Modifying the helical structure of DNA by design: recruitment of an architecture-specific protein to an enforced DNA bend

Scot A Wolfe[1], Ann E Ferentz[1,‡], Viara Grantcharova[1], Mair EA Churchill[3*] and Gregory L Verdine[1,2*]

[1]Department of Chemistry and [2]Program for Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02138, USA and [3]Department of Cell and Structural Biology and Biophysics Program, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Background: Proteins can force DNA to adopt distorted helical structures that are rarely if ever observed in naked DNA. The ability to synthesize DNA that contains defined helical aberrations would offer a new avenue for exploring the structural and energetic plasticity of DNA. Here we report a strategy for the enforcement of non-canonical helical structures through disulfide cross-linking; this approach is exemplified by the design and synthesis of an oligonucleotide containing a pronounced bend.

Results: A localized bend was site-specifically introduced into DNA by the formation of a disulfide cross-link between the 5' adenines of a 5'-AATT-3' region in complementary strands of DNA. The DNA bend was characterized by high-resolution NMR structure determination of a cross-linked dodecamer and electrophoretic mobility assays on phased multimers, which together indicate that the cross-linked tetranucleotide induces a helical bend of ~30° and a modest degree of unwinding. The enforced bend was found to stimulate dramatically the binding of an architecture-specific protein, HMG-D, to the DNA. DNase I footprinting analysis revealed that the protein is recruited to the section of DNA that is bent.

Conclusions: The present study reports a novel approach for the investigation of non-canonical DNA structures and their recognition by architecture-specific proteins. The mode of DNA bending induced by disulfide cross-linking resembles that observed in structures of protein–DNA complexes. The results reveal common elements in the DNA-binding mode employed by sequence-specific and architecture-specific HMG proteins.

Chemistry & Biology April 1995, **2**:213–221

Key words: disulfide cross-link, HMG proteins, DNA bending, architecture-specific recognition

Introduction

The structural plasticity of DNA is inextricably linked with its biological function. Structures that deviate significantly from the canonical B-form duplex not only arise as fleeting intermediates in catalytic processes [1,2], but also exist as long-lived species in non-covalent ligand–DNA complexes [3–12]. Despite the importance of non-canonical DNA structures, few have been characterized at high resolution in the absence of proteins. Because gross helical aberrations are generally absent in duplex oligonucleotides, it has proven difficult to produce and study non-canonical DNA in the absence of proteins. One solution to this problem would entail the use of molecular design and chemical synthesis to create oligonucleotides containing enforced helical structures [13,14]. Here we report the use of interstrand disulfide cross-linking [15–17] to alter the helical structure of DNA rationally.

One of the most widely observed modes of helical distortion, DNA bending, is essential for biological processes ranging from chromatin condensation to recombination and transcriptional activation [18–21]. Severely bent DNA has been directly observed in several co-crystal structures of protein–DNA complexes [3–6], but not in structures of DNA alone. Nonetheless, biochemical evidence suggests that some DNA sequences possess an intrinsic bend in solution [22,23]. Slight helical bends, (15° or less) occasionally observed in crystal structures of oligonucleotides, are believed to result primarily from crystal packing forces [24,25]. In the present study we have carried out the design, synthesis, structural elucidation and biochemical characterization of DNA containing an enforced bend.

In general terms, our strategy relies on the ability of a cross-link, introduced into DNA at a specific site, to lock the helix into a distorted structure. Although the distortion induced by cross-linking will destabilize the duplex enthalpically, this effect is counterbalanced by the entropic stabilization gained by interstrand cross-linking [15–17]. Since the cross-link is intended merely to fix two points of the DNA structure at a defined distance, but otherwise to act as a bystander,

*Corresponding author. ‡Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

 ISSN 1074-5521

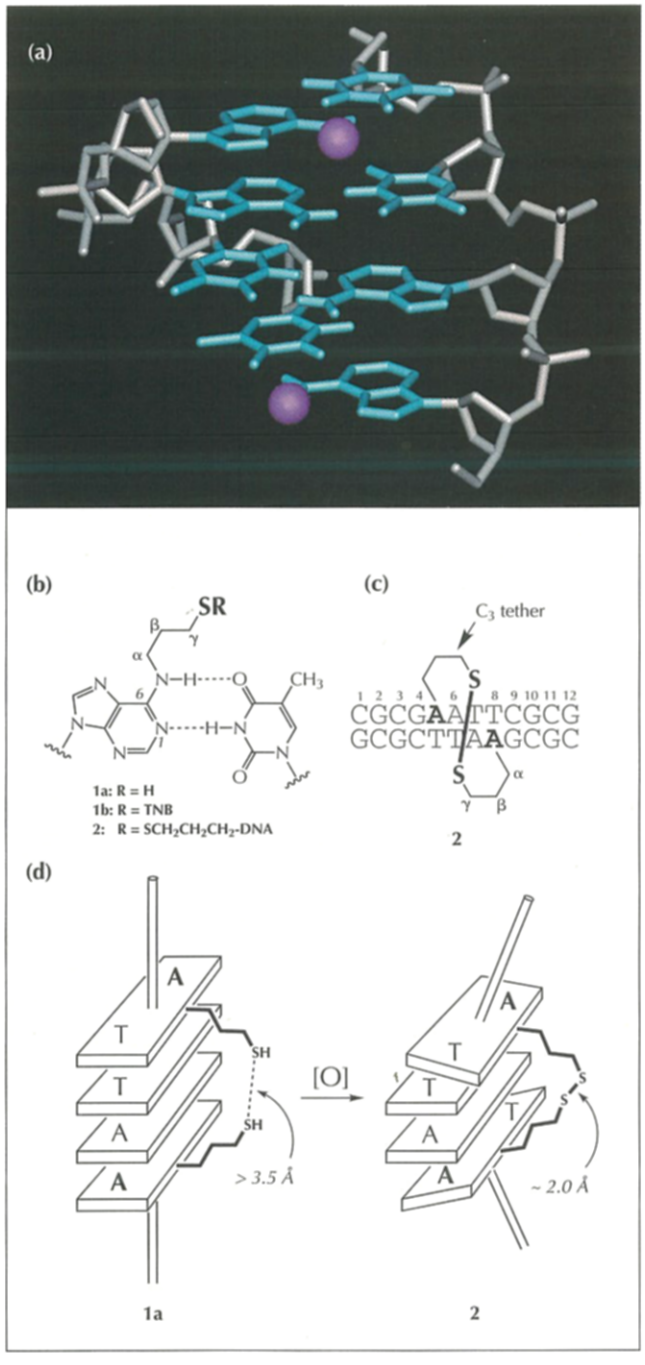

Fig. 1. **(a)** View into the major groove of the 5'-AATT-3' sequence in canonical B-form DNA. The magenta spheres represent the available positions on the exocyclic-N of the outer adenines; gray, phosphodiester backbone; blue, bases. **(b)** Structure and base-pairing of the tethered adenine residue (**A**) used in this study. Activated species **1b** is the mixed disulfide formed between **1a** and 5-thio-2-nitrobenzoic acid (denoted TNB). **(b)** and **(c)** Structure **2** is a disulfide-linked dimer containing two units of monomer **1a**. **(d)** Schematic illustration of the rationale for disulfide-induced bending of DNA. Left: the **A**ATT sequence in canonical B-form DNA, in which the thiol tethers are located too far apart to form a disulfide. Right: formation of a disulfide is envisioned to cause tipping of the modified base-pairs toward each other, thereby inducing a bend along the helix axis (depicted by cylindrical rod). For reasons of graphical simplicity, helical twist is neglected in (d) and the base-pairs are approximated as planes.

we elected to avoid the use of cross-linkers that themselves interact extensively with DNA or cause obligate helical distortion [26–28].

An alternative option was suggested by the finding that a disulfide-linked alkane tether, attached to exocyclic-nitrogens of dA, dC and dG residues in DNA, can act as a tunable inducer of DNA distortion, producing effects ranging from the negligible to the dramatic in a manner that depends on the length of the tether [13–17]. The application of this strategy to induce DNA bending was envisioned as follows: in the 5'-ANNT-3' sequence, the exocyclic N–H bonds of the outer adenines lie roughly atop one another, when viewed along the helix axis (Fig. 1a). Thiol-bearing tethers attached to these positions project directly into the major groove, where they should be capable of joining together to form a disulfide bond (Fig. 1b,c). Molecular modeling suggests, however, that the sulfur atoms of the two tethers are separated by at least ~1.5 Å more than the optimal S–S bond distance (Fig. 1d, structure **1a**). Hence, in order for a disulfide to form, the tethered adenines should have to tip toward each other, thereby inducing a localized bend toward the major groove (Fig. 1d, structure **2**).

Results and discussion

Disulfide cross-linking of a DNA duplex

To test these design principles, we synthesized a dodecamer, 5'-d(CGCG**A**ATTCGCG)-3' [29–32], which self-dimerizes to generate a duplex having a propanethiol-tethered adenine (**A**) on each strand (Fig. 1c). Whereas unstrained dithiol-containing duplexes readily undergo air oxidation to the corresponding disulfide cross-link [13–17], aerobic incubation of structure **1a** yielded only poor conversion (typically less than 10 %) to structure **2**, suggesting that the product of disulfide cross-linking is strained. However, efficient cross-linking could be achieved by mixing structure **1a** with an equimolar amount of the dodecamer activated as a mixed disulfide with 5-thio-2-nitrobenzoic acid (TNB). This activated species is referred to as species **1b**.

The effect of disulfide cross-linking on the stability of duplex DNA was first addressed through thermal denaturation experiments. In earlier work, we showed that attachment of the bis(propanethiol) cross-link to the inner adenines of the same sequence, 5'-d(CGCGA**A**TTCGCG)-3' (hereafter called the **A**T cross-linked dodecamer), increases its half-melting transition temperature (T_m) by 22 °C [13,14]. In contrast, the T_m of structure **2** was only 2 °C higher than that of the native dodecamer, indicating the presence of strain in the **A**ATT cross-linked molecule. The difference in strain of the two cross-linked dodecamers was also apparent in the susceptibility of their disulfide bond toward reduction: the **A**ATT cross-linked dodecamer was reduced by 2-mercaptoethanol roughly 50 times faster than the **A**T cross-linked dodecamer (performed at room temperature; data not shown).

Characterization of the disulfide cross-linked DNA duplex

The integrity of duplex structure in the **A**ATT cross-linked dodecamer was first analyzed by ^{1}H NMR and ^{31}P NMR spectroscopy. The imino region of the ^{1}H NMR spectrum contains five signals (300 °K), indicating that the **A**ATT cross-linked dodecamer is symmetric and fully base-paired. The imino proton signal from the **A**•T base-pairs is broadened, presumably by increased exchange with solvent. The imino proton signal of the central A•T base-pair is shifted upfield by 0.6 ppm in the **A**ATT cross-linked dodecamer relative to that in the unmodified control, but all others were comparable to within 0.1 ppm. ^{31}P NMR spectra of DNA are influenced strongly by even modest changes in structure, and can thus provide a highly sensitive indicator of induced structural distortion [33,34]. The ^{31}P signals of the phosphates in the **A**ATT cross-linked dodecamer are very similar in chemical shift to those of the unmodified duplex, except for the equivalent phosphates in the **A**-p-A steps, which are shifted upfield by ~1 ppm. Taken together, these NMR spectra reveal that the induced deformation in structure **2** is localized to the immediate vicinity of the cross-linked **A**ATT tetranucleotide.

To assay biochemically for the presence of an induced bend in the cross-linked **A**ATT tetranucleotide, we carried out electrophoretic mobility assays on multimerized cross-linked oligonucleotides [35]. The presence of helical deformations such as a DNA bend can be detected by adjusting the 'phasing' between adjacent bent sites with respect to the helical repeat, then comparing the electrophoretic mobility of these phased bends with that of a standard sequence. DNA bends that are oriented in phase — those that have an integral number of helical repeats between the centers of bending — show reduced mobility relative to that of a control oligonucleotide having the same length and sequence; bends that are oriented out of phase — those that have a half-integral number of helical repeats between the centers of bending — have a mobility similar to that of the control. Construction of phased sets of repeats is readily accomplished by homo-concatamerization of a series of duplex oligonucleotide monomers possessing a deformed site flanked by a spacer of variable length. A series of oligonucleotides was designed such that, upon ligation, the centers of their cross-linked **A**ATT sequences would be separated by 10, 12, 14, 16 or 18 base-pairs in the repeating unit. Unmodified oligonucleotides of the same sequence were synthesized and self-ligated to produce standards. The electrophoretic mobilities of the cross-linked and unmodified concatamerized oligonucleotides were then compared.

As shown in Figure 2, the observation of phase-dependent differences in mobility between the control and cross-linked multimers is consistent with the latter being bent. The greatest difference in mobility is observed with the 12 base-pair DNA molecule (12-mer), indicating the bends are in-phase when separated by 12 base-pairs. On the other hand, the least difference is observed with the

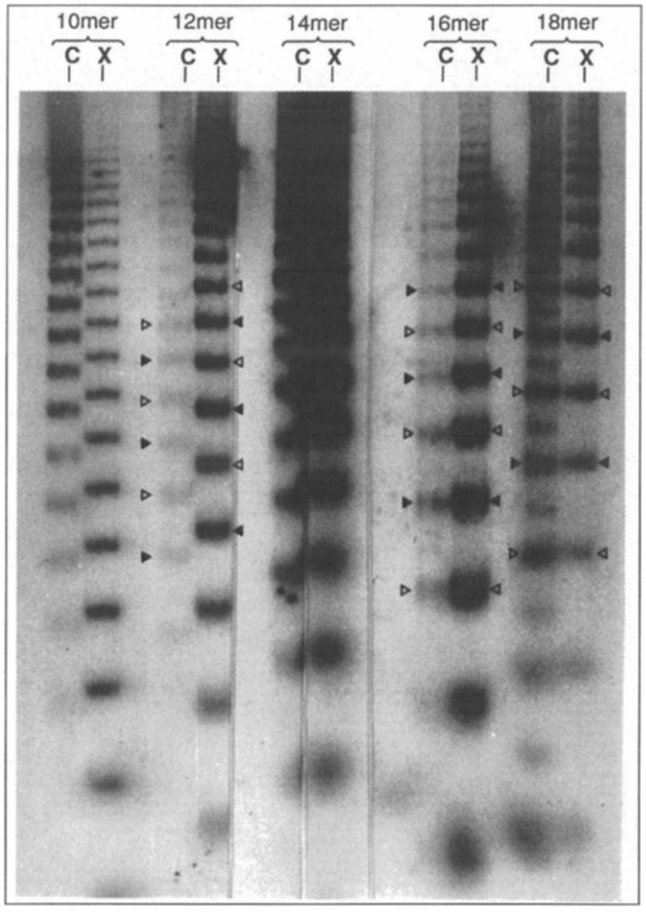

Fig. 2. Non-denaturing polyacrylamide gel electrophoresis (PAGE) of multimers resulting from self-ligation of either **A**ATT cross-linked (X) or unmodified control (C) oligonucleotides. The number above each pair of lanes represents the length of the monomer used: 10-mer, 5'-d(CGCG**A**ATTCG)-3'; 12-mer, 5'-d(GCGCG**A**ATTCGC)-3'; 14-mer, 5'-d(ATCGCG**A**ATTCGCG)-3'; 16-mer, 5'-d(ACGTCGCG**A**ATTCGCG)-3'; and 18-mer, 5'-d(CTAGTCGCG**A**ATTCGCGA)-3'; **A** = disulfide-tethered adenine (X) or native adenine (C). These were ^{32}P end-labeled and ligated enzymatically as described in [35], except that reducing agents were omitted, then the samples were subjected to non-denaturing PAGE. Alternating open and closed triangles denote bands differing in length by one unit. Extra bands in the 18 base-pair lane of the control result from hairpin formation. Decreased electrophoretic mobility is characteristic of DNA sequences that possess an inherent bend. When more than one bent sequence is juxtaposed in-phase with the helical repeat, the net bending is additive, and a further decrease in mobility is observed, (compare X and C in 12-mer lanes). When more than one bent sequence is juxtaposed out-of-phase with the helical repeat, the bends tend to cancel each other out, and no retardation of mobility is observed (compare X and C in 18-mer lanes).

18-mer, indicating that the bends are out-of-phase when separated by 18 base-pairs. This behavior reveals that the helical repeat of the cross-linked oligonucleotide is 12 base-pairs, rather than the ~10 base-pair repeat of canonical B-form DNA.

NMR structural determination

A detailed NMR analysis of the **A**ATT cross-linked dodecamer **2** was carried out to elucidate the structural features of its induced bend. Distance and backbone

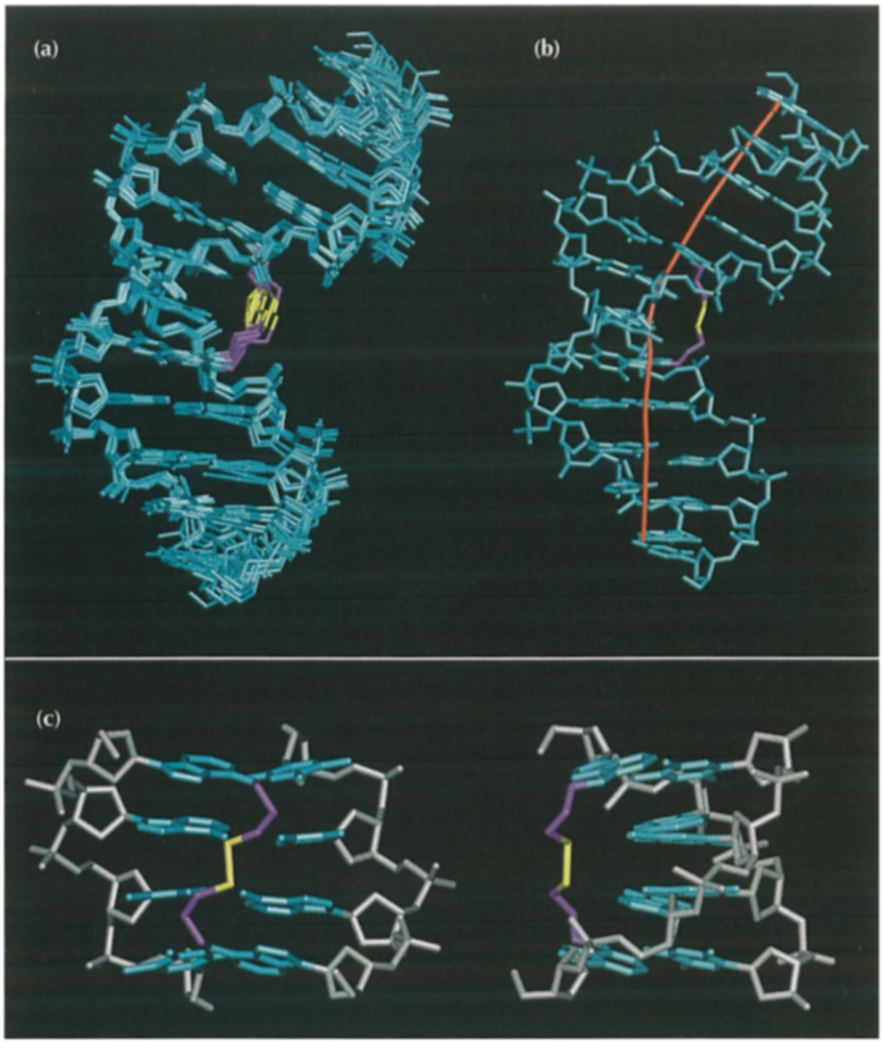

Fig. 3. (a) Superposition of nine refined structures of species **2** generated from a starting A-form helix. All non-hydrogen atoms are shown: blue, DNA; magenta, carbon atoms of tether; yellow, disulfide. Over the central 10 base pairs, the root-mean-square deviation of these structures from the average is 0.5 Å, indicating good convergence; the terminal base-pairs are poorly constrained due to a smaller number of NOE constraints. Of the 10 structures calculated from an A-form DNA starting point, one was discarded because it contained twice the number of significant (>0.5 Å) NOE violations; the remaining nine are shown and were used in the computation of structural statistics. **(b)** Structure from (a) with closest fit to the average. The helical axis, depicted as a red cylinder running vertically through the structure, reveals a bend angle of ~30°. Outside of the central AATT region, the helical parameters are generally B-form in nature. **(c)** Close-up view of the central **A**ATT region from the structure in (b). Left: looking straight into the major groove, the tipping of the cross-linked adenines is clearly apparent. Right: same view with 90° rotation about the helical axis, showing the displacement of the central A•T base-pairs toward the minor groove.

torsional constraints derived from nuclear Overhauser spectroscopy (NOESY) and correlation spectroscopy (COSY) spectra, respectively, were used as a basis set in restrained molecular dynamics calculations. The resulting family of structures (Fig. 3a) agreed well with the nuclear Overhauser effect (NOE) data and with each other (Table 1), except for the terminal base-pairs, which are poorly constrained. The most prominent feature of the refined structures is the localized ~30° bend of the helix toward the major groove (Fig. 3b).

In the structural calculations, the bend is primarily distance-determined, since bent structures are generated even when the disulfide-bearing tether is not included in the refinement protocol. The bend appears to arise primarily from tilting of the tethered **A**•T base-pairs toward each other in the major groove (as judged by a positive roll angle at the central ApT step) and displacement of the intervening A•T base pairs into the minor groove by ~1 Å (Fig. 3c). These structural features are observed directly in the NMR spectra as atypically strong NOEs between the thymine methyl groups and base protons of the 5' neighbors, as well as the atypically weak NOEs between several H-1s and the base protons of their 3' neighbors. Although this displacement may be caused by the compressive force exerted by the cross-link on the central four base pairs (rise per base-pair < 2.9 Å for the central **A**pApTpT region), it is noteworthy that similar displacements have been observed in co-crystal structures of proteins bound to bent DNA [3–6]. In addition to being bent, the dodecamer is also slightly underwound, having a helical twist of 11.7 bases per turn; this value is in close agreement with the results of the electrophoretic

Table 1. Structural parameters for families of DNA structures.

	Average $R^{1/6}$ residual indices for each family of structures		**RMS deviation (Å) of each family to its mean structure**	
Starting helical form	A	B	A	B
Starting structure	0.210	0.104	N/A	N/A
After restrained molecular dynamics	0.093	0.094	0.741	1.053
After volume refinement	0.069	0.071	0.650	1.028

Average $R^{1/6}$ residual index values and root-mean-square (rms) deviations (Å) for each family of structures at various stages of the refinement. Average rms deviations are for each family relative to its average structure.

mobility assays cited above. Outside the region of the cross-linked tetranucleotide, the helical parameters are characteristic of B-form DNA, providing further confirmation that the disulfide-induced distortion is localized to the immediate site of cross-linking.

In all calculated structures, including those generated from the B family, the torsion angle of the disulfide bond deviates significantly from the optimal value of 90° (144 ± 20°), despite the fact that a high penalty for deviation from the optimal angle of 90° was used in the structural calculations. Constraining the disulfide torsion angle to 90° during the calculations doubled the number of substantial (> 0.5 Å) NOE violations in the resulting structures. The presence of a strained disulfide in the AATT cross-linked oligonucleotide is consistent with the difficulty in forming structure **2** through air-oxidation of duplex **1a** (Fig. 1d), and the unusually high susceptibility of the disulfide toward reduction (see above). Thus, both the disulfide and the DNA in species **2** are torsionally strained; indeed, because of their geometric relationship, the DNA helix and disulfide bond are akin to opposed springs: relief of strain in either one increases the strain in the other.

Affinity of HMG-D for DNA containing a disulfide-induced bend

Having produced DNA with a well defined and highly localized helical aberration, we next set out to determine whether this constrained molecule possesses any unique functional characteristics. In particular, we and others have been interested in proteins that recognize particular helical architectures rather than sequence *per se* [36–38]. Although several of these proteins have been shown to bind DNA containing gross helical deformations such as four-way junctions and drug–DNA adducts [39–41], it has been difficult to determine whether they recognize more physiologically representative variants of the canonical B-form duplex. The non-histone high mobility group (HMG) proteins are prototypical architecture-specific proteins [42,43]. HMG-D, an HMG homolog expressed in *Drosophila melanogaster* [44,45], is a particularly challenging example because of its relatively high non-specific affinity for linear, duplex DNA and poor specificity for helical disruptions such as cruciforms and cis-platin adducts (2- to 3-fold increase in affinity [46]). With this in mind, the affinity of HMG-D for DNA containing a disulfide-enforced AATT bend was studied using electrophoretic mobility shift assays (EMSAs).

Because HMG-D binds dodecamers — including species **2** — weakly, we constructed an AATT cross-linked 20-mer and examined its binding to HMG-D (Table 2), in comparison with an unbent, AT cross-linked 20-mer, a 2-(methylthio)ethyl-tethered (uncross-linked) control and an unmodified control. The EMSAs revealed that the presence of the AATT cross-link enhances the affinity of HMG-D by ~15-fold (K_d ~7 nM) relative to that observed for the native 20-mer. This enhancement was due to enforced bending rather than simple attachment of the tether, since the affinity of HMG-D for an unbent 20-mer containing the AT cross-link or an (uncross-linked) methylthioethyl tether was indistinguishable from that of the native control (Table 2). The level of binding enhancement observed here is exceedingly high for this class of proteins; indeed, it is comparable to the enhancement observed for the sequence-specific HMG proteins LEF-1 and SRY binding to a cognate versus non-cognate site [47–49].

Table 2. Equilibrium dissociation constants (K_ds) in nM for HMG-D bound to modified 20-mers as determined by EMSA.

Native	AT Cross-link	AATT Cross-link	MTE
96.5±34.2	63.5±20.3	6.9±1.4	62.4±8.1

All four 20-mer oligonucleotides contained the sequence 5'-AGTTACTG**AATT**ACGCTCAT-3', but differed in the nature of the central **AATT** sequence. The AT and AATT cross-links are identical to those described above for the dodecamers. The (2-methylthio)ethyl tether (MTE) is linked to the outer adenines of the AATT region, acting as a control for steric effects of the tethers. Native denotes unmodified oligonucleotide.

DNase I footprinting of HMG-D bound to DNA containing a disulfide-induced bend

Previous attempts to characterize the binding mode of architecture-specific HMG proteins to duplex DNA have been thwarted by the difficulty of detecting specifically bound complexes through DNase I footprinting. Since this problem is likely to arise from association of the HMG protein with multiple sites on the DNA, one solution could be to introduce a single pre-bent site into the DNA, which might serve to focus the binding of HMG-D. Indeed, the DNase I footprint of HMG-D bound to an AATT cross-linked 44-mer exhibits a clear region of protection (Fig. 4a, brackets). Quantitative subtraction analysis [50] of the protection pattern reveals that the footprint encompasses ~15 base-pairs, centered over the cross-linked tetranucleotide. These data are fully consistent with specific targeting of HMG-D to the locus of disulfide-induced bending.

Additional aspects of the protein–DNA interaction can be inferred from the footprint. The protection patterns of the two strands, while similar, are offset by ~2–3 base-pairs (Fig. 4b). When displayed on a cylindrical projection map (Fig. 4c), it becomes apparent that the strongest contiguous region of protection wraps around the minor groove on the 'back' side of the cross-linked site. Taken together, these data suggest that HMG-D binds primarily over the minor groove, bending the DNA toward the major groove. Additional strong protection is observed over both grooves on the cross-linked face (front side) of the helix. This protection might result from protein contacts, or from steric inaccessibility of the DNase I to the crowded inside of the bent helix; the latter scenario would predict that HMG-D induces an increase in the bending angle of the cross-linked oligonucleotide.

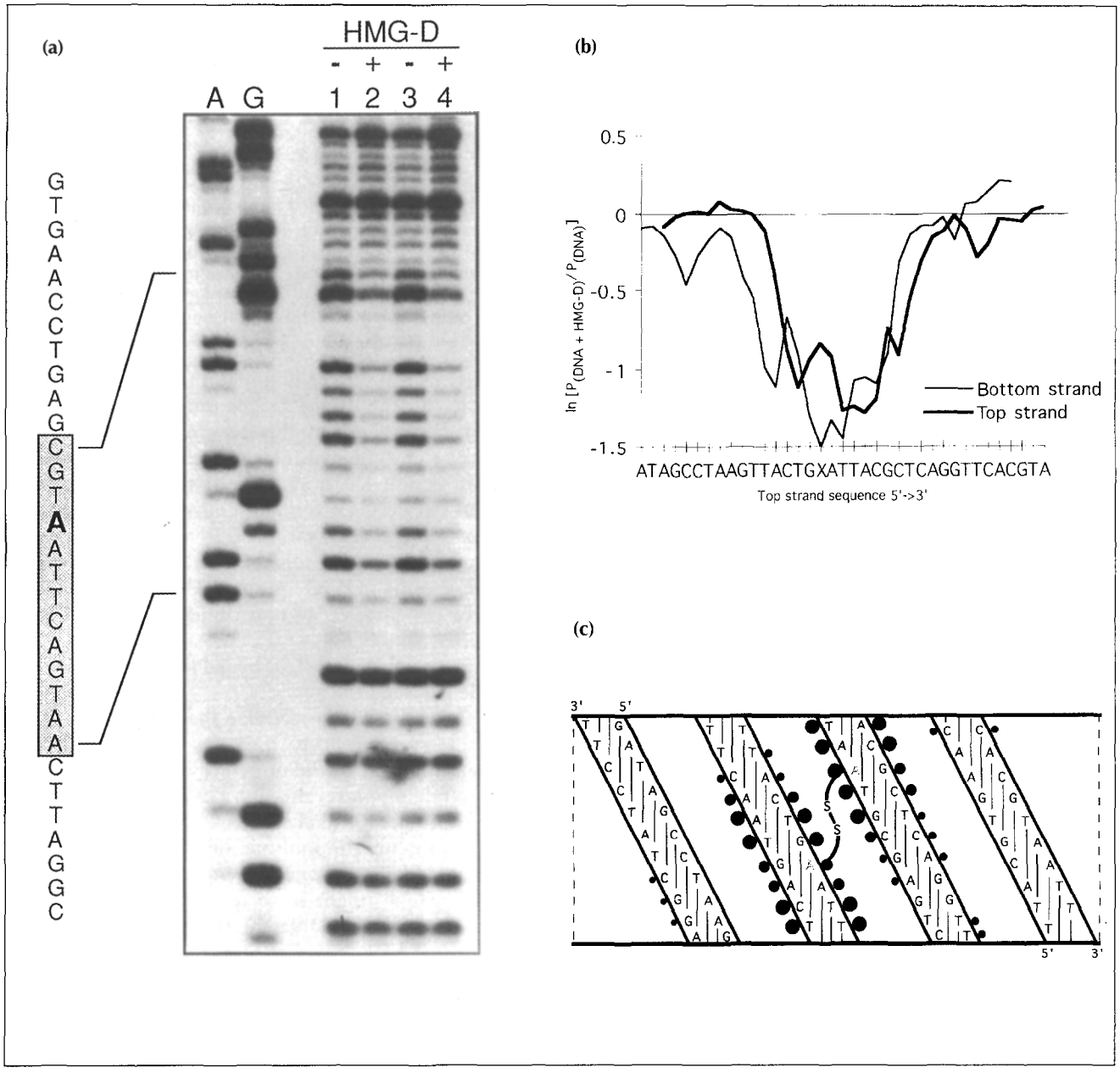

Fig. 4. DNase I protection assay. **(a)** DNase I footprint for one strand (bottom strand) of the crosslinked AATT sequence embedded in a duplex 44-mer, in the absence (lanes 1 & 3) or presence (lanes 2 & 4) of HMG-D. The shaded DNA sequence to the left of the autoradiogram illustrates the region of strongest protection by HMG-D. The modified adenine is shown in bold. Due to the high degree of cooperativity in the binding of HMG-D to DNA, a common property of all HMG proteins, it was necessary to use less than saturating concentrations of HMG-D in the footprints to avoid the formation of higher order complexes. Under these conditions, not all available binding sites are bound, resulting in the appearance of some background cleavage in the protected regions. **(b)** Results of quantitative subtraction analysis [50] of the DNase I footprints of HMG-D on both strands of the crosslinked 44-mer are summarized in a difference logarithm probability plot of two strands plotted together on the same scale; the plots of the top and bottom strands are the average of three and two experiments, respectively. The modified adenine is represented by an X in the DNA sequence. **(c)** Regions of DNA protected from DNase I by HMG-D are illustrated on a cylindrical projection of a DNA helix. The relative extent of protection at each position is denoted by the position and size of black dots. The modified adenine is represented by the outlined **A** in the DNA sequence.

The preferential interaction of HMG-D with DNA bent toward the major groove and the pattern of protection from DNase I suggest a similar DNA-binding mode to that deduced for the sequence-specific HMG protein SRY [51,52]. These results suggest that certain residues that are conserved among members of the HMG protein family serve to establish a common scaffolding for the recognition of altered helical architectures; superimposed on this scaffolding are, in some cases, amino acid residues that confer sequence-specificity through the establishment of base-specific contacts [51,52].

Significance

A persistent problem in the study of architecture-specific HMG proteins has been the difficulty of

recruiting the protein to a single site in DNA. Progress on this front has been made through the discovery that certain HMG proteins will bind DNA structures containing gross helical deformations, such as cis-platin adducts or cruciforms. The goal of obtaining stable complexes between architecture-specific HMG proteins and non-canonical structures of DNA that approximate those ordinarily formed *in vivo* has remained elusive, however. To address this problem, we have pursued a rational approach in which synthetic chemistry is used to enforce a bend in DNA, without disrupting the continuity of the helix. This bent locus binds the single-domain HMG protein, HMG-D, with high affinity and exquisite selectivity. By contrast, HMG-D shows relatively poor selectivity for cis-platin adducts and none for cruciforms. This preference is consistent with the notion that the cross-linked DNA more closely approximates the physiological substrate for HMG-D and perhaps most other architecture-specific HMG proteins, namely smoothly bent sites in the genome.

The ability to recruit HMG-D to a specific site in DNA now makes it possible to undertake studies aimed at characterizing this unusual and poorly understood mode of DNA recognition. Along these lines, we have reported evidence here that the overall disposition of the HMG protein with respect to its recognition site is similar in sequence-specific and structure-specific members of this protein family. These studies bring the ultimate goal of elucidating architecture-specific DNA recognition at high resolution one step closer to fruition.

Materials and methods

Oligonucleotide cross-linking

The propanethiol-modified dodecamer was synthesized by the convertible nucleoside approach as described for a closely related sequence [16]. The activated strand **1b** was prepared by incubation of **1a** with a 10-fold molar excess of 5-5'-dithiobis(2-nitrobenzoic acid) (DTNB). Disulfide cross-linked duplex **2** was formed by incubation of equimolar amounts of **1a** and **1b** (0.1 mM DNA, 10 mM Tris pH 9.0, 1 mM EDTA) for ~12 h. Structure **2** was purified by denaturing polyacrylamide gel electrophoresis. The presence of the disulfide cross-linked dimer of N^6-(3-thiopropyl)-dA was confirmed by HPLC and mass-spectral analysis of enzymatic digestion products [16]. Melting temperatures of the duplexes were obtained as described previously [13].

NMR spectroscopy and structure determination

NMR experiments were recorded on a Bruker AM-500. Measurements were made on a ~1.25 mM sample, buffered to pH 6.5 (10 mM $Na_xH_yPO_4$, 100 mM NaCl, 0.1 mM EDTA) at 300 °K. Sequential ^1H assignments for the **A**ATT cross-linked dodecamer were made by standard methods from ^1H-^1H double quantum filtered correlation spectroscopy (DQF-COSY) and ^1H-^1H NOESY experiments. The backbone ^{31}P NMR resonances were assigned using ^{31}P-^1H total correlation spectroscopy [53,54] (hetero-TOCSY) based upon the previously assigned proton signals. The C_2 symmetry of this molecule causes equivalent positions to exhibit identical chemical shifts in the NMR spectra. H_2O spectra, recorded in 90:10 H_2O:D_2O, observed only five of the six nondegenerate imino protons, due to fraying of the ends of the helix at the temperature at which these spectra were acquired.

These NMR experiments were used to generate 278 experimental distance restraints, 32 hydrogen bonding constraints and 168 broad torsion angle constraints for use in restrained molecular dynamics calculations (X-PLOR 3.0). The energy of the disulfide at a torsion angle of 180° was specified to be 8 kcal mol^{-1} higher than at the optimum torsion angle of 90°. This barrier is ~2 kcal mol^{-1} higher than the currently accepted value [55]. Two families of ten structures each were generated starting from either canonical A- or B-form DNA. During the refinement, constraints were slowly incorporated over 22 psec at 400 K. After cooling to 300 K, molecular dynamics was carried out for 10 psec and the coordinates from the last 4 psec were averaged and energy-minimized. The resulting structures were then refined directly against the volumes of the 50 and 125 msec NOESY cross-peaks (274 and 276 volumes respectively) to better account for the effects of spin diffusion. Distance constraints were reduced to zero as the refinement potential was increased at 400 K. After cooling to 300 K, the coordinates from the final 1.5 psec of dynamics were averaged and energy minimized to give a final structure. The resulting structures of each family demonstrated a good fit to the NOE data, as judged by $R^{1/6}$ values (Table 1), as well as a good fit to each other (Table 1). Because of the lower $R^{1/6}$ and root mean square deviation (RMSD) value for the A family of structures, it was used to analyze the helical parameters of the duplex. The program Dials and Windows [56] was used for analysis of all helical parameters described in this text. One of the ten structures from the A family was discarded due to excessive NOE violations. Structural calculations omitting the tether were carried out using the same set of constraints with the exception of those involving the tether itself.

EMSA assays with HMG-D

The equilibrium dissociation constants (K_ds) for HMG-D bound to four modified 20-mers were determined by EMSA. EMSA assays were done by incubating ~2.5 fmol of ^{32}P labeled DNA with increasing concentrations of HMG-D in a 20 ml reaction containing 150 mM NaCl, 10 mM $MgCl_2$, 10 mM Hepes pH 7.5, 1 mM EDTA, 0.05 % v/v NP-40, 0.2 mg ml^{-1} BSA, and 4 % glycerol, at 4 °C for 30 min. These reactions were then electrophoresed on a 7.5 % native polyacrylamide gel run in 0.5 x TBE at 4 °C. After drying the gel, the bound vs unbound species were assayed using a FUJI BAS-2000 phosphorimaging system. K_ds were calculated by a least-squares method using Kaleidagraph 2.1.3.

DNase I footprinting of HMG-D bound to cross-linked DNA

The cross-linked 44-mer (**X**=propanethiol modified adenine):

```
5'-AAGGATAGCCTAAGTTACTGXATTACGCTCAGGTTCACGT-3'
   3'-TATCGGATTCAATGACTTAXTGCGAGTCCAAGTGCATTAA-5'
```

was 3'-^{32}P end-labeled by filling in the 5' overhangs using Klenow polymerase. Protection assays were done by incubating 5 fmol of labeled DNA in a 10 ml reaction containing 150 mM NaCl, 10 mM $MgCl_2$, 10 mM Hepes pH 7.5, 1 mM EDTA, 0.05 % v/v NP-40, 0.2 mg/ml BSA, at 4 °C, either

with or without HMG-D (final concentration of 20–80 nM) for 30' before addition of DNase I (1 ml of a 40 mg ml^{-1} solution for ~3 min). Reactions were quenched with EDTA, reduced with DTT, and ethanol precipitated before redissolving in 80 % formamide and loading on a 20 % denaturing polyacrylamide gel. A and G cleavage reactions are shown to the left of the DNase I patterns.

Acknowledgements: We thank Mallika Sastry and Dinshaw J Patel for their generous assistance with the NMR structure determination; Sandra Searles and Heidi Hefner for technical assistance in production of HMG-D clones and purification of HMG-D protein; Thomas Keating for experimental assistance; and Larry McLaughlin for use of the T_m instrument. This work was supported by grants from the National Institutes of Health, the National Science Foundation (Presidential Young Investigators Program), and the Chicago Community Trust (Searle Scholars Program) (to G.L.V.). S.A.W. was partially supported by a National Science Foundation predoctoral fellowship. The Harvard NMR facility was supported by NIH (Grant # 1-S10—RR04870-01) and NSF (Grant # CHE88-14019).

References

1. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). Hha I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.
2. Verdine, G.L. (1994). The flip side of DNA methylation. *Cell* **76**, 197–200.
3. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.
4. Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing minor groove of a TATA element. *Nature* **365**, 520–527.
5. Winkler, F., *et al.*, & Wilson, K. (1993). The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781–1795.
6. Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science* **253**, 1001–1007.
7. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907.
8. Pavletich, N.P. & Pabo, C.O. (1993). Crystal structure of a five-finger GLI–DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701–1707.
9. Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science* **265**, 346–355.
10. Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2:1 distamycin A•d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723–5727.
11. Zhang, X. & Patel, D.J. (1990). Solution structure of the nogalamycin–DNA complex. *Biochemistry* **29**, 9451–9466.
12. Ughetto, G., Wang, A.H.-J., Quigley, G.J., van der Marel, G.A., van Boom, J.H. & Rich, A. (1985). A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res.* **13**, 2305–2323.
13. Wolfe, S.A. & Verdine, G.L. (1993). Ratcheting torsional stress in duplex DNA. *J. Am. Chem. Soc.* **115**, 12585–12586.
14. Erlanson, D.A., Chen, L. & Verdine, G.L. (1993). DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc.* **115**, 12583–12584.
15. Ferentz, A.E. & Verdine, G.L. (1991). Disulfide cross-linked oligonucleotides. *J. Am. Chem. Soc.* **113**, 4000–4002.
16. Ferentz, A.E., Keating, T.A. & Verdine, G.L. (1993). Synthesis and characterization of disulfide cross-linked oligonucleotides. *J. Am. Chem. Soc.* **115**, 9006–9014.
17. Ferentz, A.E. & Verdine, G.L. (1994). The convertible nucleoside approach: structural engineering of nucleic acids by disulfide cross-linking. In: *Nucleic Acids and Molecular Biology*. (Eckstein, F. & Lilley, D.M.J. eds), New York, Springer-Verlag.
18. Wang, H., Zuiderweg, E.R.P. & Glick, G.D. (1995). Solution structure of a disulfide cross-linked DNA hairpin. *J. Am. Chem. Soc.* **117**, 2981–2991.
19. Travers, A.A. (1989). DNA conformation and protein binding. *Annu. Rev. Biochem.* **58**, 427–452.
20. Travers, A.A. (1991). DNA bending and kinking — sequence dependence and function . *Curr. Opin. Struct. Biol.* **1**, 114–122.
21. Crothers, D.M. (1993). DNA-bending proteins: Architectural elements in nucleoprotein complexes. *Curr. Biol.* **3**, 675–676.
22. Crothers, D.M., Haran, T.E. & Nadeau, J.G. (1990). Intrinsically bent DNA. *J. Biol. Chem.* **265**, 7093–7096.
23. Paolella, D.N., Palmer, C.R. & Schepartz, A. (1994). DNA targets for certain bZip proteins distinguished by an intrinsic bend. *Science* **264**, 1130–1133.
24. Goodsell, D.S., Kopka, M.L., Cascio, D. & Dickerson, R.E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. USA* **90**, 2930–2934.
25. DiGabriele, A.D. & Steitz, T.A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024–1039.
26. Bellon, S.F. & Lippard, S.J. (1990). Bending studies of DNA site-specifically modified by cisplatin, *trans*-diamminedichloroplatinum(II) and *cis*-$[Pt(NH_3)_2(N3\text{-cytosine})Cl]^+$. *Biophys. Chem.* **35**, 179–188.
27. Tomic, M.T., Wemmer, D.E. & Kim, S.-H. (1987). Structure of a psoralen cross-linked DNA in solution by nuclear magnetic resonance. *Science* **238**, 1722–1725.
28. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science* **253**, 784–786.
29. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S. & Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of DNA. *Nature* **287**, 755–758.
30. Nerdal, W., Hare, D.R. & Reid, B.R. (1989). Solution structure of the *Eco*RI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations. *Biochemistry* **28**, 10008–10021.
31. Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E. & Sussman, J.L. (1988). The three-dimensional structure of a DNA duplex containing looped-out bases. *Nature* **334**, 82–84.
32. Kalnik, M.W., Li, B.F.L., Swann, P.F. & Patel, D.J. (1989). O^6-Ethylguanine carcinogenic lesions in DNA: an NMR study of O^6etG•C pairing in dodecanucleotide duplexes. *Biochemistry* **28**, 6182–6192.
33. Gorenstein, D.G., Schroeder, S.A., Fu, J.M., Metz, J.T., Roongta, V. & Jones, C.R. (1988). Assignments of ^{31}P NMR resonances in oligodeoxyribonucleotides: origin of sequence-specific variations in the deoxyribose phosphate backbone conformation and the ^{31}P chemical shifts of double-helical nucleic acids. *Biochemistry* **27**, 7223–7237.
34. Roongta, V.A., Jones, C.R. & Gorenstein, D.G. (1990). Effect of distortions in the deoxyribose phosphate backbone conformation of duplex oligodeoxyribonucleotide dodecamers containing GT, GG, GA, AC, and GU base-pair mismatches on ^{31}P NMR spectra. *Biochemistry* **29**, 5245–5258.
35. Koo, H.-S., Wu, H.-M. & Crothers, D.M. (1986). DNA bending at adenine•thymine tracts. *Nature* **320**, 501–506.
36. Travers, A.A., Ner, S.S. & Churchill, M.E.A. (1994). DNA chaperones: a solution to a persistence problem. *Cell* **77**, 167–169.
37. Ner, S.S., Travers, A.A. & Churchill, M.E.A. (1994). Harnessing the writhe: a role for DNA chaperones in nucleoprotein-complex formation. *Trends Biochem. Sci.* **19**, 185–187.
38. Churchill, M.E.A. & Travers, A.A. (1991). Protein motifs that recognize structural features of DNA. *Trends Biochem. Sci.* **16**, 92–97.
39. Pil, P.M. & Lippard, S.J. (1992). Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin. *Science* **256**, 234–237.
40. Brown, S.J., Kellett, P.J. & Lippard, S.J. (1993). Ixr1, a yeast protein that binds to platinated DNA and confers sensitivity to cisplatin. *Science* **261**, 603–605.
41. Bianchi, M.E., Beltrame, M. & Paonessa, G. (1989). Specific recognition of cruciform DNA by nuclear protein HMG1. *Science* **243**, 1056–1059.
42. Landsman, D. & Bustin, M. (1993). A signature for the HMG-1 box DNA-binding proteins. *Bioessays* **15**, 539–546.
43. Grosschedl, R., Giese, K. & Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet.* **10**, 94–100.
44. Ner, S.S. & Travers, A.A. (1994). HMG-D, the *Drosophila melanogaster* homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. *EMBO J.* **13**, 1817–1822.

45. Wagner, C.R., Hamana, K. & Elgin, S.C.R. (1992). A high-mobility-group protein and its cDNAs from *Drosophila melanogaster*. *Mol. Cell. Biol.* **12**, 1915–1923.
46. Churchill, M.E.A., Jones, D.N.M., Glaser, T., Hefner, H., Searles, M.A. & Travers, A.A. (1995). HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J.* **14**, 1264–1275.
47. Giese, K., Cox, J. & Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69**, 185–195.
48. Giese, K., Amsterdam, A. & Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5**, 2567–2578.
49. Harley, V.R., *et al.*, & Goodfellow, P.N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science* **255**, 453–456.
50. Churchill, M.E.A., Tullius, T.D. & Klug, A. (1990). Mode of interaction of the zinc finger protein TFIIIA with a 5S RNA gene of *Xenopus*. *Proc. Natl. Acad. Sci. USA* **87**, 5528–5532.
51. King, C.-Y. & Weiss, M.A. (1993). The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. USA* **90**, 11990–11994.
52. van de Wetering, M. & Clevers, H. (1992). Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson–Crick double helix. *EMBO J.* **11**, 3039–3044.
53. Kellogg, G.W. (1992). Proton-detected hetero-TOCSY experiments with application to nucleic acids. *J. Magn. Reson.* **98**, 176–182.
54. Kellogg, G.W., Szewczak, A.A. & Moore, P.B. (1992). Two-dimensional hetero-TOCSY-NOESY. Correlation of ^{31}P resonances with anomeric and aromatic ^{1}H resonances in RNA. *J. Am. Chem. Soc.* **114**, 2727–2728.
55. Jiao, D., Barfield, M., Combariza, J.E. & Hruby, V.J. (1992). *Ab initio* molecular orbital studies of the rotational barriers and the ^{33}S and ^{13}C chemical shieldings for dimethyl disulfide. *J. Am. Chem. Soc.* **114**, 3639–3643.
56. Ravishanker, G., Swiminathan, S., Beveridge, D.L., Lavery, R. & Sklenar, H. (1989). Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: 'curves,' dials and windows. *J. Biomol. Struct. Dyn.* **6**, 669–699.

Received: **17 Mar 1995**; revisions requested: **1 Apr 1995**;
revisions received: **4 Apr 1995**. Accepted: **4 Apr 1995**.