X-ray and NMR Studies of the DNA Oligomer d(ATATAT): Hoogsteen Base Pairing in Duplex DNA†

Nicola G. A. Abrescia,*,‡,§ Carlos González,|| Catherine Gouyette,# and Juan A. Subirana‡

Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Av. Diagonal 647, E-08028 Barcelona, Spain, Instituto de Química Física Rocasolano, CSIC, Serrano 119, E-28006 Madrid, Spain, and Unité de Chimie Organique, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris, France

Received August 24, 2003; Revised Manuscript Received December 10, 2003

ABSTRACT: We present and analyze the structure of the oligonucleotide d(ATATAT) found in two different forms by X-ray crystallography and in solution by NMR. We find that in both crystal lattices the oligonucleotide forms an antiparallel double helical duplex in which base pairing is of the Hoogsteen type. The double helix is apparently very similar to the standard B-form of DNA, with about 10 base pairs per turn. However, the adenines in the duplex are flipped over; as a result, the physicochemical features of both grooves of the helix are changed. In particular, the minor groove is narrow and hydrophobic. On the other hand, d(ATATAT) displays a propensity to adopt the B conformation in solution. These results confirm the polymorphism of AT-rich sequences in DNA. Furthermore, we show that extrahelical adenines and thymines can be minor groove binders in Hoogsteen DNA.

In this paper, we present a new crystal structure of the oligonucleotide d(ATATAT). It forms antiparallel duplexes in which Watson−Crick pairing has been replaced by Hoogsteen pairing **(**Figure 1**)**. These results support our previous study (*1*) and demonstrate that Hoogsteen pairing in DNA duplexes may be found under a variety of conditions. In fact, a recent theoretical analysis (2) shows that the energies of Hoogsteen and B-form DNAs are practically identical. We analyze in detail the conformation of Hoogsteen DNA by comparing the structures found in two different crystal forms. We have also studied the solution structure of this oligonucleotide by NMR.

Soon after the model of the double helix was proposed (*3*) efforts were made to demonstrate that the different features of the model were correct. A surprise came when Hoogsteen (*4*) and other investigators (*5*) found that adenine and thymine paired in crystals in a different way, called Hoogsteen pairing. However, a large number of studies, including the X-ray structure of many protein−DNA complexes and, in particular, X-ray and NMR studies of oligonucleotides, have shown that the normal way of pairing in duplex DNA is the one suggested by Watson and Crick. To our surprise, we have found (*1*) that the hexanucleotide d(ATATAT) could fully crystallize in the Hoogsteen mode.

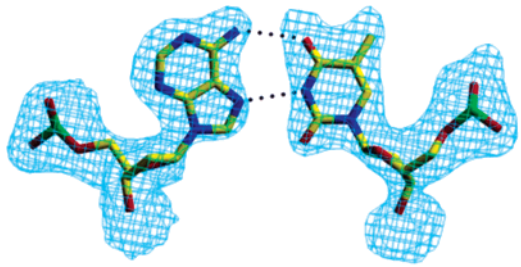

FIGURE 1: SIGMAA-weighted $2F_o - F_c$ electron density contoured at 1σ of base pair A3·T10 in the $P6_5$ crystal structure. The Hoogsteen conformation is evident, it does not depend on the presence of either bromine or methyl in the pyrimidine base (*1*).

Despite the lack of detailed crystallographic studies, it is known that DNA sequences rich in adenine and thymine are strongly polymorphic (*6*−*9*), even in the presence of standard Watson−Crick base pairs. Alternative base pairing schemes have been found in other cases (*10*−*12*). Isolated Hoogsteen base pairs are formed in some special cases as in the closing loop of DNA hairpins (*13*−*15*) and in circular oligonucleotides (*16*). Also when the nucleic acid structure is chemically modified (*17*, *18*) or the DNA is locally influenced by either drug intercalation (*19*) or some protein interactions (*20*, *21*). Hoogsteen base pairs are found in triple helices (*22*) and have also been suggested in models for fiber diffraction patterns of dTpdA (*23*).

METHODS AND MATERIALS

Crystallization and Structure Determination. The palindromic deoxyhexanucleotide d(ApTpApTpApTp) and its bromo-uracil derivative at position four were crystallized at 13 °C using the hanging-drop vapor diffusion technique. Crystals could be obtained from sodium cacodylate buffer (25 mM, pH 6) containing spermine (3 mM) and either Na^+

† This work has been supported by Grants BIO2002-00317 of the Ministerio de Ciencia y Tecnología and 2001SGR 00250 from the Generalitat de Catalunya.

* To whom correspondence should be addressed: Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. Telephone: 44 1865 287549. FAX: 44 1865 287547. E-mail: nicola@strubi.ox.ac.uk.

‡ Universitat Politècnica de Catalunya.

|| Instituto de Química Física Rocasolano, CSIC.

Institut Pasteur.

§ Present address: Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

10.1021/bi0355140 CCC: $27.50
Published on Web 03/17/2004

Table 1: Summary of Data Collection and Refinement Statistics[a]

sequence	d(ApTpApTpApT)	d(ApTpAp^{Br}UpApT))	d(ApTpApTpApT)
space group	$P2_1$	$P2_1$	$P6_5$
unit cell (Å)	$a = 23.66$, $b = 48.37$	$a = 23.96$, $b = 48.95$	$a = b = 32.31$
	$c = 32.23$, $\beta = 92.82°$	$c = 32.24$, $\beta = 93.27°$	$c = 117.77$
resolution limit (Å)	2.82	2.50	2.17
R work (%)		22.0	21.8
R free (%)		23.4	26.2

[a] Additional data have been given in ref *1* for $P2_1$ and ref *25* for $P6_5$.

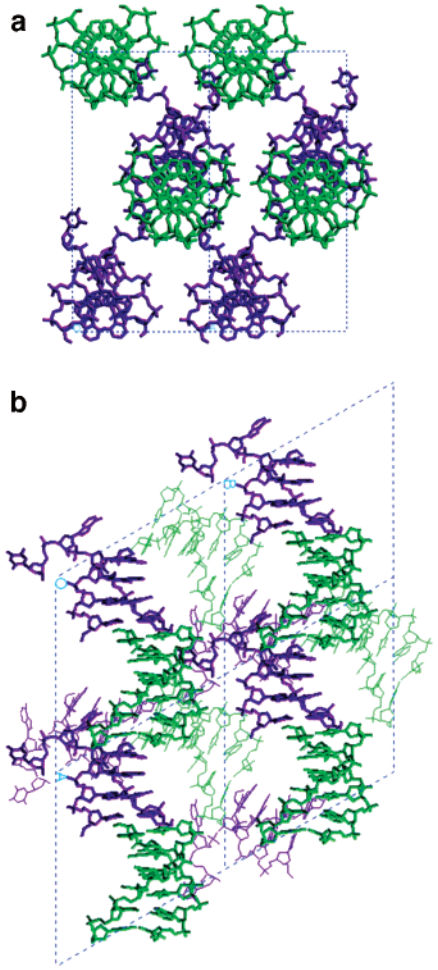

FIGURE 2: Packing of d(ATATAT) in two crystal forms: (a) Two $P2_1$ unit cells projected onto a plane perpendicular to the *z* axis; (b) view of two layers of molecules in the $P6_5$ crystal projected onto a plane perpendicular to the *z* axis. A section of four unit cells is shown in the upper layer. In the lower layer, the molecules are shown with thinner lines. Parallel columns of molecules in consecutive layers cross at an angle of 60°. The hexamer duplex is shown in green; the tetramer duplex with its extrahelical bases is in purple. All base pairs are in the Hoogsteen conformation.

(18 mM) or K^+ (9 mM). 2-Methyl-2,4-pentanediol (MPD) was used as a precipitant.

Diffraction X-ray data were collected at either DESY (Hamburg, Germany) or ESRF (Grenoble, France) synchrotron radiation sources. Full data collection and structure refinement statistics have been given elsewhere for the $P2_1$ structure (*1*). A summary is given in Table 1. The $P2_1$ structure was solved by using anomalous diffraction data of the Br atoms. On the other hand, the $P6_5$ structure was solved by molecular replacement with AMoRe (*24*) using as search model the structure of the same sequence previously solved at 2.5 Å in a different space group (*1*). The extrahelical bases were omitted. Because of twinning and pseudosymmetry, the solution was not straightforward. The details of the successful protocol as well as the initial refinement procedure are described elsewhere (*25*). Further refinement was performed with CNS (*26*) using the scripts for the merohedral twinning case and monitored through the free-R test (*27*). Attempts to refine two of the extrahelical terminal bases in the weak residual density observed in the minor groove did not result in either an improvement of the electron density maps or in a significant lowering of *R* values. So the final model contains only two extrahelical bases, and the other two were omitted, as shown in Figure 2a. A few water molecules were added at this stage. The α twin fraction was recalculated using the current model. Simulated annealing was performed to eliminate any model bias introduced by the starting model. It was followed by a gradient conjugate minimization. The water molecules were fixed and, hydrogen bonds were restrained. A final atomic B-factor refinement ended the process. Further details on refinement statistics of the $P6_5$ structure are given in Table 1. An example of the electron density map is given in Figure 1. Coordinates for both models have been deposited at NDB (code UD0035 for $P2_1$ and UD0049 for $P6_5$).

NMR Spectroscopy. NMR samples were prepared in 500 μL of either D_2O or 9:1 H_2O/D_2O. The resulting buffer solution (25 mM phosphate, 1 M KCl, pH = 7.0) was about 2 mM in duplex. The spectra were acquired in a Bruker DMX spectrometer operating at 600 MHz, and processed with the UXNMR software. DQF-COSY, TOCSY (*28*), and NOESY (*29*) experiments were recorded in D_2O and in 90% H_2O/10% D_2O. All 2D experiments were carried out at −2 °C. The NOESY spectra in D_2O were acquired with mixing times of 200 ms, and 100 ms in H_2O. TOCSY spectra were recorded with 80-ms mixing times. A jump-and-return pulse sequence (*30*) was employed to observe the rapidly exchanging protons in H_2O experiments. The spectral analysis program XEASY (*31*) was used for the assignment of the NOESY cross-peaks.

RESULTS

Crystal Structure. In the $P2_1$ structure previously described (*1*), it was found that the DNA duplexes form infinite parallel columns of pseudocontinuous helices as shown in Figure 2. Hexamer and tetramer duplexes alternate in such columns. The tetramers have two extrahelical AT bases, with T inserted in the minor groove of neighbor hexamers. The $P6_5$ structure, which we now report, is also shown in Figure 2. It has similar infinite columns of tetramer/hexamer duplexes, but in this case parallel columns are organized in planes which cross each other at 60° angles. One of the extrahelical thymines enters the minor groove of a neighbor column in the same plane, as shown in Figures 2−4. The other extra-

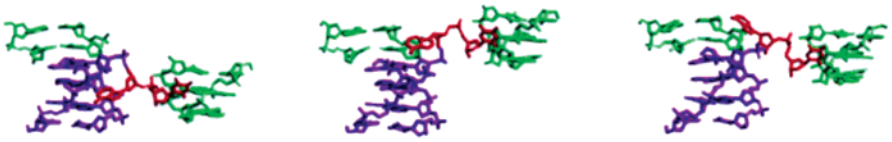

FIGURE 3: Comparison of the position of the extrahelical bases (shown in red) in the $P2_1$ (left, middle) and $P6_5$ crystal structures (right). At the right of each drawing, the three terminal bases of an hexamer duplex are shown, with thymine in the minor groove. Note that the position of adenines and the relative height of neighbor hexamers is different in each case. Thymines occupy practically identical positions associated by hydrogen bonding with the second thymine at either end of the hexamer duplex, as shown in detail in Figure 4.

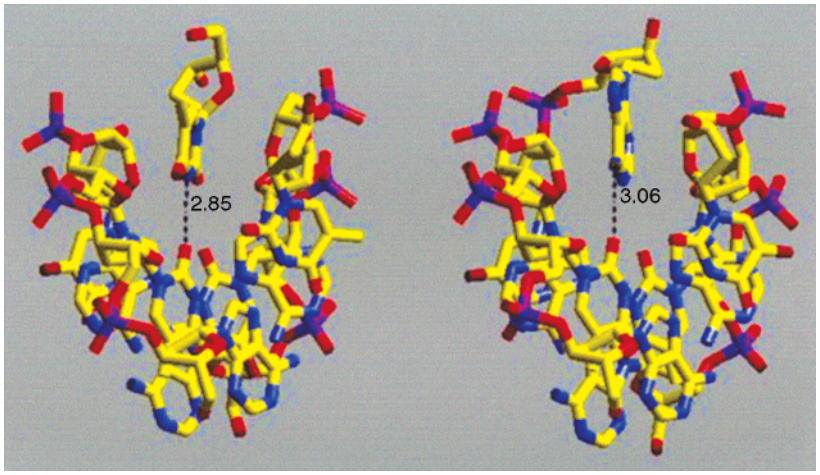

FIGURE 4: Close-up view of the looped-out bases and details of the interactions of the minor groove with either thymine in the $P6_5$ structure (left) or adenine in the $P2_1$ structure (right). In both cases, a nitrogen atom of the external base forms a hydrogen bond with an O2 atom of thymine from the duplex. Note the strong similarity of both types of interaction. The looped out thymines in $P2_1$ show the same structure, as it is apparent in Figure 3. In the case of adenine, the phosphodiester backbone folds back and the base enters into the minor groove of the same duplex, while the terminal thymine enters into a neighbor duplex (Figure 3, left).

helical thymine also enters the minor groove of a neighbor hexamer, but in the next plane of duplexes: it appears to occupy different alternative positions, since it could not be uniquely located in the electron density map.

It should be noted that in both cases the crystal structure presents several channels full of solvent, which explain the limited resolution of the data collected. The volume of crystal occupied by a base pair is 1478 Å^3/base pair in $P6_5$ and 1573 Å^3/base pair in $P2_1$, significantly larger than in the usual B-form crystals (about 1300 Å^3/base pair). No water molecules could be detected forming duplex−duplex bridges between the layers of the $P6_5$ structure, a fact that indicates a limited order in this interlayer region. The latter observation is consistent with the lack of order found in the AT bridge which extends from one layer to the next.

Hoogsteen Pairing. The electron density found in our structure can only be matched when the bases are placed in a Hoogsteen conformation. Furthermore, the C1′−C1′ distance is only 8.12 Å on the average, as expected for a Hoogsteen conformation, and quite different from the 10.85 Å distance found in B-form DNA. Hydrogen bonds have normal distances. Their values are given in Table 2. The O2 of thymine also forms a weak hydrogen bond with C8 of adenine, which is longer than normal, as also shown in Table 2. Perhaps due to this fact the Hoogsteen base pairs may have a greater variability of hydrogen bond distances as indicated by the standard deviations given in Table 2. An additional point to note is that the presence of bromo-uracil instead of thymine does not have any noticeable effect on either the dimensions of the Hoogsteen base pairs or the duplex structure in general. An obvious result of Hoogsteen base pairing is that the nucleotide conformation changes from *anti* to *syn*, with values of the glycosidic angle in the range 60−90°. A similar change is found in Z-form DNA, where the guanine base undergoes a similar rotation.

Table 2: Geometry of the Hoogsteen Base Pairs[a]

	distances (Å)			
structure	N6(A)−O4(T)	N7(A)−N3(T)	C8(A)−O2(T)	C1′−C1′
$P2_1$	3.16 (0.16)	2.88 (0.11)	3.39 (0.13)	8.11 (0.14)
$P6_5$	2.99 (0.17)	2.85 (0.04)	3.34 (0.10)	8.13 (0.10)

[a] The average values for base pairs in the crystal structure are given, with standard deviations in parentheses. Base pairs A5·T8, A6·T7, and A19·T16 in $P6_5$ are not included: they have significantly higher values in the last two columns (about 0.3 Å) due to distortion of the base pair.

Duplex Structure. The overall appearance of the Hoogsteen duplex is very similar to standard B-form DNA, as it is apparent from Figure 2. A space filling model is given in Figure 5. Its helical parameters are also very similar (Table 3). Within the duplex hexamers and tetramers, the average twist is about 34.6°, which corresponds to 10.4 base pairs/turn, practically identical to B-form DNA. The minor groove is narrow due to the shorter C1′−C1′ distances, as it is apparent in Figure 5. It ranges between 9.3 and 11.1 Å. Similar values are also found in some AT-rich oligonucleotides crystallized in the B-form (*32*).

The main difference with B-form DNA lies in the change in the accessibility of the grooves toward interaction with solvent/proteins. The N3 atom of adenine, which lies in the minor groove in B-form DNA, is now moved to the major groove. As a result the minor groove becomes less elec-

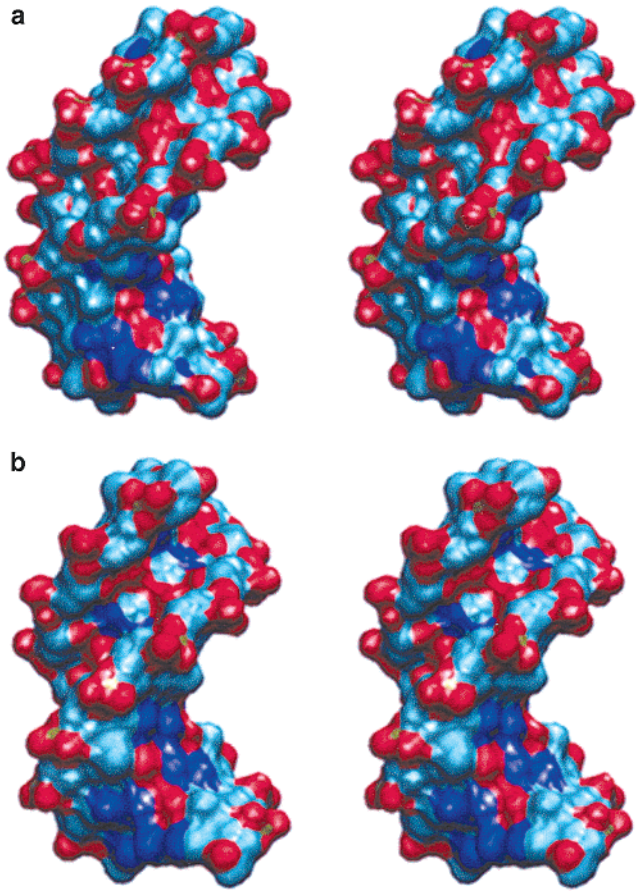

FIGURE 5: Stereo picture of (a) Hoogsteen DNA and (b) ideal AT B-DNA (10 base pairs) rendered by surface accessibility (probe 1.4 Å radii) and colored by atom. Differences in the distribution of oxygen and nitrogen along the minor and major grooves between the two conformations are apparent. A similar comparison with stick models was presented elsewhere (Figure 2 in ref *1*).

Table 3: Selected Helical Parameters[a]

crystal	duplex	twist (°)	rise (Å)	slide (Å)
$P2_1$	hexamer	35.2 (1.9)	2.77 (0.40)	−2.31 (0.40)
$P2_1$	tetramer	34.1 (1.0)	3.32 (0.22)	−2.35 (0.09)
$P2_1$	inter[b]	40.8	4.21	−4.96
$P2_1$	overall[c]	36	3.22	
$P6_5$	hexamer	34.2 (2.1)	3.37 (0.18)	−2.34 (0.34)
$P6_5$	tetramer	34.5 (2.1)	3.02 (0.29)	−2.35 (0.09)
$P6_5$	inter[b]	42.4	3.25	−4.90
$P6_5$	overall[c]	36	3.24	

[a] Standard deviations are given in parentheses. [b] The values given correspond to the steps between the tetramer and hexamer duplexes. [c] Average values considering the tetramer and hexamer together as a continuous duplex.

tronegative (Figure 5); it has a single hydrogen bond acceptor atom (O2 of thymine). This fact, together with the narrowness of the minor groove, indicates that it may become an appropriate target for interaction with hydrophobic groups, as it is actually found in our structures and will be discussed below.

There are a few other features that differentiate Hoogsteen DNA from standard B-form DNA: the position of the helical axis, the degree of overlap of consecutive base pairs and some conformational angles. The helical axis is now found at the edge of the base pairs, approximately at the midpoint of the hydrogen bond between adenine N6 and thymine O4.

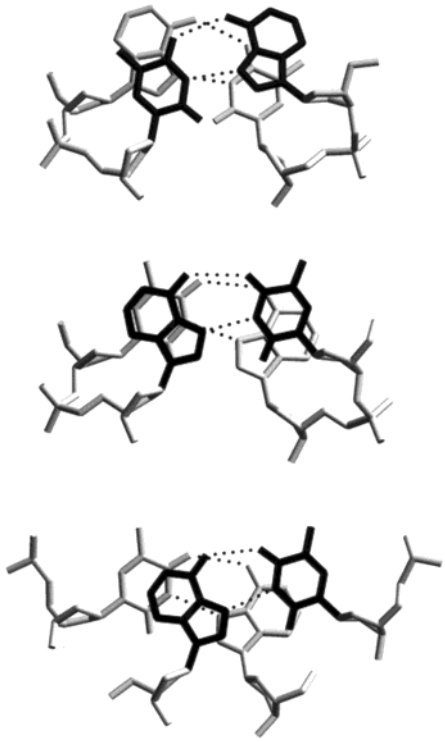

FIGURE 6: Comparison of the three types of stacking found in Hoogsteen DNA $P2_1$ crystals: top, an AT step (A3-^{Br}U4); middle, a TA step (T20-A21). Note that in the latter case, the C5′ atom is moved toward the bases as discussed in the text. Bottom, the end-to-end interaction between tetramers and hexamers, which is analogous to a TA step. By comparison with the middle figure, it is apparent that the upper base pair slides toward the right. All equivalent base steps in both $P2_1$ and $P6_5$ structures are practical identical with those shown in this figure.

The major groove is in the center of the helix, and the phosphates are externally located, a situation also found in A-form DNA.

The oligonucleotide we have studied has an alternating sequence, and therefore it has two types of base steps: AT and TA. In B-form DNA the latter base steps differ considerably in the twist parameter (*32*), whereas in the Hoogsteen case the twist value is similar in both cases. The stacking of base pairs is also different in both DNA forms. In Hoogsteen DNA, there is a very good overlap of the aromatic rings in the TA steps and very poor in the AT steps (Figure 6), whereas the opposite is true in B-form DNA (*32*). Furthermore, in the Hoogsteen form in some of the TA base steps the conformational α/γ angles appears in a g^+/g^- conformation, which is seldom found in B-form DNA, although it is common in the Z-form (*33*). The net result of this change in conformation is that the C5′ atom is placed in an inner position in the duplex, whereas in the usual g^-/g^+ conformation it lies in a more external position, as it is apparent in Figure 6.

Hydration. The pattern of external hydrogen bonding sites in Hoogsteen DNA is clearly different from the other canonical DNA forms. In the Hoogsteen structure, the base atoms which can form hydrogen bonds are mainly on the major groove side. The latter is covered by a network of water molecules, both in the $P2_1$ and $P6_5$ structures (Figure 7). They form hydrogen bonds among themselves, with phosphate oxygens and with the exposed nitrogen and oxygen atoms of the bases. In particular, most of the N1

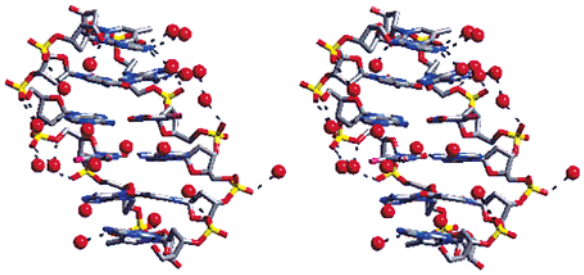

FIGURE 7: Stereopair of major groove hydration of the $P2_1$ structure. Typical hydrogen bonds cover the N1 atom of adenine and phosphates, among other atoms.

atoms of adenines are hydrogen bonded to a water molecule. In the Hoogsteen conformation, the N1 atom lies on the major groove side, whereas in Watson−Crick base pairs it is hydrogen bonded to thymine. The minor groove has only the O2 atom of thymine which could accept a hydrogen bond from water. Furthermore, the minor groove is occupied by extrahelical bases, which prevent hydration. Thus, we have found a single water molecule in the center of the minor groove in the $P2_1$ hexamer, hydrogen bonded to an O2 thymine atom and a sugar O4′. Two of the water molecules we have found in the $P2_1$ structure form a bridge between the external N3 atom of adenine and a neighboring phosphate group, which may contribute to stabilize the Hoogsteen structure.

Extrahelical Residues. An additional feature of the results we have obtained is that some of the bases occupy extrahelical positions. In both crystal structures, there are 20 bases in a duplex conformation and four extrahelical bases. As shown in Figures 2−4, terminal thymine residues enter the minor groove of a neighbor duplex. In all cases, the N3 atom of the extrahelical thymine is hydrogen bonded to the O2 atom of another thymine in the duplex, as shown in detail in Figure 4. One of the adenines in the $P2_1$ structure also shows a similar interaction. In the latter case, the extrahelical arm of the oligonucleotide folds back and adenine enters the minor groove of the same duplex. This adenine occupies a fixed position, its B-factors are similar to those shown by adenines in duplex conformation. Here it is the N6 atom of adenine which is hydrogen bonded to the O2 atom of a thymine in the duplex (Figure 4). Both extrahelical bases are stabilized by van der Waals contacts with neighbor atoms in the minor groove. This form of interaction is reminiscent of minor groove binding drugs. In the present case, such interactions contribute to the stability of the crystal, since they form bridges between neighbor columns of duplexes (Figure 2).

Whereas extrahelical thymine residues occupy a similar position in all cases, each adenine shows a different conformation. In both $P2_1$ and $P6_5$ one of the extra helical adenines is found in an external position on the major groove of the next duplex in the crystal (Figure 3). They are stabilized by van der Waals forces and their position is not strongly fixed: they have the highest B-factors in both $P2_1$ and $P6_5$ structures. In fact, one group of extrahelical AT base in the $P6_5$ case appears to be disordered and could not be localized with certainty in the electron density map. Nevertheless, a peak of electron density was modeled in this region as a water molecule, which turned out to have the lowest B-factor in the whole structure.

End-to-End Interactions. A characteristic feature of the crystal structures found for d(ATATAT) is that, despite the different packing arrangements in space groups $P2_1$ and $P6_5$, the asymmetric unit is very similar in both cases. It contains two duplexes, a hexamer and a tetramer, the latter with two extrahelical AT arms. An additional similarity is found in end-to-end interactions between tetramer and hexamer (Figure 6). Both duplexes form a pseudocontinuous helix, despite the absence of phosphate in the steps between the terminal T of one duplex and the starting A of the next duplex in the column. The value of the twist parameter in this step is about 41° (Table 3), larger that the average value of 34.4° inside the duplex. Such difference should be considered rather small, given the absence of any phosphate binding both duplexes. A more striking feature is the large value of the slide parameter (−4.9 Å), which indicates that the hexamer and tetramer helical axes are shifted. This is most clearly seen in Figure 2b: in the central column of duplexes it is obvious that the tetramer (purple) is displaced upward with respect to the hexamer (green). As a result of this shift, the overlap of bases in neighbor base pairs is modified. In the end-to-end TA step, the base pairs slide laterally (Figure 6, bottom), whereas in the inner base pairs there is a very good overlap of the aromatic rings of thymine and adenine in all TA steps (Figure 6, center). In the end-to-end interaction, only a partial overlap of adenines is found. Such interactions are practically identical in both $P2_1$ and $P6_5$ structures. It is worth comparing such interaction with that found in equivalent situations in the B-form, although there are only a few crystal structures with a terminal A·T base pair (*34*, *35*). The five structures reported by the latter authors are not isomorphous, but all of them contain columns of duplexes with end-to-end interactions equivalent to TA steps as in the case of the Hoogsteen structures reported here. Inspection of the coordinates of all five structures shows that all of them present a similar left-handed twist between terminal base pairs of about −20°. As a result there is a very good overlap of A and T bases of neighbor duplexes. Interestingly, in the structure of the ras_{12} oligonucleotide (*36*), the TG/CA base step between duplexes also shows the same negative twist. In conclusion, end-to-end interactions in pseudo continuous duplexes are quite different in Hoogsteen and B-form DNA crystals. This is probably due to the different dipole moments associated with either Hoogsteen or Watson−Crick base pairs. End-to-end interactions between duplexes may contribute to the stabilization of the Hoogsteen conformation in the crystal.

Structure of d(ATATAT) in Solution. The presence of duplexes in Hoogsteen conformation in d(ATATAT) crystals was quite unexpected. Therefore, we decided to study the behavior in solution of this oligonucleotide by NMR. The ^{1}D spectrum in 1 M KCl at −2 °C is shown in Figure 8. Three imino resonances are observed between 13 and 14 ppm, indicating that all the thymines are base paired. In the NOESY spectra, these imino signals present NOE cross-peaks with adenine H2 protons, which were identified by their relatively long spin−lattice relaxation times. Other nonexchangeable protons were assigned using established techniques for right-handed, double-stranded nucleic acids using DQF-COSY, TOCSY, and 2D NOESY spectra (*37*). Sugar spin systems were identified in the DQF-COSY and TOCSY spectra and connected with their own base and their

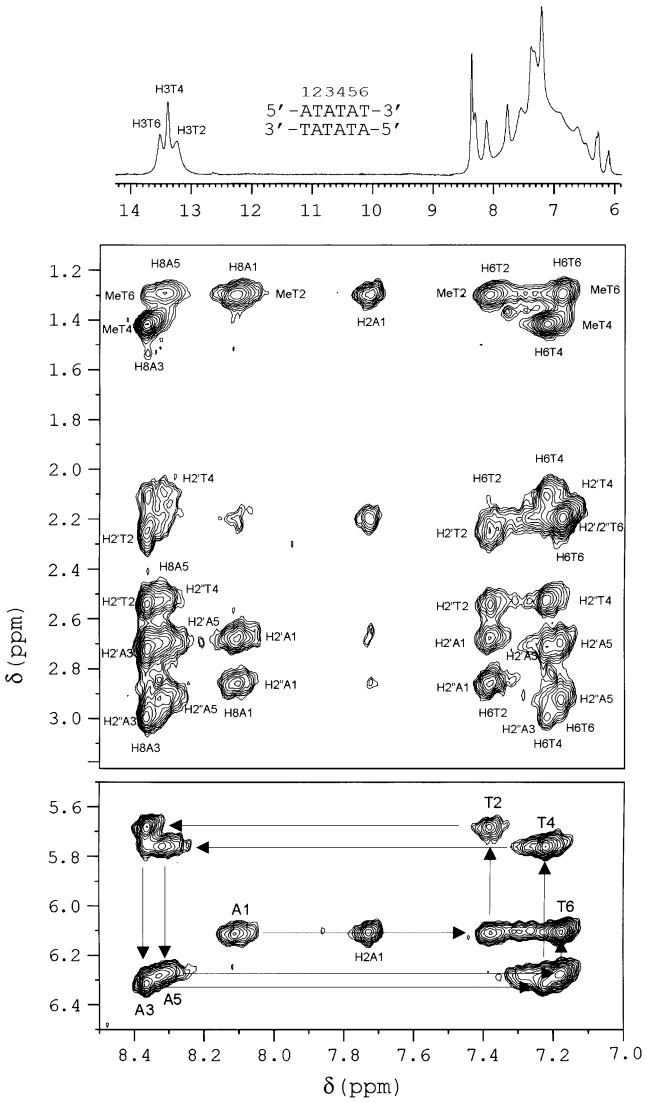

FIGURE 8: Top: Exchangeable proton region of the NMR spectra of $d(ATATAT)_2$ in H_2O (1 M KCl, $T = -2$ °C, pH = 7). Middle and bottom: Two regions of the NOESY spectra in D_2O (τ_m = 200 ms). Cross-peaks are labeled according to the numbering scheme shown on the top panel. Sequential assignment pathways are shown in the H1′−H6/H8 region (bottom panel).

Table 4: Assignment Table of NMR Resonances (ppm)

	NH	NH_2 (1)	NH_2 (2)	H6/H8	H2/Me	H1′	H2′	H2″	H3′	H4′
A1		n.o.	n.o.	8.12	7.75	6.11	2.67	2.85	4.83	4.23
T2	13.49			7.39	1.32	5.68	2.24	2.52	4.92	4.24
A3		7.55	6.47	8.37	7.28	6.30	2.73	2.98	5.02	4.47
T4	13.36			7.22	1.43	5.75	2.08	2.50	4.92	4.25
A5		n.o.	n.o.	8.30	7.28	6.26	2.70	2.90	5.01	4.47
T6	13.26			7.19	1.32	6.10	2.18	2.18	5.09	4.56

n.o.: not observed.

5′-neighbor in the NOESY spectra. The assignment pathways could be followed in the base-H1′, and in the base-H2′/H2″ region, as it is also shown in Figure 8. Resonance positions are given in Table 4.

The intensity of the intraresidual NOE contacts between H1′ and the H8 of the adenines, and between H1′ and H6 of the thymines are very similar, indicating that all glycosidic angles are mainly in an *anti* conformation. Other experimental distances, like sequential H6/H8 or H8-methyl are also characteristic of a right-handed Watson−Crick duplex. Except for the amino resonances of adenines 1 and 5, all the exchangeable DNA protons could be assigned. The NOE patterns observed for the exchangeable protons indicate that all bases are forming Watson−Crick pairs throughout the duplex. Thus, we can conclude that under the experimental conditions we have used, d(ATATAT) in solution has a standard B-form conformation.

A few unusual NOEs involving the H2 of A1 are observed (see Figure 8). Most probably, they are associated with an enhanced fraying effect in the terminal adenines. It should be noticed that this short duplex is not very stable, as confirmed by melting studies (*38*). The imino resonances are not observed at 5 °C, indicating that the duplex is mainly denatured at this temperature. Given that the NOESY experiments were conducted at −2 °C, some mobility effect in the terminal nucleotides is not surprising.

NMR spectra were also acquired for other DNA duplexes of related sequences, like $d(ATATATATAT)_2$ and $d(ATATATAT)_2$. Details on the NMR study of these molecules will be presented elsewhere. Different experimental conditions were tested to simulate the crystal environment, such as adding increasing quantities of MPD. In an attempt to simulate the effect of the extrahelical thymines that interact with the minor groove of the Hoogsteen duplex in the crystal, deuterated thymidines were added to the NMR samples. However, in all cases the NMR data show that these oligonucleotides mainly form standard Watson−Crick duplexes in solution. The UV spectra and melting curves did not show either any feature that might indicate the presence of Hoogsteen DNA (*38*).

Since thus far we have not found a Hoogsteen duplex in solution, it may appear that packing interactions in the crystal could induce the formation of Hoogsteen DNA. Nevertheless, a similar duplex structure is found in crystals with different symmetry, as shown in Figure 2. In any case, the Hoogsteen conformation is an additional conformation available for duplex DNA, at least for alternating AT sequences. In fact, energy calculations and molecular dynamics studies indicate that the Hoogsteen and B-form of duplex DNA have a similar stability (*2*).

DISCUSSION

Previous Studies with AT-Rich DNA. As discussed in the introduction, AT sequences are strongly polymorphic. Thus, it is not surprising that we have found a Hoogsteen structure for d(ATATAT) in crystals, whereas our NMR studies show a Watson−Crick structure in solution. We cannot exclude that the use of other solvents might allow us to detect the Hoogsteen conformation in solution. We should note that the structures we have described in this paper are the only ones available for oligonucleotides that are 100% A·T and contain six or more bases. Longer oligonucleotides that contain A·T sequences have been studied, but their crystal structure is strongly influenced by the presence of guanine bases at the ends of the oligonucleotides which present characteristic intermolecular interactions (*39*).

Hoogsteen DNA in Protein/DNA Complexes. There are very few protein/DNA complexes that have been studied by X-ray crystallography and contain stretches of A·T base pairs. In a detailed study (*20*) of TATA-box binding proteins,

all A·T base pairs show standard Watson−Crick hydrogen bonds, even when an alternating $(TA)_3$ sequence is embedded in the oligonucleotide. In some repressor−operator complexes an oligonucleotide sequence containing $(AT)_4$ was used (*40*, *41*). Unfortunately, the resolution was limited (3.2 Å), and the bases were not well resolved. Another feature of interest is the eventual presence of Hoogsteen base pairs embedded in oligonucleotides with mixed sequence. Given the comparatively low resolution attained in many DNA/protein complexes, it is possible that eventual Hoogsteen base pairs may have not been detected. Only recently (*21*), a Hoogsteen base pair has been found embedded in an undistorted B-DNA. Surprisingly, the adenine in the *syn* conformation has no direct interactions with the protein side chains, although the oligonucleotide is constrained by two α-helices of the protein which have many contacts with neighbor base pairs. This observation indicates that transition to a Hoogsteen conformation does not imply a large energetic penalty for a DNA duplex, as confirmed by theoretical studies (2). On the other hand, the presence of Hoogsteen base pairs will have a strong influence on protein/DNA recognition, given the strong change of properties of both grooves in the duplex.

Hoogsteen pairing has also been detected in some C·G base pairs (*20*). Formation of such base pairs requires protonation of cytosine. One of the hydrogen bonds between the bases is absent, whereas a hydrogen bond between the N2 atom of guanine and one phosphate oxygen is formed, so that on the whole no hydrogen bond is lost. In this case, the formation of one of the Hoogsteen base pairs is clearly stabilized by hydrophobic interaction of the protein with the minor groove of the oligonucleotide. These results (*20*) indicate that C·G base pairs could also be present in Hoogsteen DNA duplexes. In fact, it has been found (*42*) that poly(dCG) at moderate low pH may show a transition to a Hoogsteen form. The pH at which the transition occurs depends on the ionic strength of the buffer used.

Interactions that Favor Hoogsteen DNA. We may wonder why we have obtained a Hoogsteen DNA conformation in the crystals we have described. The structure is independent from the packing arrangement, since it has been found in two different crystalline forms. In both cases, a hydrophobic environment is created in the minor groove which could explain the presence of extrahelical bases (Figures 3 and 4). In fact, isolated base pairs in Hoogsteen conformation (*20*, *21*) are also found in a hydrophobic environment. In some hairpin loops studied by NMR, a thymine base was also present in the minor groove of Hoogsteen A·T pairs (*14*, *16*). Thus, we may expect that Hoogsteen DNA may be found in protein/DNA complexes in which the protein provides a hydrophobic environment for the DNA minor groove.

Another related feature of our crystallization trials is that under many conditions liquid crystals were observed (unpublished results). AT-rich oligonucleotides in the B-form will have a spine of hydration in the minor groove and may form highly hydrated unstable crystals. The crystals we have obtained at a relatively high precipitant concentration (37% MPD) and low temperature (13°) may represent a less soluble, more stable crystal structure. What is obvious from our crystallographic study is that Hoogsteen base pairing is not isolated to a single base pair. Thus, the formation of an antiparallel duplex with all base pairs in Hoogsteen conformation is an accessible state for DNA. Nevertheless, no safe conclusions can be derived from the crystallization conditions used in each particular case. It is well-known that A- and B-form crystals are obtained under very similar conditions: the particular form obtained (either A or B) is mainly determined by the length and sequence of the oligonucleotide that is used.

Table 5: Examples of Repeated Sequences of AT Base Pairs in Different Genomes[a]

organism	repeats found
Mycoplasma genitalium	A_{19}
Mycobacterium leprae	$(AT)_{18}$
Dictyostelium discoideum	A_{206}; $(AT)_{214}$; $(AAT)_{88}$
Saccharomyces cerevisiae	A_{42}; $(AT)_{20}$; $(AAT)_{35}$
Arabidopsis thaliana	A_{47}; $(AT)_{35}$; $(AAT)_{59}$
Homo sapiens	A_{60}; $(AT)_{47}$; $(AAT)_{20}$
Drosophila melanogaster	A_{71}; $(AT)_{40}$; $(AAT)_{84}$; $(AAAT)_{110}$

[a] The repeats have been detected with a program developed by Roset et al. (*66*). Repeats containing C·G base pairs are also found.

An interesting case is the CCCG tetraloop which closes an oligonucleotide hairpin (*15*). In that case, a transition of a C·G base pair from Watson−Crick to Hoogsteen is detected by a moderate change in pH. It is the only example we have found that shows such a transition of a single base pair in the same DNA structure. This example demonstrates that such changes are possible and might be found in other situations.

The Potential Role of Hoogsteen DNA in the Cell Nucleus. Thus far there is no evidence that Hoogsteen DNA may occur "in vivo". However, there are a number of chromatin functions which are poorly understood in which Hoogsteen DNA may play a role, and we will briefly review them. We will emphasize those cases in which an unknown structure for AT-rich sequences has been suggested. Such sequences are common in many organisms (Table 5). For example, the scaffold-associated regions (SARs, also called MARs) are highly AT-rich regions of the genome (*43*−*45*). Specific proteins are apparently found in those regions, but the structure of their complexes with DNA is unknown.

Centromere DNAs always have AT-rich regions (*46*). They are usually embedded in longer heterochromatic regions, also AT-rich. In particular, yeast centromeres have a central CDEII element about 85-bp long and 93% AT (*47*). Drosophila centromeres contain long (AATAT) repeats, accompanied by other types of sequences (*48*). In Arabidopsis (*49*, *50*) and human α-satellite (*51*) the centromeric sequences are more complex, but are always enriched in AT base pairs. Nevertheless, it is not clear which features of centromere DNA determine its unique function (*52*). It appears that a high percentage of AT and some undetermined features of sequence are required for function. An alternative view is that only specific proteins are required for centromere function; DNA sequence does not play a role (*53*). In fact, DNA sequence and protein structure may have evolved together in centromeres (*54*). The arguments in favor of such contrasting views have been recently reviewed by Lamb and Birchler (*55*).

There are many other DNA and chromatin features that involve AT-rich regions in which Hoogsteen DNA might play a role, such as simple sequence repeats (*56*), replication

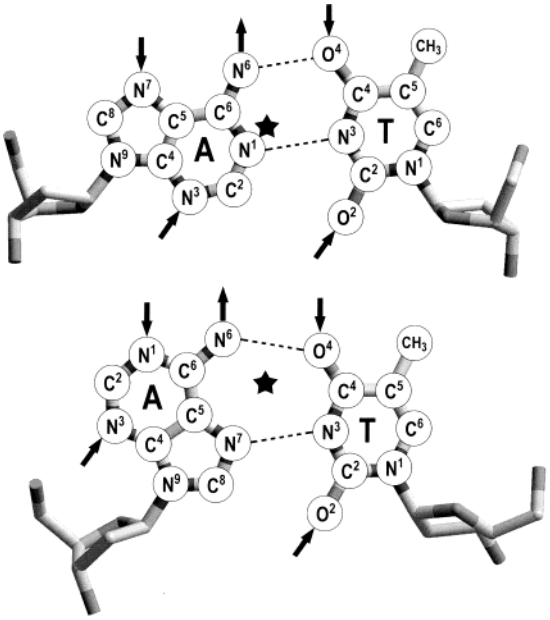

FIGURE 9: Comparison of Watson−Crick (WC) (top) and Hoogsteen (bottom) base pairs in duplex conformation. In the picture, the minor groove is facing downward. The position of the helical axis in each case is indicated by a star. Hydrogen bond donors and acceptors are indicated by arrows. Note that the minor groove is narrower in the Hoogsteen case and has lost a hydrogen bond acceptor atom. The major groove has a similar appearance, but in the Hoogsteen case an additional external N3 atom is present. It is placed close to the phosphodiester backbone and may form water bridges with a phosphate group. An example is apparent in Figure 7. Such bridges contribute to the stability of the Hoogsteen base pair. The conformation of the glycosidic angle is *syn* in the Hoogsteen adenine base and *anti* in all other cases.

origins (*57*, *58*), fragile site genesis (*59*), bending (*60*), HMG interaction sites (*61*), etc. Unusual structures present in repeated DNA may trigger gene silencing (*62*). TATA boxes may become silent and not recognized by TATA-box binding proteins (TBPs) when found in the Hoogsteen conformation. We should also note that the biological role of satellite repeats remains unknown (*63*); many of them are AT-rich (*64*), with the extreme case of some crab DNAs which contain over 90% of an alternating AT sequence (*65*).

CONCLUSION

A summary of the differences between B-form and Hoogsteen DNA is presented in Figure 9. Hoogsteen DNA provides a strong structural message: hydrogen bonding is altered, the minor groove changes from hydrophilic to hydrophobic. Its stability is similar to B-form DNA, although it may be more rigid (2). It is an optimal target for proteins to recognize AT-rich regions. New approaches are required to determine under what conditions this unexpected conformation of the double helix plays a role in the cell, which may have gone undetected so far. In particular, further high-resolution studies of AT-rich DNA and its protein complexes by NMR, X-ray crystallography and other methods should be undertaken.

ACKNOWLEDGMENT

We are most thankful to Drs. F. Azorín and M. Chiva for discussion and suggestions.

REFERENCES

1. Abrescia, N. G. A., Thompson, A., Huynh-Dinh, T., and Subirana, J. A. (2002) Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc. Natl. Acad. Sci. U.S.A. 99*, 2806−2811.
2. Cubero, E., Abrescia, N. G. A., Subirana, J. A., Luque, F. J., and Orozco, M. (2003) Theoretical Study of a new DNA structure: the antiparallel Hoogsteen duplex. *J. Am. Chem. Soc. 125*, 14603−14612.
3. Watson, J. D., and Crick, F. H. C. (1953) A Structure for Deoxyribose Nucleic Acid. *Nature 171*, 737−738.
4. Hoogsteen, K. (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr. 12*, 822−823.
5. Voet, D., and Rich, A. (1970) The crystal structures of purines, pyrimidines and their intermolecular complexes. *Prog. Nucl. Acid Res., Mol. Biol. 10*, 183−265.
6. Leslie, A. G. W., Arnott, S., Chandrasekaran, R., and Ratliff, R. L. (1980) Polymorphism of DNA double helices. *J. Mol. Biol. 143*, 49−72.
7. Viswamitra, M. A., Shakked, Z., Jones, P. G., Sheldrick, G. M., Salisbury, S. A., and Kennard, O. (1982) Structure of the deoxytetranucleotide d-pApTpApT and a sequence-dependent model for poly(dA-dT). *Biopolymers 21*, 513−533.
8. Davies, D. R., and Baldwin, R. L. (1963) X-ray studies of two synthetic DNA copolymers. *J. Mol. Biol. 6*, 251−255.
9. Kypr, J., Chládková, J., Arnold, L., Sági, J., Szemzö, A., and Vorlícková, M. (1996) The Unusual X-form DNA in oligodeoxynucleotides: dependence of stability on the base sequence and length. *J. Biomol. Struct. Dyn. 13*, 999−1006.
10. Rippe, K., and Jovin, T. M. (1992) Parallel-stranded duplex DNA. *Methods Enzymol. 211*, 199−220.
11. Parvathy, V. R., Bhaumik, S. R., Chary, K. V. R., Govil, G., Liu, K., Howard, F. B., and Miles, H. T. (2002) NMR structure of a parallel-stranded DNA duplex at atomic resolution. *Nucleic Acids Res. 30*, 1500−1511.
12. Raghunathan, G., Miles, H. T., and Sasisekharan, V. (1994) Parallel nucleic acid helices with Hoogsteen base pairing: symmetry and structure. *Biopolymers 34*, 1573−1581.
13. Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1989) Effects of base sequence on the loop folding in DNA hairpins. *Biochemistry 28*, 7491−7498.
14. Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1991) The three-dimensional structure of a DNA hairpin in solution. Two-dimensional NMR studies and structural analysis of d(ATCCTATTTATAGGAT). *Eur. J. Biochem. 201*, 33−51.
15. Van Dongen, M. J. P., Wijmenga, S. S., van der Marel, G. A., van Boom, J. H., and Hilbers, C. W. (1996) The transition from a neutral-pH double helix to a low-pH triple helix induces a conformational switch in the CCCG tetraloop closing a Watson−Crick stem. *J. Mol. Biol. 263*, 715−729.
16. Escaja, N., Gómez-Pinto, I., Rico, M., Pedroso, E., and González, C. (2003) Structure and stability of small DNA dumbbells with Watson−Crick and Hoogsteen base pairs. *ChemBioChem 4*, 101−110.
17. Hakoshima, T., Fukui, T., Ikehara, M., and Tomita, K.-I. (1981) Molecular structure of a double helix that has non-Watson−Crick type base pairing formed by 2-substituted poly(A) and poly(U). *Proc. Natl. Acad. Sci. U.S.A. 78*, 7309−7313.
18. Isaksoon, J., Zamaratski, E., Maltseva, T. V., Agback, P., Kumar, A., and Chattopadhyaya, J. (2001) The first example of a Hoogsteen base paired DNA duplex in dynamic equilibrium with a Watson−Crick base paired duplex − A structural (NMR), kinetic and thermodynamic study. *J. Biomol. Struct., Dyn. 18*, 783−806.
19. 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. U.S.A. 86*, 3006−3010.
20. 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.
21. Aishima, J., Gitti, R. K., Noah, J. E., Gan, H. H., Schlick, T., and Wolberger. C. (2002) A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res. 30*, 5244−5252.

22. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) Triplex DNA structure. *Annu. Rev. Biochem. 64*, 65−95.
23. Radwan, M. M., and Wilson, H. R. (1982) Fibre and molecular structure of thymidylyl-3′,5′-deoxyadenosine. *Int. J. Biol. Macromol. 4*, 145−149.
24. Navaza, J., and Saludjian, P. (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol. 276*, 581−594.
25. Abrescia, N. G. A., and Subirana, J. A. (2002) When pseudosymmetry and merohedral twinning come across: the case of the d(ApTpApTpApT) oligonucleotide in a hexagonal lattice. *Acta Crystallogr., Sect. D 58*, 2205−2208.
26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gross, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D 54*, 905−921.
27. Brünger, A. T. (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*, 472−475.
28. Bax, A., and Davies, D. J. (1985) MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson. 65*, 355−360.
29. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton−proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun. 95*, 1−6.
30. Plateau, P., and Güeron, M. (1982) Exchangeable proton NMR without baseline distortions, using new strong pulse sequences. *J. Am. Chem. Soc. 104*, 7310−7311.
31. Bartels, C., Xia, T., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR 6*, 1−10.
32. Yuan, H., Quintana, J., and Dickerson, R. E. (1992) Alternative structures for alternating poly(dA-dT) tracts: the structure of the B-DNA decamer C-G-A-T-A-T-A-T-C-G. *Biochemistry 31*, 8009−8021.
33. Schneider, B., Neidle, S., and Berman, H. M. (1997) Conformations of the sugar−phosphate backbone in helical DNA crystal structures. *Biopolymers 42*, 113−124.
34. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. S., and Shakked, Z. (1998) Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl. Acad. Sci. U.S.A. 95*, 15194−15199.
35. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D., and Shakked, Z. (2001) DNA bending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A. 98*, 8490−8495.
36. Timsit, Y., Vilbois, E., and Moras, D. (1991) Base-pairing shift in the major groove of (CA)n tracts by B-DNA crystal structures. *Nature 354*, 167−170.
37. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
38. De Luchi, D., Gouyette, C., and Subirana, J. A. (2003) The influence of size on the thermal stability of oligonucleotides. The case of AT sequences. *Anal. Biochem. 322*, 279−282.
39. Subirana, J. A., and Abrescia, N. G. A. (2000) Extra-helical guanine interactions in DNA. *Biophys. Chem. 86*, 179−189.
40. Anderson, J. E., Ptashne, M., and Harrison, S. C. (1987) Structure of the repressor-operator complex of bacteriophage 434. *Nature 326*, 846−852.
41. Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S. C. (1988). Structure of a phage 434 Cro/DNA complex. *Nature*, *335*, 789−795.
42. Segers-Nolten, G. M. J., Sijtsema, N. M., and Otto, C. (1997) Evidence for Hoogsteen GC base pairs in the proton-induced transition from right-handed to left-handed poly(dG-dC)•poly(dG-dC). *Biochemistry, 36*, 13241−13247.
43. Hart, C. M., and Laemmli, U. K. (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev. 8*, 519−525.
44. Singh, G. B., Kramer, J. A., and Krawetz, S. A. (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res. 25*, 1419−1425.
45. Subramanian, S., Mishra, R. K., and Singh, L. (2003) Genome-wide analysis of Bkm sequences (GATA repeats): predominant association with sex chromosomes and potential role in higher order chromatin organization and function. *Bioinformatics 19*, 681−685.
46. Choo, K. H. A. (1997) *The Centromere*, Oxford University Press, Oxford.
47. Fleig, U., Beinhauer, J. D., and Hegemann, J. H. (1995) Functional selection for the centromere DNA from yeast chromosome VIII. *Nucleic Acids Res. 23*, 922−924.
48. Sun, X., Le, H., Wahlstrom, J. M., and Karpen, G. H. (2003) Sequence analysis of a functional Drosophila centromere. *Genome Res. 13*, 182−194.
49. Hall, S. E., Kettler, G., and Preuss, D. (2003) Centromere satellites from Arabidopsis populations: maintenance of conserved and variable domains. *Genome Res. 13*, 195−205.
50. Heslop-Harrison, J. S., Brandes, A., and Schwarzacher, T. (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of Arabidopsis species. *Chromosome Res. 11*, 241−253.
51. Willard, H. F., and Waye, J. S. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet. 3*, 192−198.
52. Henikoff, S., Ahmad, K., and Malik, H. S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science 293*, 1098−1102.
53. Cleveland, D. W., Mao, Y., and Sullivan, K. F. (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell 112*, 407−421.
54. Malik, H. S., and Henikoff, S. (2002) Conflict begets complexity: the evolution of centromeres. *Curr. Opin. Genet., Dev. 12*, 711−718.
55. Lamb, J. C., and Birchler, J. A. (2003) The role of DNA sequence in centromere formation. *Genome Biol. 4*, 214−217.
56. Bacon, A. L., Dunlop, M. G., and Farrington, S. M. (2001) Hypermutability at a poly(A/T) tract in the human germline. *Nucleic Acids Res. 29*, 4405−4413.
57. Gilbert, D. M. (2001) Making sense of eukaryotic DNA replication origins. *Science 294*, 96−100.
58. Lee, J.-K., Moon, K.-Y., Jiang, Y., and Hurwitz, J. (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc. Natl. Acad. Sci. U.S.A. 98*, 13589−13594.
59. Hewett, D. R., Handt, O., Hobson, L., Mangelsdorf, M., Eyre, H. J., Baker, E., Sutherland, G. R., Schuffenhauer, S., Mao, J., and Richards, R. I. (1998) FRA 10B structure reveals common elements in repeat expansion and chromosomal fragile site genesis. *Mol. Cell 1*, 773−781.
60. Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982) Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci. U.S.A. 79*, 7664−7668.
61. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem. Sci. 26*, 167−174.
62. Dernburg, A. F., and Karpen, G. H. (2002) A chromosome RNAissance. *Cell 111*, 159−162.
63. Csink, A. K., and Henikoff, S. (1998) Something from nothing: the evolution and utility of satellite repeats. *Trends Genet. 14*, 200−204.
64. Brutlag, D. L. (1980) Molecular arrangement and evolution of heterochromatic DNA. *Annu. Rev. Genet. 14*, 121−144.
65. Gray, D. M., and Skinner, D. M. (1974) A circular dichroism study of the primary structures of three crab satellite DNA's rich in A:T base pairs. *Biopolymers 13*, 843−852.
66. Roset, R., Subirana, J. A., and Messeguer, X. (2003) MREPATT: detection and analysis of exact consecutive repeats in genomic sequences. *Bioinformatics 19*, 2475−2476.

BI0355140