5244–5252 *Nucleic Acids Research, 2002, Vol. 30, No. 23*

A Hoogsteen base pair embedded in undistorted B-DNA

Jun Aishima[1], Rossitza K. Gitti[2,3], Joyce E. Noah[4], Hin Hark Gan[2,4,5], Tamar Schlick[2,4,5] and Cynthia Wolberger[1,2,*]

[1]Department of Biophysics and Biophysical Chemistry and [2]Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205-2185, USA, [3]Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21250, USA, [4]Department of Chemistry and [5]Courant Institute of Mathematical Sciences, New York University, 31 Washington Place, Room 1021 Main, New York, NY 10003, USA

Received July 11, 2002; Revised and Accepted October 3, 2002 PDB no. 1K61

ABSTRACT

Hoogsteen base pairs within duplex DNA typically are only observed in regions containing significant distortion or near sites of drug intercalation. We report here the observation of a Hoogsteen base pair embedded within undistorted, unmodified B-DNA. The Hoogsteen base pair, consisting of a *syn* adenine base paired with an *anti* thymine base, is found in the 2.1 Å resolution structure of the MATα2 homeodomain bound to DNA in a region where a specifically and a non-specifically bound homeodomain contact overlapping sites. NMR studies of the free DNA show no evidence of Hoogsteen base pair formation, suggesting that protein binding favors the transition from a Watson–Crick to a Hoogsteen base pair. Molecular dynamics simulations of the homeodomain–DNA complex support a role for the non-specifically bound protein in favoring Hoogsteen base pair formation. The presence of a Hoogsteen base pair in the crystal structure of a protein–DNA complex raises the possibility that Hoogsteen base pairs could occur within duplex DNA and play a hitherto unrecognized role in transcription, replication and other cellular processes.

INTRODUCTION

Most duplex DNA, whether in A-, B- or Z-form, is composed of complementary strands that associate solely through Watson–Crick base pairing. In a small number of DNA structures containing intercalating drugs (1,2) or pronounced protein-induced distortion (3), Hoogsteen base pairs have been found embedded in duplex DNA. The Hoogsteen base pair geometry, which was first observed in crystal structures of monomeric adenine and thymine base derivatives (4), is characterized by hydrogen bonds between the side of the purine base that faces the major groove and the Watson–Crick base pairing face of the pyrimidine. In B-DNA, formation of a Hoogsteen base pair would require rotation of the purine base about the glycosidic χ bond from the *anti* to the *syn* conformation and, in the case of a guanine-cytosine base pair, protonation of the N3 of cytosine. Hydrogen bonds in the Hoogsteen base pair are formed between the purine N7 to the pyrimidine N3 and either the adenine N6 to the thymine O6 or the guanine O4 to the cytidine N4 (4).

Hoogsteen base pairs have been observed in several crystal structures of distorted double-stranded B-DNA. DNA complexed with intercalating binding drugs, such as triostin-A, contain Hoogsteen base pairs in regions of underwound B-DNA (1). Hoogsteen base pairs can also occur in regions of DNA that are highly distorted by bound protein. In the structure of TATA-binding protein (TBP) bound to DNA (3), a Hoogsteen base pair is observed in the region of DNA underwinding and intercalation by a phenylalanine side chain from TBP. Finally, Hoogsteen base pairs have also been observed at free ends of end-to-end stacked oligonucleotides, such as in the structure of integration host factor bound to DNA (5). The presence of distortions or flexible end regions in B-form DNA may lower the energy barrier for rotation of the purine base about the glycosidic χ bond by changing the base stacking arrangements or the helical properties of the DNA. The lower energy barrier may allow the purine base to rotate from the normal *anti* conformation to the *syn* conformation, leading to the formation of hydrogen bonds characteristic of a Hoogsteen base pair.

We report here the observation of a Hoogsteen base pair embedded within undistorted B-DNA. The Hoogsteen base pair, formed by a *syn* adenine base and an *anti* thymine base, occurs within a 2.1 Å resolution crystal structure containing four MATα2 homeodomains bound to DNA. The 21 bp DNA used for the crystal structure contains two binding sites for the MATα2 homeodomain and was previously used in a crystal structure of the MATα2 homeodomain–DNA complex (6). In

*To whom correspondence should be addressed at 7 Johns Hopkins University School of Medicine, Baltimore, MD 21205-2185, USA. Tel: +1 410 955 0728; Fax: +1 410 614 8648; Email: cwolberg@jhmi.edu
Present address:
Jun Aishima, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 4-230, Berkeley, CA 94720, USA

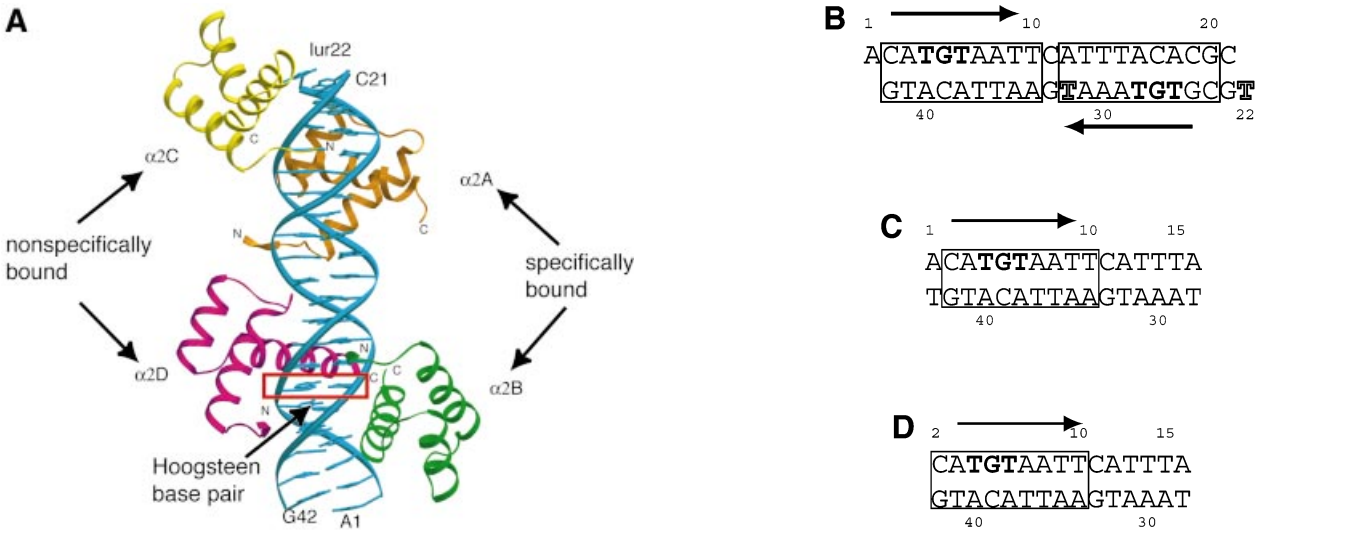

Figure 1. (**A**) Crystal structure of the α2 homeodomain contains four α2 proteins bound to two α2 binding sites in the DNA. Base pair A7-T37 (red box) is a Hoogsteen base pair. Figure produced with SETOR (37). (**B**) The oligonucleotide duplex used in the crystal structure contains two α2 binding sites (box) with a 5′ overhanging base at each end. (**C**) The DNA fragment used in the NMR experiments is blunt-ended and contains five fewer base pairs. (**D**) The DNA fragment used in the molecular dynamics simulations.

the current structure, two of the MATα2 homeodomains bind to the MATα2 binding sites, while the other two MATα2 proteins are bound non-specifically to the DNA (7).

Simulated annealing omit maps (8) clearly show the unambiguous presence of the Hoogsteen base pair in the crystal structure. The Hoogsteen base pair appears to be stabilized by an extra inter-base pair hydrogen bond and base stacking within the DNA, as well as by contacts made by one of the non-specifically bound MATα2 homeodomains with the sugar–phosphate backbone of the *syn* adenine base. None of the observed stabilizing contacts are out of the ordinary, suggesting that Hoogsteen base pairs and other non-Watson–Crick base pairs may occur more commonly within undistorted DNA than has previously been thought.

MATERIALS AND METHODS

Protein and oligonucleotide synthesis and purification

The procedure for the purification of the α2 homeodomain and DNA are described elsewhere (7). Briefly, the α2 homeodomain (residues 132–191) was synthesized by the Fmoc solid state peptide synthesis method and purified by C4 reverse phase column (Vydac). Mass spectrometry was used to verify the molecular weight of the protein. The DNA was synthesized at the HHMI–Keck Biopolymer Facility at Yale University and purified with a PureDNA reverse phase column (Rainin). Oligonucleotides were combined in a 1:1 molar ratio and annealed by heating to 90°C, then cooling overnight. Both protein and DNA were quantitated by spectroscopic methods. The DNA concentration was quantitated with the formula 1 OD_{260} unit = 50 μg/ml. The protein concentration was verified by the Lowry method (Sigma). Protein and DNA were combined at a molar ratio of 1.8:1.

Phasing and refinement

Full details of the X-ray data collection, phasing and refinement procedures are described elsewhere (7). Initial trials of molecular replacement (using a model of two α2 proteins bound to DNA from the original structure; 6) on the 2.1 Å resolution data set were unsuccessful, resulting in distorted DNA during refinement with Xplor (9) and CNS (10). Subsequently, electron density maps calculated with multiple isomorphous replacement phases showed that the molecular replacement solution was correct. However, two unexpected, additional α2 molecules were visible in the bones-skeletonized electron density map (11). After the two additional α2 molecules were placed in the density, refinement proceeded smoothly with CNS. For all data to 2.1 Å resolution, the final R_{free} (12) and R factor are 27.8 and 22.2%, respectively.

Sample preparation

The 16 bp DNA fragment used for the NMR studies is 5 bp shorter than the crystallization oligonucleotide and does not contain base pairs 17:27 to 21:23 (Fig. 1C). Single-stranded oligonucleotides were purified twice over a Dynamax PureDNA column (Rainin) with an acetonitrile gradient in 0.1 M triethylamine acetate (pH 7). The trityl group on the oligonucleotide was removed during the second run by 0.5% TFA. Peak fractions were pooled, dialyzed into 10 mM triethylamine bicarbonate (pH 7) and annealed. Annealed oligonucleotides were quantitated by spectroscopic absorbance at 260 nm, lyophilized and stored at –80°C. NMR studies were performed on unlabeled DNA samples, redissolved after HPLC into 25 mM sodium phosphate and 25 mM NaCl buffer (pH 7.2) in D_2O. DNA samples for exchangeable proton detection were transferred in H_2O by lyophilization to a final concentration of 2.0 mM DNA in 25 mM sodium phosphate and 25 mM NaCl buffer (pH 6.4) in 95% H_2O/5% D_2O.

NMR spectroscopy

NMR experiments were carried out at 25°C on General Electric OMEGA PSG and Bruker DMX 600 MHz spectrometers equipped with *x*,*y*,*z*-shielded gradient triple resonance probes. NMR data were processed with NMRPipe (13) and analyzed with NMRVIEW (14). NOE data (mix time τ_m = 150 ms) were obtained from 2D NOESY (15,16). Water suppression was achieved with WATERGATE and field gradient pulses (17) or presaturation pulses during the relaxation delay for DNA samples in H_2O. 2D TOCSY data were obtained in D_2O with a 75 ms clean-MLEV-17 mixing period (18–20). ^1H-^{13}C HMQC (21) was recorded in natural abundance.

Energetics and dynamics calculations

The α2B and α2D proteins were used for the dynamics studies. Structures of the DNA alone, α2B–DNA complex and α2B–α2D–DNA complex were studied by first solvating with water, then neutralizing with sodium ions. The systems were propagated using a multiple timestepping integrator for Langevin dynamics (22,23) as implemented in the CHARMM molecular modeling package, v.26α2 (24).

The systems were minimized and equilibrated to 300K for 75 ps before starting production runs. The molecular simulations and energy minimization use the Cornell *et al.* (25) force fields as implemented in the CHARMM package. Energy minimization of the system was done using the steepest descent method and an adopted-basis Newton–Raphson protocol in CHARMM. (See 26,27 for further details.) The free energy, *f*, was estimated as $f(\chi) = -0.59 \ln n(\chi)$ (kcal/mol), where $n(\chi)$ is the number of counts at χ (28).

Coordinates

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession number 1K61).

RESULTS

A Hoogsteen base pair is seen in a homeodomain/DNA crystal structure

Like other homeodomains, the MATα2 homeodomain (abbreviated as α2) consists of a compact three α-helix domain with a flexible N-terminal arm that becomes ordered upon binding DNA. Homeodomains make base-specific contacts with the major groove of the DNA using residues in the third α-helix, while the N-terminal arm contacts bases in the minor groove. In addition to base-specific contacts, residues in the loop between helices 1 and 2, as well as residues in helix 3, contact the sugar–phosphate backbone. Previous crystal structures of the α2 homeodomain bound to DNA (PDB entry 1APL) (6), the **a**1–α2 heterodimer bound to DNA (PDB entry 1YRN) (29) and the MCM1/α2 heterotetramer bound to DNA (PDB entry 1MNM) (30) show that the α2 homeodomain has the same structure and makes similar DNA contacts when bound to canonical α2 binding sites.

We recently determined a higher resolution structure of the MATα2 homeodomain bound to DNA (7). This new, 2.1 Å resolution structure was determined by a combination of multiple isomorphous replacement and molecular replacement. Briefly, the structure contains four α2 homeodomains, each containing residues 132–191 of the MATα2 protein [homeodomain residues 4–60 as numbered by Qian *et al.* (31)], bound to a 21 bp fragment of duplex DNA that contains two α2 binding sites. Two of the α2 homeodomains, α2A and α2B (Fig. 1A), bind DNA at the two α2 binding sites, while the other two α2 proteins bind DNA in an atypical fashion and are considered non-specifically bound proteins. One of the non-specifically bound homeodomains, α2D, binds DNA in a previously unobserved manner, resulting in a significant reorientation of helix 3 relative to the DNA that disrupts normal side chain–major groove contacts. Instead, other side chains that are not typically involved in homeodomain–DNA interactions are found to mediate base contacts with the major groove. Despite the overall change in the homeodomain docking and major groove contacts, there is little change in the contacts with the sugar–phosphate backbone. The other non-specifically bound α2 homeodomain, α2C, binds near the junction of two stacked, crystallographically related DNA fragments.

Table 1. Twist and rise for the DNA in the crystal structure and standard B-DNA

Base pair	Twist	Rise
C2-G42	36.38	3.88
A3-T41	25.81	3.59
T4-A40	45.44	2.43
G5-C39	26.93	3.09
T6-A38	38.66	3.04
A7-T37	32.73	3.89
A8-T36	32.38	3.29
T9-A35	31.06	3.46
T10-A34	39.12	2.77
C11-G33	44.77	3.22
A12-T32	27.94	3.85
T13-A31	35.29	3.13
T14-A30	35.50	3.38
T15-A29	39.41	2.98
A16-T28	27.29	3.32
C17-G27	36.56	3.18
A18-T26	30.60	3.44
C19-G25	38.06	3.35
G20-C24	26.69	3.24
C21-G23		
Average	34.2	3.29
B-DNA	38.6	3.38

The DNA reported in the current crystal structure contains a total of 21 bp (Fig. 1B). Twenty base pairs are contained within the duplex and one base pair is formed by overhanging 5′ bases from adjacent complexes, which stack end-to-end to form a pseudocontinuous DNA helix. The DNA is B-form throughout, with sugar puckers, axial rises and twist all characteristic of B-DNA (Table 1).

All of the DNA was initially modeled with Watson–Crick base pairs. The $2F_o - F_c$ and $F_o - F_c$ electron density maps fit the Watson–Crick DNA model well except at base pair A7-T37. The difference density at A7-T37 could only be accounted for by the rotation of the A7 base about the torsion angle χ to the uncommon *syn* conformation, yielding an A(*syn*)·T(*anti*) Hoogsteen base pair. Simulated annealing omit maps of the A7-T37 base pair confirm the presence of the Hoogsteen base pair at this position (Fig. 2A). To accommodate the short C1′–C1′ distance across the base pair [8.5 Å for this base pair, as opposed to 10.5 Å for Watson–Crick base

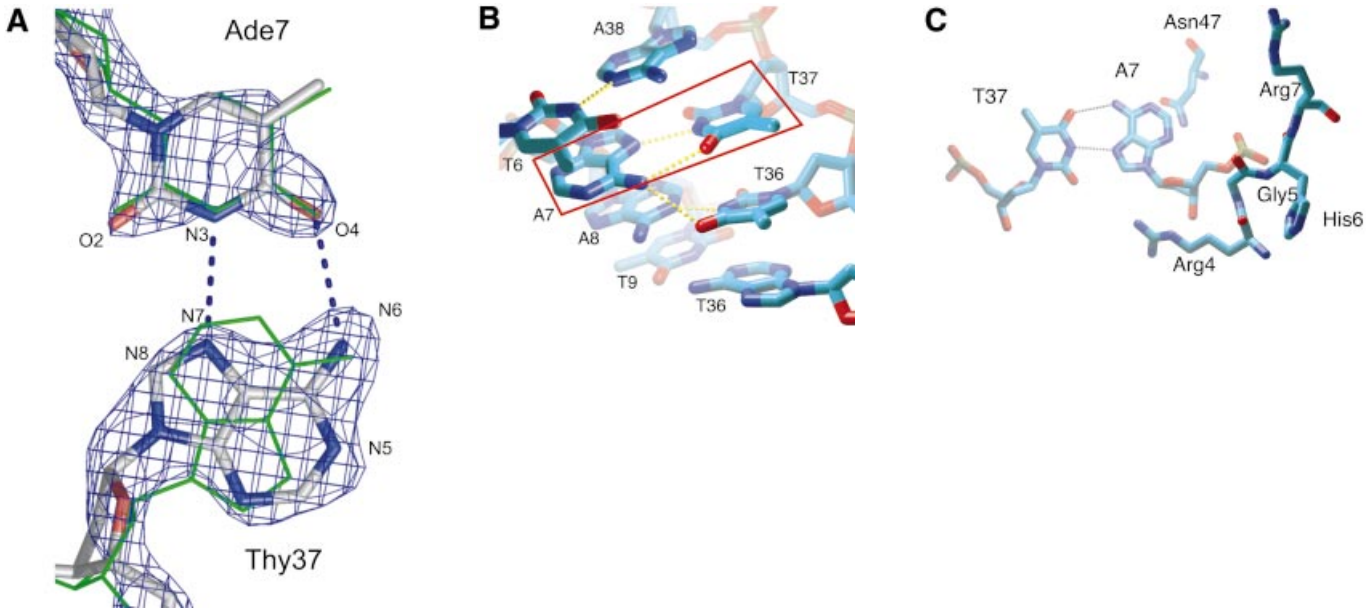

Figure 2. The Hoogsteen base pair is stabilized by intra-DNA and protein–DNA interactions in the crystal structure of the α2 homeodomain–DNA complex. (**A**) Simulated annealing omit map of the A7-T37 base pair. The two hydrogen bonds between the N7 of base A7 and N3 of base T37, as well as between the N6 of A7 and the O4 of T37 characterize a Hoogsteen base pair. The same base pair modeled in the Watson–Crick configuration (green) clearly does not fit the electron density. Figure prepared with Pymol (38). (**B**) The A7 N6 group makes a bifurcated hydrogen bond with both T37 and T36. Base stacking interactions between T6 and A7, as well as A8 and T9, may contribute to stabilization of the Hoogsteen base pair. (**C**) The Arg4 residue of the α2D homeodomain packs against the sugar–phosphate backbone at bases A7 and T6. (B) and (C) prepared with VMD (39).

pairs (4)], the α and γ torsion angles (about bonds P–O5′ and C5′–C4′, respectively) in the sugar–phosphate backbone are in the unusual gauche$^+$/gauche$^-$ conformation about base A7. In normal B-DNA, the α and γ torsion angles are in the gauche$^-$/gauche$^+$ conformations. There are no changes in the conformation of the T37 phosphate backbone α or γ torsion angles and there are no other irregularities in the backbone of the DNA.

The Hoogsteen base pair seen in this crystal structure appears to be a property of this protein–DNA complex and is observed in different crystals grown from independent DNA syntheses. In addition to the 5-iodouracil derivative crystal used for refinement, we also examined the simulated annealing omit maps of the A7-T37 base pair using data from the native crystal, which extends to 2.4 Å resolution. Electron density maps calculated from the native data set also reveal the presence of the Hoogsteen base pair, showing that the 5-iodouracil substitution in the DNA does not affect the formation of the Hoogsteen base pair and that the Hoogsteen base pair occurs in different crystals of this protein–DNA complex. In contrast, neither the DNA from the previous structure of the **a**1–α2–DNA complex nor the DNA in the other α2 binding site in the current α2–DNA crystal structure contains a Hoogsteen base pair. The Hoogsteen base pair therefore appears to be unique to this α2–DNA complex and to the A7-T37 position within the DNA.

The unexpected observation of a Hoogsteen base pair in the present structure raised the possibility that the presence of Hoogsteen base pairs may have been missed in previous structures due to misinterpretation of ambiguous electron density. We examined several structures of protein–DNA complexes in the Protein Data Bank (32) to determine whether any structures contained Hoogsteen base pairs that were mistakenly modeled as Watson–Crick base pairs. Crystal structures of proteins bound to straight DNA determined from crystals diffracting to 2.5 Å resolution or better and with available structure factor files were examined (PDB entries 1FJL, 1B72, 1DP7, 2CGP and 1LAT). $2F_o - F_c$ and $F_o - F_c$ electron density maps were generated with CNS (10) and viewed in xfit from the XtalView package (33). In no case was density indicative of a Hoogsteen base pair seen in the DNA.

In the absence of protein, the DNA in solution does not contain a Hoogsteen base pair

Is the Hoogsteen base pair somehow intrinsic to this DNA sequence? To answer this question we used NMR spectroscopy to examine the structure of a 16 bp DNA fragment in solution. This 16 bp DNA fragment (Fig. 1C) is a shorter version of the crystallized DNA and lacks the 5 bp farthest from the A7-T37 site, but it is sufficiently long to retain local interactions that could potentially have induced formation of the Hoogsteen base pair. In the crystal structure of the α2–DNA complex, only the base pairs immediately adjacent to base pair A7-T37 make stabilizing contacts with the Hoogsteen base pair.

We examined base conformations in solution by following the standard sequential connectivities involving the sugar–base H1′–H6/H8 and H2′/H2″–H6/H8 pathways for all 32 bases using both 2D NOESY (nuclear Overhauser and exchange spectroscopy) and TOCSY (total correlation spectroscopy) experiments in D_2O (see Materials and Methods). If the A7 base is in the *syn* conformation in solution, there should be a strong NOE between A7-H1′ and A7-H8 as the distance between these two protons is 1.5 Å closer than in the *anti* conformation (34). The entire 2D NOESY-D_2O spectrum in

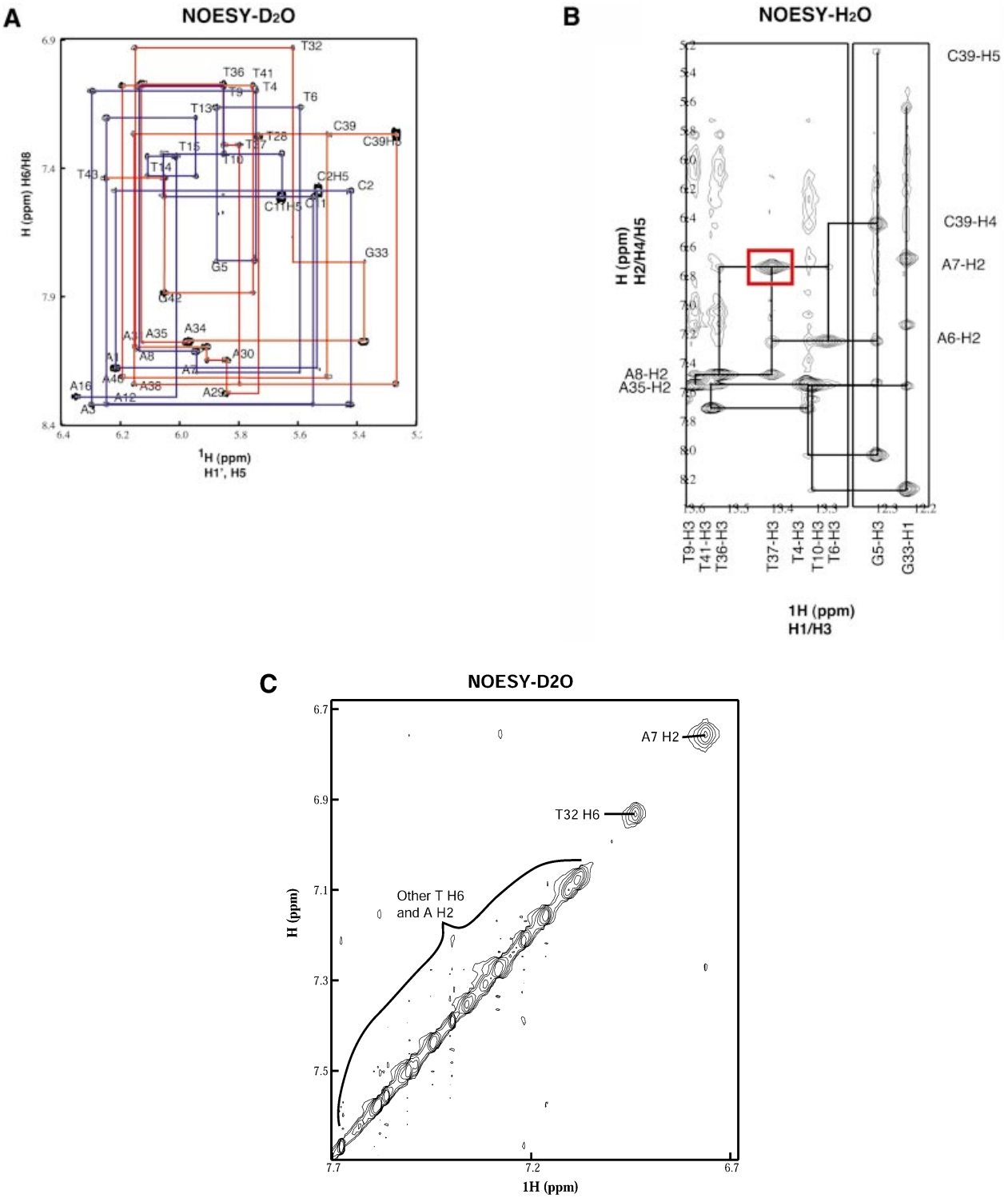

Figure 3. NMR studies of the DNA clearly show a Watson–Crick base pair in solution. (**A**) The NOESY-D_2O spectrum was easily assignable in the H1′–H6/H8 region. The weak A7 H1′–A7 H8 NOE correlation peak is consistent with A7 in the *anti* conformation. (**B**) The NOESY-H_2O spectrum was assignable in the amino–imino region. The A7 H2-T37 H3 NOE correlation (red box) is strong, consistent with a Watson–Crick base pair at A7-T37. (**C**) In the NOESY-D_2O spectrum, the A7-H2 peak is shifted upfield and is broader than other adenine H2 peaks, consistent with a flexible A7 base.

the H1′–H6/H8 (Fig. 3A) and H2′/H2″–H6/H8 (data not shown) regions was well resolved and easy to assign.

The presence of a weak NOE correlation peak between A7-H1′ and A7-H8 indicates that the A7 base is in an *anti* conformation in solution at least 95% of the time, in contrast to the *syn* A7 base conformation that was observed in the crystal structure of the α2–DNA complex. Furthermore, the A7-H2′/H2″ to A7-H8 NOE correlation peaks are strong,

which is another indication of the presence of a Watson–Crick base pair. An A-T Hoogsteen base pair would be characterized by strong imino thymine H3 to adenine H8 NOE cross-peak correlations (2), in contrast to the observed strong imino thymine H3 to adenine H2 NOE cross-peak correlations that are consistent with Watson–Crick base pairing.

The A7-H8 and A7-H2 resonances for the 16 bp DNA construct were reliably assigned using a combination of the 2D NOESY in H_2O and natural abundance ^1H-^{13}C HMQC in D_2O experiments, which also allowed for the assignment of 12 out of 16 bp. The A7-H2 assignments were further confirmed by a strong NOE correlation between A7-H2 and A6-H2 (from base pair A6-T38). Also, we observed no detectable A7-H8 to T37-H3 correlation that would be expected in the Hoogsteen base pair. As shown in Figure 3B, the strong NOE correlation observed between A7-H2 and imino T37-H3 is consistent with A-T Watson–Crick base pair formation. This, together with the lack of a strong A7-H1′ to A7-H8 NOE correlation provides strong evidence for an *anti* conformation for the A7 base in solution. We therefore conclude that the 16 bp DNA fragment itself does not have a propensity to form a Hoogsteen base pair in solution.

While analyzing the NOESY-H_2O peaks, we noticed that the NOE peaks of some adenine H2 protons were broad and shifted upfield. The adenine H2 proton NOE peaks were partially assigned with the help of the ^1H-^{13}C HMQC spectrum. One of the adenine H2 peaks, the A7-H2 peak, is very broad and clearly shifted upfield relative to the other H2 protons by at least 0.2 p.p.m. (Fig. 3C). This broad, upshifted peak is characteristic of the adenine H2 proton of TpA base doubles. The adenine bases of TpA base doubles often have high mobility about the glycosidic χ bond, resulting in the broad peak seen in this and other NMR experiments (35). The high mobility of the TpA base double at T6-A7 may allow the A7 base to flip more easily than in other base doublets, increasing the probability of forming the Hoogsteen base pair upon formation of the α2–DNA complex containing both specifically and non-specifically bound proteins that is observed in this structure.

Protein–DNA and base–base interactions that may stabilize the Hoogsteen base pair

The fact that the Hoogsteen base pair does not form in the absence of bound protein suggests that the α2 proteins must make favorable contacts with the DNA in the crystal that contribute to the formation of the Hoogsteen base pair, A7-T37. In the crystal structure, Arg4 of the non-specifically bound α2D monomer makes van der Waals contacts with the A7 base and the sugar–phosphate backbone of bases T6 and A7 (Fig. 2C). This van der Waals contact may stabilize the Hoogsteen base pair by preventing the sugar–phosphate backbone from moving out to the gauche$^-$/gauche$^+$ conformation for the α and γ torsion angles. When the A7-T37 base pair is modeled as a Watson–Crick base pair, several unfavorable steric clashes are observed between the α2D molecule and the A7 base. In the hypothetical Watson–Crick base pair model, the A7 C3′ deoxyribose atom and the α2D Gly5 main chain N atoms are only 2.2 Å from one another. Additionally, the C8 atom of an A7 base modeled in the *anti* conformation would be within 2.8 Å of the α2D Asn47 ND2 amino group (Fig. 2C). Many of these unfavorable contacts in the Watson–Crick model are also present in the molecular dynamics simulations described below. No other α2 proteins contact the Hoogsteen base pair. These potentially unfavorable contacts could favor a Hoogsteen conformation for the A7-T37 base pair. In contrast, the A29-T15 base pair that is related to A7-T37 by pseudo-two-fold symmetry of the α2 binding sites in the DNA adopts the typical Watson–Crick geometry. The A29-T15 base pair is embedded within the same sequence as the A7-T37 base pair, differing only in the absence of a non-specifically bound homeodomain counterpart to α2D.

Overall, the A7-T37 base pair in the Hoogsteen conformation has several favorable interactions within the DNA around the Hoogsteen base pair, and between the non-specifically bound α2D protein and the A7-T37 base pair, that would not be present were this base pair in the Watson–Crick conformation. Besides the protein–DNA interactions that favor the Hoogsteen base pair, interactions with adjacent base pairs may help stabilize the Hoogsteen base pair. In addition to the hydrogen bonds that constitute the Hoogsteen base pair interaction, the *syn* conformation of base A7 and negative roll of base T37 allow the adenine 7 N6 group to make an additional hydrogen bond with thymine 36 O4 in the adjacent A8-T36 base pair (Fig. 2B). This intra-base pair hydrogen bond does, however, come at the cost of base pair hydrogen bonds that are lengthened beyond 3.2 Å in the T6-A38 and A8-T36 base pairs, resulting in only one hydrogen bond between them instead of the two hydrogen bonds in a normal A-T base pair. The Hoogsteen base pair may be further stabilized by favorable base stacking interactions between the A7 and T6 bases, and the A8 and T9 bases, where the six-membered rings of the adenine and the thymine stack favorably on one another (Fig. 2B). Thus, several favorable interactions within the DNA and between the non-specifically bound α2D protein and DNA could preferentially stabilize the Hoogsteen base pair over the Watson–Crick base pair.

Energetics calculations and molecular dynamics simulations of the protein–DNA system

Molecular dynamics simulations were used to study the influence of the α2 homeodomains on the stability of the Hoogsteen base pair. The system studied by molecular dynamics consisted of either 16 bp DNA alone, the DNA with the specifically bound α2B homeodomain, or DNA with the specifically bound α2B and the non-specifically bound α2D homeodomains. The DNA in the simulations was 1 bp shorter than the oligonucleotide duplex used in the NMR experiments (Fig. 1D). Each system was fully solvated and neutralized with sodium ions. The dynamics simulations were run for 1 ns using Langevin dynamics (22,23) implemented in CHARMM (24). Fluctuations of the dihedral angles and distances are used to estimate free energies (not on an absolute scale; see Materials and Methods), not the relative height of the free energy barriers.

The presence of the non-specific α2D protein minimizes the free energy of the base pair at $\chi = 52°$, as expected for a Hoogsteen base pair (Fig. 4A). Additionally, the distribution of energies is much sharper in the presence of the α2D protein, with a potential well width of 5° (compared to >20° without α2D), indicating that the Hoogsteen base pair is more stable in the presence of the α2D protein. The large fluctuations of the

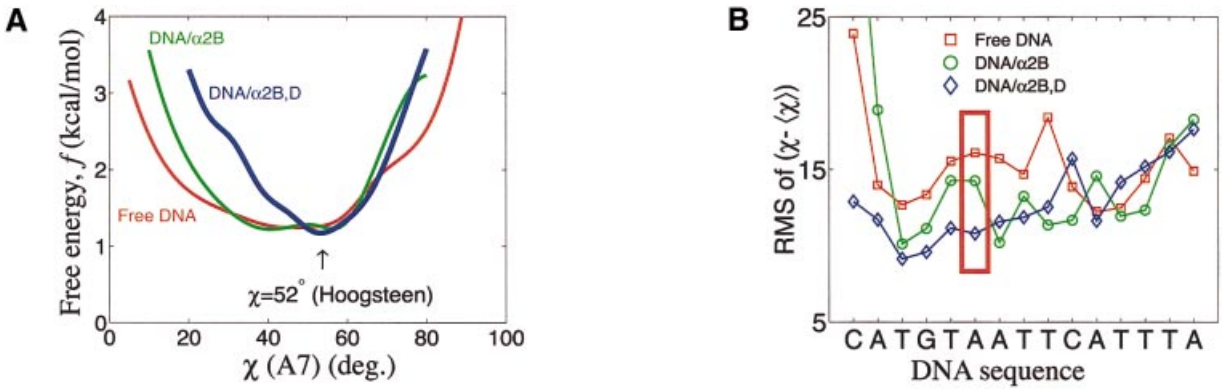

Figure 4. (**A**) Energetics of the A7 base calculated from glycosidic angle χ in molecular dynamics simulations. The presence of bound α2 proteins stabilizes the χ angle at 52°, while free DNA is free to rotate over a broad range. (**B**) The fluctuations of the χ angle around the A7 base (red box) decrease significantly when the α2 proteins are bound to the DNA.

DNA in the region of the A7 base suggest a possible low transition energy barrier for flipping the A7 base in the DNA in the absence of bound α2 proteins.

Molecular dynamics simulations confirm many of the contacts between the non-specifically bound α2D protein and the DNA in the crystal structure, as well as revealing additional stabilizing contacts that favor the Hoogsteen base pair. The dynamics simulations show that there are many contacts between the α2D Arg4 residue and the sugar–phosphate backbone at base A7. Additional hydrogen bonds not seen in the crystal structure form between α2D Arg4 and the O3′ and O4′ atoms of base A7 during the simulations. Furthermore, a simulation run with the A7-T37 base pair in the Watson–Crick conformation shows that the peptide backbone of the α2D N-terminal arm moves away from the minor groove to avoid steric clashes, leading to loss of hydrogen bonds and van der Waals contacts between the α2D protein and DNA. These steric clashes are similar to the clashes observed when the A7-T37 base pair was modeled in the Watson–Crick configuration in the crystal structure.

The fluctuations of the glycosidic χ angle at the A7 base pair in the molecular dynamics simulations show that the TpA base doubles, which show characteristic large fluctuations of the χ angle, may be contributing to the formation of the Hoogsteen base pair. In addition to the broadened peaks observed at A7-H2 in the NMR experiments, the simulations also show a 16° root mean square fluctuation of the A7 χ angle. Furthermore, stacking energy calculations using the AMBER force field yield low stabilization energies for base pairs between base pairs T4-A40 and T10-A34 (Fig. 4B). These low stabilization energies and large fluctuations may be favorable for flipping the adenine base in solution, and the Hoogsteen base pair may not subsequently revert back to the Watson–Crick base pair due to the favorable α2D–DNA and intra-DNA contacts for the Hoogsteen base pair.

DISCUSSION

Implications for protein–DNA interactions

We have observed a Hoogsteen base pair embedded in the structure of undistorted dsDNA. Hoogsteen base pairs have previously been observed in crystal structures of drug–DNA complexes (such as triostin A–DNA) with underwound DNA (1), in DNA severely distorted by binding of the transcription factor TBP (3) and at the ends of DNA oligonucleotides used in crystallization (5). In all of these cases, the base stacking energy barrier to forming the Hoogsteen base pair rather than the more typical Watson–Crick base pair is presumably lowered by the intercalation of drugs, bending and unwinding of the DNA, or the presence of free, unpaired bases, all of which have the flexibility to allow the flipping of a purine base from the normal *anti* conformation to the *syn* conformation. To our knowledge, a Hoogsteen base pair has not previously been seen in oligonucleotide structures that lack any of these kinds of distortions.

The crystal structure described here contains both intra-base pair interactions and protein–DNA interactions that could stabilize the Hoogsteen base pair. However, in the absence of bound α2 proteins, the DNA does not contain a Hoogsteen base pair, as shown in solution NMR studies of the free DNA. The NMR studies instead show a broadened A7-H2 peak that may indicate that the A7 base is free to rotate about the glycosidic χ bond. We speculate that the flexible A7 base may spontaneously flip about the χ bond to the unusual *syn* conformation, leading to formation of the Hoogsteen base pair. Molecular dynamics simulations of the DNA and protein–DNA systems confirm the stability of the complex of α2 proteins bound to DNA containing the Hoogsteen base pair. The dynamics simulations also show that the A7 base indeed has an inherent flexibility, with significant fluctuations about its glycosidic χ bond angle. From these experiments, however, we cannot determine whether the Hoogsteen base pair is a result of the binding of the non-specifically bound α2D protein or whether the presence of the Hoogsteen base pair may stabilize the α2D contacts with DNA.

It appears that the energetics of the particular configuration of proteins and DNA in the present structure gives rise to this unusual base pair, despite the absence of DNA distortions previously observed to be required for Hoogsteen base pair formation within duplex DNA. A combination of van der Waals interactions, hydrogen bonds and base stacking interactions may allow the stabilization of the A7-T37 Hoogsteen base pair by the α2 proteins in this structure. These base fluctuations of the A7 base between *syn* and *anti* may happen

at a very low frequency due to base flipping out of the DNA and reinsertion into the DNA. Such base flipping has been observed and predicted to require an energy of 25 kcal/mol (36), much less than the 100 kcal/mol or greater predicted to flip the A7 base within DNA (data not shown). One such base flipping event, a Hoogsteen base pair that appears to require stabilization by an α2 protein bound to the DNA, has been observed in the current crystal structure. Because the Hoogsteen base pair is only present in the crystal structure and not in the DNA alone in NMR experiments, we cannot determine whether the presence of the Hoogsteen base pair could influence the binding affinity of the α2 protein for the DNA.

Our observation of a Hoogsteen base pair within otherwise undistorted B-DNA that is either induced or stabilized by protein–DNA contacts raises the possibility that Hoogsteen base pairs could occur within cellular DNA and play a role in protein–DNA interactions. The particular configuration of proteins and DNA reported here is undoubtedly influenced by the non-physiological concentrations of α2 protein in the crystal drops and does not reflect the arrangement of binding sites found upstream of genes regulated by α2 *in vivo*. Nevertheless, it is possible that the local conditions under which the present Hoogsteen base pair forms could be duplicated for other proteins at *in vivo* regulatory sites. The presence of multiple overlapping binding sites is common in chromosomal DNA and could give rise to a configuration of proteins analogous to that observed in the crystal. The open question is whether such an arrangement of proteins either binds preferentially to transiently formed Hoogsteen base pairs or favors Hoogsteen base pair formation in order to form optimal interactions. Since we were unable to detect measurable Hoogsteen base pair formation in free DNA, it was not possible to assess the energetic contribution of Hoogsteen base pair formation by directly comparing the DNA-binding affinity of the α2 homeodomain for sites containing Hoogsteen versus Watson–Crick base pairs. However, the absence of DNA distortion and the relatively typical array of protein–DNA contacts suggests that the conditions that favor Hoogsteen base pair formation could be replicated in a cellular context. These observations raise the intriguing possibility that Hoogsteen base pair formation could potentially play a role in the binding of proteins to undistorted B-DNA, although further investigation will be required in order to determine whether this indeed occurs in the cell.

ACKNOWLEDGEMENTS

We thank W. Olson for advice and discussions and for allowing us to preview the 3DNA program, M. Summers for generously allowing the use of equipment for the NMR experiments, and C. Garvie, A. VanDemark, A. Ke, N. LaRonde-LeBlanc, P. Minary and R. Campbell for helpful discussions. This work was supported by NSF grant MCB9808412 (C.W.).

REFERENCES

1. Wang,A.H., Ughetto,G., Quigley,G.J., Hakoshima,T., van der Marel,G.A., van Boom,J.H. and Rich,A. (1984) The molecular structure of a DNA-triostin A complex. *Science*, **225**, 1115–1121.
2. Gilbert,D.E., van der Marel,G.A., van Boom,J.H. and Feigon,J. (1989) Unstable Hoogsteen base pairs adjacent to echinomycin binding sites within a DNA duplex. *Proc. Natl Acad. Sci. USA*, **86**, 3006–3010.
3. Patikoglou,G.A., Kim,J.L., Sun,L., Yang,S.H., Kodadek,T. and Burley,S.K. (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.*, **13**, 3217–3230.
4. Hoogsteen,K. (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.*, **16**, 907–916.
5. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
6. Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, **67**, 517–528.
7. Aishima,J. and Wolberger,C. (2002) Crystal structure of the MATalpha2 homeodomain-DNA complex with nonspecifically bound homeodomains. *Proteins Struct. Funct. Genet.*, in press.
8. Hodel,A., Kim,S.-H. and Brunger,A.T. (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr.*, **A48**, 851–858.
9. Brunger,A.T. (1992) *X-PLOR, Version 3.1. A System for X-ray Crystallography and NMR*, 3.84 Edn. Yale University Press, New Haven, CT.
10. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.
11. Kleywegt,G.J. and Jones,T.A. (1996) xdlMAPMAN and xdlDATAMAN—programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr.*, **D52**, 826–828.
12. Brunger,A.T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
13. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
14. Johnson,B.A. and Blevins,R.A. (1994) NMRview: a computer program for the visualization and analysis for NMR data. *J. Biomol. NMR*, **4**, 603–614.
15. Jeener,J., Maier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.
16. Macura,S. and Ernst,R.R. (1980) Elucidation of cross relaxation in liquids by two-dimensional NMR-spectroscopy. *Mol. Phys.*, **41**, 95–117.
17. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
18. Greisinger,C., Otting,G., Wuthrich,K. and Ernst,R.R. (1988) Clean TOCSY for 1H spin system identification in macromolecules. *J. Am. Chem. Soc.*, **110**, 7870.
19. Bax,A. and Davis,D.G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.*, **65**, 355–360.
20. Braunschweiler,L. and Ernst,R.R. (1983) Coherence transfer by isotropic mixing—application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**, 521–528.
21. Bax,A. and Subramanian,S. (1986) Sensitivity-enhanced two-dimensional heteronuclear shift correlation NMR-spectroscopy. *J. Magn. Reson.*, **67**, 565–569.
22. Schlick,T. (2001) Time-trimming tricks for dynamic simulations: splitting force updates to reduce computational work. *Structure*, **9**, R45–R53.
23. Barth,E. and Schlick,T. (1998) Overcoming stability limitations in biomolecular dynamics. I. Combining force splitting via extrapolation with Langevin dynamics in LN. *J. Chem. Phys.*, **109**, 1617–1632.
24. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.*, **4**, 187–217.
25. Cornell,W.D., Cieplak,P., Bayley,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and

Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
26. Qian,X., Strahs,D. and Schlick,T. (2001) Dynamic simulations of 13 tata variants refine kinetic hypotheses of sequence/activity relationships. *J. Mol. Biol.*, **308**, 681–703.
27. Strahs,D. and Schlick,T. (2000) A-tract bending: insights into experimental structures by computational models. *J. Mol. Biol.*, **301**, 643–663.
28. Kottalam,J. and Case,D.A. (1988) Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.*, **110**, 7690–7697.
29. Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.
30. Tan,S. and Richmond,T.J. (1998) Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature*, **391**, 660–666.
31. Qian,Y.Q., Billeter,M., Otting,G., Muller,M., Gehring,W.J. and Wuthrich,K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580. [Erratum (1990) *Cell*, **61**, 548.]
32. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
33. McRee,D.E. (1999) XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, **125**, 156–165.
34. Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.
35. McAteer,K. and Kennedy,M.A. (2000) NMR evidence for base dynamics at all TpA steps in DNA [In Process Citation]. *J. Biomol. Struct. Dyn.*, **17**, 1001–1009.
36. Chen,Y.Z., Mohan,V. and Griffey,R.H. (2000) Spontaneous base flipping in DNA and its possible role in methyltransferase binding. *Phys. Rev.*, **E62**, 1133–1137.
37. Evans,S.V. (1993) SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, **11**, 134–138, 127–138.
38. Delano,W.L. (2002) *The PyMOL Molecular Graphics System*. Delano Scientific, San Carlos, CA.
39. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD—visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.