.*, **11**, 4497–4506.
Folta-Stogniew,E. and Russu,I.M. (1994) *Biochemistry*, **33**, 11016–11024.
Gerchman,S.E., Graziano,V. and Ramakrishnan,V. (1994) *Protein Exp. Purificat.*, **5**, 242–251.
Giese,K., Amsterdam,A. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 2567–2578.
Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 185–195.
Grosschedl,R., Giese,K. and Pagel,J. (1994) *Trends Genet.*, **10**, 94–100.
Haaq,C.M., King,C.-Y., Ukiyama,E., Falsafi,S., Haaq,T.N., Donahoe,P.K. and Weiss,M.A. (1994) *Science*, **266**, 1494–1500.
Harrison,S.D. and Travers,A.A. (1988) *Nucleic Acids Res.*, **16**, 11403–11416.
Haykinson,M.J. and Johnson,R.C. (1993) *EMBO J.*, **12**, 2503–2512.
Jantzen,H.M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature* 244, 830–836.
Jones,D.N.M., Searles,M.A., Shaw,G.L., Churchill,M.E.A., Ner,S.S., Keeler,J., Travers,A.A. and Neuhaus,D. (1994) *Structure*, **2**, 609–627.
Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) *Nature*, **365**, 520–527.
Kim,Y.C., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) *Nature*, **365**, 512–520.
King,C.-Y. and Weiss,M.A. (1993) *Proc. Natl Acad. Sci USA*, **90**, 11990–11994.
Kotlarz,D., Fritsch,A. and Buc,H. (1986) *EMBO J.*, **5**, 799–803.
Lahm,A. and Suck,D. (1991) *J. Mol. Biol.*, **222**, 645–667.
Landsman,D. and Bustin,M. (1991) *Mol. Cell. Biol.*, **11**, 4483–4489.
Landsman,D. and Bustin,M. (1993) *Bioessays* , **15**, 539–546.
Lazarus,L.R. and Travers,A.A. (1993) *EMBO J.*, **12**, 2483–2494.
Leblanc,B., Read,C. and Moss,T. (1993) *EMBO J.*, **12**, 513–525.
Leroy,J.L., Kochoyan,M., Huynh-Dinh,T. and Guéron,M. (1988) *J. Mol. Biol.*, **200**, 223–238.
Lutter,L.C. (1978) *J. Mol. Biol.*, **124**, 391–420.
Marion,D. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Megraw,T.L. and Chae,C.B. (1993) *J. Biol. Chem.*, **268**, 12758–12763.
Nakaseko,Y., Neuhaus,D., Klug,A. and Rhodes,D. (1992) *J. Mol. Biol.*, **228**, 619–636.
Nasrin,N., Buggs,C., Kong,X. F., Carznazza,J., Goebl,M. and Alexander-Bridges,M. (1991) *Nature*, **354**, 317–320.
Ner,S.S. (1992) *Curr. Biol.*, **2**, 208–210.
Ner,S.S. and Travers,A.A. (1994) *EMBO J.*, **13**, 1817–1822.
Ner,S.S., Churchill,M.E.A., Searles,M.A. and Travers,A.A. (1993) *Nucleic Acids Res.*, **21**, 4369–4371.
Ner,S.S., Travers,A.A. and Churchill,M.E.A. (1994) *Trends Biochem. Sci.*, **19**, 185–187.
Paull,T.T., Haykinson,M.J. and Johnson,R.C. (1993) *Genes Dev.*, **7**, 1521–1534.
Pil,P. and Lippard,S.J. (1992) *Science*, **256**, 234–237.
Pil,P.M., Chow,C.S. and Lippard,S.J. (1993) *Proc. Natl Acad. Sci USA*, **90**, 9465–9469.
Plateau,P. and Guéron,M. (1982) *J. Am. Chem. Soc.*, **104**, 7310–7311.
Pollock,R. and Treisman,R. (1990) *Nucelic Acids Res.*, **18**, 6197–6204.
Pontiggia,A., Negri,A., Beltrame,M. and Bianchi,M.E. (1993) *Mol. Microbiol.*, **7**, 343–350.
Rance,M. (1987) *J. Magn. Resonance*, **74**, 557–564.
Read,C.M., Cary,P.D., Crane-Robinson,C., Driscoll,P.C. and Norman, D.G. (1993) *Nucleic Acids Res.*, **21**, 3427–3436.
Read,C.M., Cary,P.D., Preston,N.S., Lnenicek-Allen,M. and Crane-Robinson,C. (1994) *EMBO J.*, **13**, 5639–5646.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) *J. Mol. Biol.*, **191**, 659–675.
Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) *Science*, **253**, 1001–1007.
Shaka,A.J., Barker,P.B. and Freeman,R. (1985) *J. Magn. Resonance*, **64**, 547–553.
Shaka,A.J., Lee,C.J. and Pines,A. (1988) *J. Magn. Resonance*, **77**, 274–293.
Sheflin,L.G. and Spaulding,S.W. (1989) *Biochemistry*, **28**, 5658–5664.
Sheflin,L.G., Fucile,N.W. and Spaulding,S.W. (1993) *Biochemistry*, **32**, 3238–3248.
Smith,J.M. and Thomas,D.J. (1990) *Curr. Adv. BIol. Sci.*, **6**, 93–99.
Staden,R. (1989) *Curr. Adv. BIol. Sci.*, **5**, 293–298.
Stros,M., Stokrova,J. and Thomas,J.O. (1994) *Nucleic Acids Res.*, **22**, 1044–1051.
Suck,D., Lahm,A. and Oefner,C. (1988) *Nature*, **332**, 464–468.
Travers,A.A. (1991) *Curr. Opin. Structural Biol.*, **1**, 1140122.
Travers,A.A. and Klug,A. (1987) *Phil. Trans. R. Soc. Lond. B*, **317**, 537–561.
Travers,A.A., Ner,S.S. and Churchill,M.E.A. (1994) *Cell*, **77**, 167–169.
van de Wetering,M. and Clevers,H.J. (1992) *EMBO J.*, **11**, 3039–3044.
van de Wetering,M., Oosterwegel,M., van Norren,K. and Clevers,H.J. (1993) *EMBO J.*, **12**, 3847–3854.
Vipond,I.B. and Halford,S.E. (1993) *Mol. Microbiol.*, **9**, 225–231.
von Hippel,P.H. (1994) *Science*, **263**, 769–770.
von Hippel,P.H. and Berg,O.G. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 1608–1612.
Wagner,C.R., Hamana,K. and Elgin,S. (1992) *Mol. Cell. Biol.*, **12**, 1915–1923.
Weir,H.M., Kraulis,P.J., Hill,C.S., Raine,A., Laue,E.D. and Thomas,J.O. (1993) *EMBO J.*, **12**, 1311–1319.
Wisniewski,J.R. and Schulze,E. (1994) *J. Biol. Chem.*, **269**, 10713–10719.
Zuiderweg,E.R.P., Hallenga,K. and Olejniczak,E.T. (1986) *J. Magn. Resonance*, **70**, 336–343.

Received on November 11, 1994; revised on December 28, 1994