Published online September 27, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 17 ***5087–5095***
doi:10.1093/nar/gkh844

DNA adopts normal B-form upon incorporation of highly fluorescent DNA base analogue tC: NMR structure and UV-Vis spectroscopy characterization

K. Cecilia Engman, Peter Sandin[1], Sadie Osborne[2], Tom Brown[2], Martin Billeter, Per Lincoln[1], Bengt Nordén[1], Bo Albinsson[1] and L. Marcus Wilhelmsson[1,*]

Department of Chemistry, Göteborg University, PO Box 462, SE-40530 Gothenburg, Sweden, [1]Physical Chemistry Section, Department of Chemistry and Bioscience, Chalmers University of Technology, SE-41296 Gothenburg, Sweden and [2]School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received June 29, 2004; Revised and Accepted September 3, 2004 PDB accession no. 1TUQ

ABSTRACT

The influence of the highly fluorescent tricyclic cytosine base analogue (tC) on duplex DNA conformation is investigated. The duplex properties are characterized by absorbance and circular dichroism (CD) for all combinations of neighbouring bases to tC, and an NMR structure is determined for one tC-containing sequence. For the oligonucleotides with one tC incorporated instead of cytosine, the melting temperature is increased on average by 2.7°C above that for the unmodified ones. CD spectra are practically identical for modified and unmodified sequences, indicating an unperturbed B-DNA conformation. The NMR structure determination of the self-complementary sequence 5′-CTC(tC)ACGTGGAG shows a DNA conformation consistent with B-form for the whole duplex. The root-mean-square distance for the nucleotides of the eight central base pairs between the 10 structures with lowest CYANA target functions and a mean structure is 0.45 ± 0.17 Å. The NMR data confirm correct base pairing for tC by the observation of both intrastrand and interstrand imino proton NOEs. Altogether, this suggests that tC works well as a cytosine analogue, i.e. it is situated in the base stack, forming hydrogen bonds with G in the complementary strand, without distorting the DNA backbone conformation. This first example of an artificial, highly fluorescent DNA base that does not perturb the DNA conformation could have valuable applications for the study of the structure and dynamics of nucleic acid systems.

INTRODUCTION

There has recently been an enormous increase in the application of fluorescence methods for studying and detecting biochemical macromolecules. Among the advantages of the use of fluorescence methods are the straightforwardness with which samples can be studied and the small amount of sample needed, even down to single molecules (1). In order to investigate nucleic acid containing systems using fluorescence techniques, fluorescent labels have to be introduced, as the intrinsic quantum yield of the naturally occurring bases is very low. These labels, for example, can be fluorescent dyes that are covalently linked to the DNA backbone or fluorescent base analogues replacing a normal base. Because of their natural linkage to DNA and site-specific incorporation into DNA, fluorescent base analogues have a great advantage compared to covalently attached fluorescent dyes and have therefore been receiving increasing attention [for review see (2)]. Most importantly, when probing the structure and dynamics of nucleic acids by using an internal optical probe, it is desirable that the probe does not introduce any perturbations into the structure.

Among the crucial properties required for fluorescent base analogues are their ability to form specific base pairs with the naturally occurring complementary base, minimal influence on the overall DNA structure and sufficient fluorescent quantum yield, not only when free in solution but also when studied individually in single- and double-stranded DNA as well as in complex with DNA-binding proteins. The most utilized fluorescent base analogue to date has been 2-aminopurine (2-AP) (3–13), which forms stable base pairs with thymine but can also form fairly stable base pairs with cytosine (14). Moreover, the high intrinsic fluorescence quantum yield of 2-AP ($\phi_f = 0.68$) is reduced dramatically (100-fold) when incorporated into nucleic acids (3). These properties make 2-AP less suited as a probe for molecular dynamics and DNA–protein interaction using fluorescence anisotropy and energy transfer, for example.

Other commercially available fluorescent base analogues are the pteridines [for review see (15)] 3-methylisoxanthopterin (3-MI) and 6-methylisoxanthopterin (6-MI) (16–19). Like 2-AP, these guanine analogues have high fluorescence

*To whom correspondence should be addressed. Tel: +46 (0) 31 7723051; Fax: +46 (0) 31 7723858; Email: mawi@chembio.chalmers.se

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

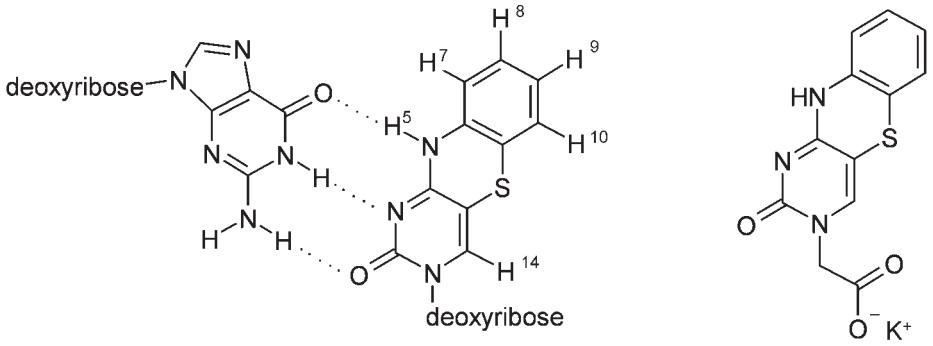

Figure 1. Structure of G–tC base pair to the left (atom numbering as used for NMR assignments) and to the right the potassium salt of 1,3-diaza-2-oxophenothiazin-3-yl acetic acid (KtC).

quantum yields when free in solution (0.88 and 0.70, respectively) but display a highly reduced (up to 25-fold) emission intensity upon incorporation into DNA (17). Concerns for the base pairing ability and a larger base flip tendency of 3-MI due to the bulky methyl group positioned in its hydrogen-bonding motif have been confirmed in duplex melting temperature studies, where it was shown that 3-MI reduced the thermal stability and is approximately equivalent to a single base pair mismatch (17). This kind of lowered thermal stability has also been reported for a new promising family of fluorescent base analogues, the benzoquinazolines (20–22).

Recently we reported on the interesting fluorescence properties of a novel fluorescent base analogue, 1,3-diaza-2-oxophenothiazine, tC (Figure 1) (23). It has been previously reported that this tricyclic cytosine analogue discriminates well between G and A targets (24), and in measurements on a few sequences it has also been shown to increase the thermal stability of DNA–RNA (24), DNA–peptide nucleic acid (PNA) (25) and PNA–PNA duplexes (25), probably as a result of improved base stacking. We have shown that an acetic acid derivative of tC (KtC; Figure 1) is in its normal base pairing form between pH 4 and 12 and exhibits a high fluorescence quantum yield ($\phi_f \sim 0.2$) (26) when free in solution. The fluorescence quantum yield is essentially unaffected by incorporation into a PNA single strand, and it is not changed by further annealing of a complementary DNA single strand to the PNA strand (23). We have also established that the fluorescence quantum yield of tC is virtually unaffected when incorporated into DNA single strands with all possible neighbouring bases and when a complementary DNA strand is hybridized to the tC-modified single-stranded oligomer (P. Sandin, B. Albinsson, P. Lincoln, T. Brown, B. Nordén and L. M. Wilhelmsson, manuscript in preparation).

The tC base has an absorption band with maximum absorption at ~375 nm (26). This transition is well separated from both normal DNA bases and amino acids and can be used for selective excitation of the tC chromophore. Furthermore, as an effect of this transition the tC base can be used, with advantage, as a fluorescence resonance energy transfer (FRET) acceptor in a pair with tryptophan. We have previously shown that this absorption band is the result of a transition moment directed ~35° from the long axis of tC (26). Furthermore, we have reported that tC can serve as an excellent FRET donor in a pair with rhodamine as acceptor (23). The rigid and well-defined geometry expected for tC stacked within the duplex will be a major advantage for future quantitative FRET experiments accurately measuring distances within molecular systems, for fluorescence anisotropy measurements and also for investigating local conformational changes in DNA base stacking with circular dichroism (CD).

To establish that the fluorescent base analogue tC functions well as a probe in DNA with minimal disturbance to the local conformation, overall structure and duplex stability, we report here on the influence on DNA structure when a normal cytosine is replaced by tC. CD spectra show that incorporation of tC will leave the DNA virtually unaffected in its B-form. Also, DNA duplex melting temperature measurements show that tC, in general, slightly increases the thermal stability of the duplex. Furthermore, we use NMR spectroscopy to determine the structure of a double-stranded DNA oligomer containing one tC in each strand. From NMR data, we confirm that a self-complementary DNA dodecamer, with tC in position four, adopts a B-form DNA conformation.

MATERIALS AND METHODS

Oligonucleotide synthesis

Oligonucleotides containing tC were synthesized on an ABI 394 DNA/RNA synthesizer on the 0.2 μmol scale by standard automated solid-phase methods using β-cyanoethyl phosphoramidites. Cleavage from the solid support and deprotection were carried out in the normal manner using concentrated aqueous ammonia at 55°C for 5 h. Oligonucleotides were purified by reverse-phase high performance liquid chromatography on a C8 (octyl) column, eluting with a gradient of 0.1 M ammonium acetate to 25% acetonitrile in 0.1 M ammonium acetate buffer (27). Desalting was carried out by Sephadex gel filtration (Nap 10, Pharmacia). Synthesis of the 5′-O-DMT protected β-cyanoethyl phosphoramidite derivative of the 2-deoxyribonucleoside of 1,3-diaza-2-oxophenothiazine (tC) used in the synthesis above will be presented elsewhere (P. Sandin, B. Albinsson, P. Lincoln, T. Brown, B. Nordén and L. M. Wilhelmsson, manuscript in preparation). Unmodified oligonucleotides, not containing tC, were purchased from Eurogentec.

Concentration determination

The standard buffer in all measurements, unless stated otherwise, was a sodium phosphate buffer, pH 7.5, with a total sodium concentration of 50 mM. The concentrations of the oligonucleotides (sequences as listed in Table 1) were

Table 1. DNA sequences and their duplex melting temperatures for oligonucleotides containing a tC (T_m^{tC}) and only normal DNA bases (T_m)

DNA sequence[a]	Neighbouring bases = name of oligomer	T_m^{tC} (°C)[b]	T_m (°C)[b] normal sequence	(ΔT_m/tC) (°C)[c]
5′-CGCA**T(tC)**ATCG-3′	TA	45	41	+4
5′-CGCA**A(tC)**ATCG-3′	AA	42	41	+1
5′-CGCA**T(tC)**TTCG-3′	TT	44	39	+5
5′-CGCA**G(tC)**ATCG-3′	GA	44	45	−1
5′-CGCA**C(tC)**TTCG-3′	CT	49	44	+5
5′-CGCA**T(tC)**GTCG-3′	TG	50	45	+5
5′-CGCA**C(tC)**ATCG-3′	CA	52	46	+6
5′-CGCA**G(tC)**GTCG-3′	GG	50	50	+/−0
5′-CGCA**C(tC)**CTCG-3′	CC	53	48	+5
5′-CGCA**G(tC)**CTCG-3′	GC	46	49	−3
5′-CT**C(tC)**ACGTGGAG-3′	CA(NMR)	57	49	+4

[a]In the DNA oligomers containing only normal DNA bases, tC should be replaced by C.
[b]T_m values were determined from the temperature at the peak height maximum of the derivative of the melting curves. Measurements were performed in a phosphate buffer (50 mM Na^+, pH 7.5) at a duplex concentration of 5 μM.
[c]Effects on duplex melting temperature per incorporated tC.

determined by ultraviolet (UV) absorption measurements at 260 nm. The extinction coefficients for the modified oligonucleotides were approximated by the linear combination of the extinction coefficients of the natural nucleotides and the extinction coefficient of the potassium salt of 1,3-diaza-2-oxophenothiazin-3-yl acetic acid (KtC) (Figure 1). To account for the base stacking interactions, this linear combination was multiplied by 0.9 to give the final extinction coefficients for the oligomers. For the sequence used in the NMR experiments, the multiplication factor was 0.8 because of the self-complementarity. The individual extinction coefficients at 260 nm used were ε_T = 9300 M^{-1} cm^{-1}, ε_C = 7400 M^{-1} cm^{-1}, ε_G = 11 800 M^{-1} cm^{-1}, ε_A = 15 300 M^{-1} cm^{-1} (28) and ε_{KtC} = 13 500 M^{-1} cm^{-1}. This resulted in the following extinction coefficient at 260 nm for the modified oligonucleotides: ε_{TA} =97650 M^{-1} cm^{-1}, ε_{AA} = 103050 M^{-1} cm^{-1}, ε_{TT} =92250 M^{-1} cm^{-1}, ε_{GA} = 99 900 M^{-1} cm^{-1}, ε_{CT} = 90 540 M^{-1} cm^{-1}, ε_{TG} =94 500 M^{-1} cm^{-1}, ε_{CA} = 95 940 M^{-1} cm^{-1}, ε_{GG} = 96 750 M^{-1} cm^{-1}, ε_{CC} = 88 830 M^{-1} cm^{-1}, ε_{GC} = 92 790 M^{-1} cm^{-1} and $\varepsilon_{CA(NMR)}$ = 105 680 M^{-1} cm^{-1} (indices of εs as in Table 1). The extinction coefficients used for the purchased, unmodified oligonucleotides were those given by Eurogentec except for the CA(NMR) sequence, the extinction coefficient of which was multiplied by 0.9 due to its self-complementarity.

UV-Vis spectroscopy and circular dichroism

Equimolar amounts of the single-strand oligonucleotides in buffer were mixed at room temperature. The samples were heated to 90°C and thereafter annealed by slow cooling to 25°C.

The melting temperatures were measured on a Varian Cary 4B spectrophotometer equipped with a programmable multicell temperature block. The samples were heated from 5 to 90°C at a maximum rate of 0.5°C min^{-1}. The temperature was held at 90°C for 5 min whereupon the samples were cooled to 5°C at the same rate as before. Absorption at 260 nm was measured with a temperature interval of 1°C.

CD spectra were recorded on a Jasco J-720 spectropolarimeter at 25°C using a 5 mm quartz cell. The spectra of oligonucleotides containing tC were recorded between 200 and 550 nm, and the spectra of oligonucleotides not containing tC were recorded between 200 and 350 nm. All spectra were corrected for background contributions.

NMR sample preparation, spectra recording, processing and analysis

The synthesized oligonucleotide 5′-CTC(tC)ACGTGGAG [CA(NMR)] was lyophilized twice and dissolved in 50 mM potassium phosphate buffer, pH 6.0. Before recording the NMR spectra, the sample was heated to a temperature at which the imino signals were not visible and then slowly cooled to 15°C (the temperature at which all NMR experiments were performed).

NMR spectra were recorded on a Varian Inova 800 MHz spectrometer equipped with a triple resonance probe. The following spectra were recorded in D_2O: DQF-COSY (16 transients, 512 complex points in t1), two TOCSYs (16 transients, 512 complex points in t1, 50 and 80 ms mixing time) and two NOESYs (16 transients, 512 complex points in t1, 100 and 250 ms mixing time). A weak presaturation pulse was used for water suppression. Additional NOESY spectra (16 transients, 512 complex points in t1, 100 and 250 ms mixing time) were recorded in a mixture of 90% ^1H_2O:10% D_2O using the Watergate sequence for water suppression. A ^1H–^{31}P spectrum (256 transients and 256 complex points in t1) was recorded on a Varian Inova 600 MHz spectrometer equipped with a triple resonance probe.

All data were processed with the program NMRPipe (29). The data were typically zero filled once in each dimension and Fourier transformed. A polynomial baseline was used in the time domain. XEASY (30) was used for analysis, assignment and peak integration. The sequence was assigned by DQF-COSY, TOCSY and NOESY using standard procedures (31). J-couplings were measured in DQF-COSY and ^{31}P–^1H using the program XEASY.

Characterization of local structure

The ribose ring puckers were characterized from scalar coupling constants and NOE intensity comparisons. Two independent analyses can be performed to differentiate between C2′-endo and C3′-endo conformations (32). For the C2′-endo form, the $^3J_{HH}$ scalar couplings between H2″–H3′ and H3′-H4′ are both small; the corresponding cross peaks in the DQF-COSY may become unobservable. In contrast, these couplings are large for the C3′-endo conformation. The alternative basis for determining the sugar conformation is the

examination of NOE intensities for the H2″–H4′ and H2′/H2″–H(6/8/14) cross peaks. A weak H2″–H4′ NOE (d ~ 4 Å) and a shorter distance between H2′ and H(6/8/14) than between H2″ and H(6/8/14) also reflects a C2′-endo conformation.

Information about the torsion angles β and ε can be obtained from the observation of J-couplings between the phosphorus and the H5′ and H5″ nuclei and between the phosphorus and the H3′ nuclei, respectively, in the ^{31}P-^{1}H spectrum. Two very small P-H5′/H5″ couplings (<5 Hz) indicate a *trans* conformation as in B-DNA; small but measurable differences between the two couplings may reflect local deviations from ideal B-DNA. Similarly, a small P–H3′ coupling indicates a *trans* conformation of the β angle as in B-DNA. Stronger NOEs between H3′ and H5′ than between H3′ and H5″ in combination with small or undetectable couplings between H4′ and both H5′ and H5″ suggest a *gauche*+ conformation for the γ angle. Finally, observation of weaker intra-residual NOEs between the aromatic proton H(6/8/14) and the H1′ proton than between H(6/8/14) and the H2′ proton can be taken as evidence for *anti* conformation around all glycosidic bonds [in *syn* conformation, the H(6/8) to H1′ distance is short, while H(6/8) to H2′/H2″ is long].

Experimental restraints for structure calculation

In the structure calculation, the ribose rings were locked in the C2′-endo conformation. This and direct constraints on the angles β, γ and ε (see above) leave only the angles α and ζ unrestricted during the calculation. In general, NOEs in the crowded sugar regions were primarily assigned to intra-ribose connectivities and then considered as already used and therefore excluded from the structure calculations.

The maximum intensities for the peaks from the NOESY spectra in D_2O were translated into upper distance limits using the procedure 'calibrate' included in the CYANA package (33). A function of the type NOE = c/d^6, where c is a calibration constant and d is the distance, was used. The calibration constant was selected to be as tight as possible while avoiding systematic violations. For each of the two NOESYs, a different calibration constant was determined. Imino and amino protons were not treated differently, since the small number of NOEs involving these types of protons (17 in total) represents a minor contribution to the final set of constraints. Any possible exchange process with solvent would reduce the NOE, leading to a longer upper distance limit and thus no error would be introduced. For the structure calculation with CYANA, a total of 254 NOE-derived distance restraints were used. In addition, a total of 92 angle restraints derived from J-couplings according to Wijmenga *et al*. (34) were included, together with hydrogen bond restraints for base pairs 3–10. The upper limits for hydrogen bond restraints were set to 3.25 and 2.25 Å for the N to O/N and H to O/N distances, respectively. Corresponding lower limits were 3.15 and 2.15 Å. From a total of 100 random start structures, the 10 conformers with lowest target functions were analysed.

RESULTS

UV melting experiments

The 11 sequences of modified and natural DNA single strands studied in this work are presented in Table 1: a group of 10 decamer sequences and one self-complementary dodecamer, the latter designed for the NMR experiments. The only variation made between the first 10 sequences is the choice of the two bases surrounding the site where tC is incorporated. Table 1 summarizes the duplex melting temperatures of the different modified (T_m^{tC}) and natural (T_m) oligonucleotides and the differences in melting temperature per modification (ΔT_m/tC) (melting profiles in Supplementary Material S1). As can be seen, the 10 natural sequences can be divided, not unexpectedly, into three distinct classes of sequences according to T_m: those containing no G:C (TA, AA and TT), one G:C (GA, CT, TG and CA) and two G:C (GG, CC and GC) base pairs surrounding the site in the sequence where tC is incorporated, with T_ms of ~40, 45 and 49°C, respectively. The incorporation of tC instead of a C gives, with a few exceptions (GA and GC), a higher or equally high T_m suggesting that tC, generally, promotes a stabilization of the duplexes. For the group of 10 sequences, the average increase in melting temperature resulting from a C to tC substitution is +2.7°C, and the increase in T_m for the CA(NMR) sequence is approximately the same (+4°C per incorporated tC).

Circular dichroism

CD spectra of the duplexes containing a tC-modified single strand and its complementary strand are shown in Figure 2 (individual CD spectra in Supplementary Material S2). As can be seen in the figure, there are positive bands at ~375, 275 and below 210 nm as well as negative bands at ~315 and 250 nm. There is also a conspicuous spectral contribution at 225 nm, which for some duplexes shows a slightly positive and for some duplexes a slightly negative CD. The general appearance of the CD signal between 300 and 200 nm resembles that of a normal B-form DNA, which is normally characterized by a positive band centred at 275 nm, a negative band at 240 nm, a band which could be positive or negative at 220 nm and just below that a narrow negative peak followed by a large positive peak at 190–180 nm (35). The CD bands at longer

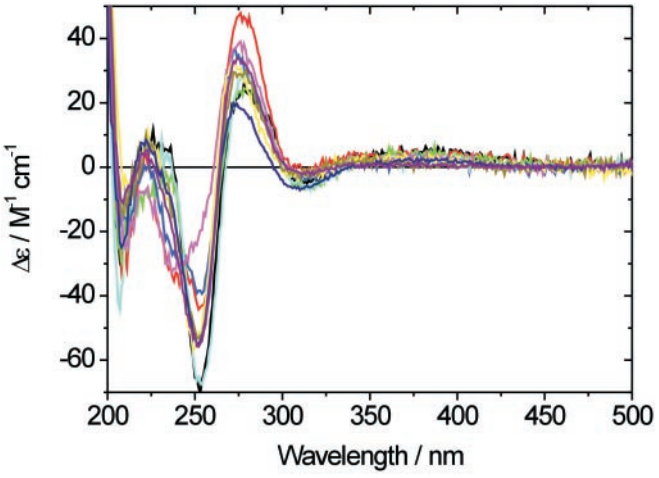

Figure 2. CD spectra of duplexes of all 10 different tC-modified DNA-oligomers: TA (yellow), AA (navy), TT (purple), GA (black), CT (red), TG (dark yellow), CA (blue), GG (cyan), CC (magenta), GC (green), hybridized to their complementary natural DNA strand. Measurements were performed at a duplex concentration of 2.5 μM in a phosphate buffer (50 mM Na^+, pH 7.5) at 25°C. M on Δε-axis refers to total concentration of single strands (5 μM).

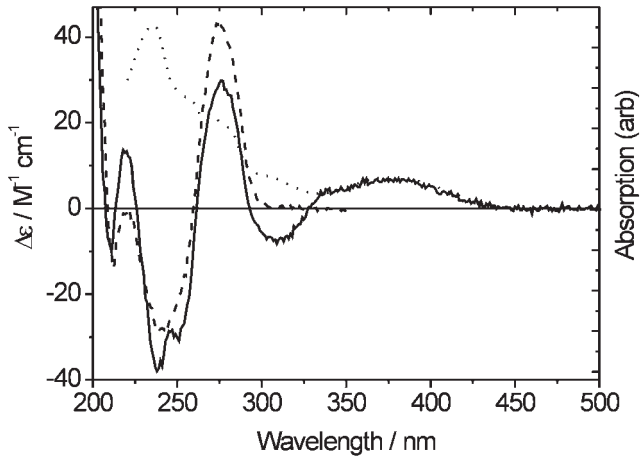

Figure 3. CD of the duplex of CA(NMR). DNA oligomer containing one tC/ strand (solid) and only normal DNA bases (dashed). Also included is an absorption spectrum of KtC (dotted). Measurements were performed in a phosphate buffer (50 mM Na^+, pH 7.5) at 25°C. The concentration of the duplex in CD experiments was 2.35 μM. M on Δε-axis refers to total concentration of single strands (4.7 μM).

wavelengths than 300 nm come as a result of electronic transitions of the tC chromophore interacting with the transitions of the chirally oriented surrounding bases. CD spectra of the 10 unmodified decamers generally show small variations compared with CD spectra of the corresponding tC-modified sequences (Supplementary Material S2).

CD spectra of the duplex of both the tC-modified (solid curve) and natural sequence (dashed curve) used in the NMR experiments [sequence CA(NMR) in Table 1] are shown in Figure 3. Also included in Figure 3, for comparison, is the normal absorption spectrum (dotted curve) of KtC (structure in Figure 1). For the tC-modified CA(NMR) sequence (solid curve), the same general spectral features as in Figure 2 can be observed. As can be seen in Figure 3, the CD spectrum for the corresponding natural sequence (dashed curve) has essentially the same spectral signature in the region below 300 nm. Another observation from Figure 3 is that the wavelengths at which there are slight differences between the two CD spectra (modified and natural) match well the wavelength positions of the absorption bands for KtC (dotted curve). Furthermore, it can be seen that the positive CD band for the tC-modified duplex in the 325–425 nm region is more intense than what is seen in Figure 2. This is a result of a higher ratio of tC in this duplex (2 tC + 22 natural nucleobases) than in the ones shown in Figure 2 (1 tC + 19 bases).

NMR assignments

The assignment of the non-exchangeable protons relied heavily on the sequential NOE connectivities from aromatic protons to either H1′ or H2′/H2″. For the protons in the base analogue, assignments were based on TOCSY and COSY. In addition, an NOE between H10 and H14 in the tC base system was used. Comparing NOE intensities helped in assigning the spin system. The identification and assignment of the remaining ribose protons was carried out using the TOCSY and DQF-COSY spectra. Complete assignment of the sugar protons was achieved. Stereospecific assignments for all H2′/H2″ and H5′/H5″ pairs were made based on the chemical shifts (31). The assignment of exchangeable protons with the help of the Watergate NOESY yielded all imino protons except H1 on G12 and H3 on T11. Connectivities between imino protons could be observed throughout the duplex. The H1 protons in guanines were used to assign the amino protons of cytosines, while the H3 protons of thymines helped in identifying the H2 protons in adenines. The base analogue tC has an imino proton substituting the amino group protons in normal cytosines (Figure 1). This proton makes numerous contacts with the imino proton on the opposite strand and also with adjacent bases.

Local conformation

Examination of the scalar couplings and the NOEs within the sugar rings, as described in Materials and Methods, lead to the conclusion that all sugars, except those of C1 and G12, adopt a C2′-endo conformation. The ribose ring conformations for C1 and G12 could not be quantitatively determined, probably as a result of increased dynamics. Cross peaks could not be observed in the ^{31}P–^{1}H spectrum for both pairs P–H5′ and P–H5″ of the nucleotides 1–2 and 6–10, which sets an upper limit of 5 Hz on the corresponding scalar coupling. This corresponds to β angles of 170–190°. For nucleotides 3–5 and 11, small but observable couplings with slightly larger values for the P–H5″ pair give β angles in the range 200–220°. The measured P–H3′ couplings are consistent with the ε angles in *trans* conformation as observed in B-form DNA structures (34). NOEs and H4′–H5′/H5″ couplings determine, as described in Materials and Methods, a *gauche*+ conformation of the γ angles for all residues except C1.

Structure calculation

In total, 433 NOEs were assigned. After the removal of the intra-ribose NOEs, used for the determination of the sugar conformations, a set of 127 NOEs were left. These were used as input into the structure calculation. For NOEs involving a methyl group, the volume was divided by three before calibration. The symmetry in the molecule caused by the palindromic sequence has two consequences. First, all NOEs can be assigned as intrastrand or interstrand connections. A NOE was judged to be interstrand when it connected protons on nucleotides separated by more than one residue; the exceptions are the central nucleotides C6 and G7, where the NOE was assigned to the shorter distance as measured in a standard B-DNA model. In the present data set, only three NOEs need to be considered if one assumes a double-stranded duplex and excludes NOEs involving sugar protons which are too far from the opposite strand when assuming reasonable base pairing. Regarding the NOE between H6 of C6 and H8 of G7 (or G107), both H6 and H8 point to their own ribose ring and any interstrand distance is too long. The other two NOEs, from H41/H42 of C6 to H1 of G7 (or G107), are parallel to the hydrogen bonds and add no new information once the base pairing is accepted. Second, the DNA symmetry means that only a set of NOEs describing half of the duplex can be observed. To obtain a full data set, all NOEs were duplicated, yielding the final set with 254 NOEs (Table 2). The same holds for the 46 observed J-couplings yielding 92 dihedral angle constraints. In addition, 22 hydrogen bonds forming the

Table 2. Statistics for NMR data and structure calculation

Experimental constraints	
Assigned NOE peaks	433
Dihedral angles obtained from J-couplings	46
Input for CYANA structure calculation	
NOE upper distance limits	254
Dihedral angle constraints	92
Constrained hydrogen bonds	22
Stereospecific assignments	
H2′/H2″ and H5′/H5″	All
Residual CYANA target function	0.40 ± 0.01 Å^2
Average maximal violations of	
Upper distance limits	0.22 ± 0.01 Å
van der Waals distances	0.07 ± 0.00 Å
Dihedral angles	1.7 ± 0.2°
RMSD for calculated structures relative to mean coordinates	
All heavy atoms in base pairs 3–10	0.45 ± 0.17 Å
All heavy atoms	0.67 ± 0.16 Å
RMSD for mean calculated structure to ideal B-DNA model	
All heavy atoms in base pairs 3–10	2.62 Å
All heavy atoms in base pairs 3–10 of one strand	1.37 Å, 1.41 Å
Heavy atoms of the bases in base pairs 3–10	1.81 Å
Heavy atoms of the bases in base pairs 3–10 of one strand	1.24, 1.26 Å

base pairs 3–10 and 24 stereospecific assignments for H2′/H2″ and H5′/H5″ were included.

The result from 100 CYANA calculations is summarized in Table 2 for the 10 conformers with the smallest target functions. The average target function for these conformers is 0.40 ± 0.01 Å^2. The structure is highly consistent with the NMR data. The only violation of an upper distance limit of 0.2 Å concerns an intranucleotide NOE in base pair 3 (and symmetrically in base pair 10). Violations of van der Waals and torsion angle constraints are negligible. The final 10 conformers exhibit structural variations corresponding to a global root-mean-square distance (RMSD) of 0.67 ± 0.16 Å for all heavy atoms, and of 0.45 ± 0.17 Å for the eight central base pairs. The RMSD between the mean calculated structure of base pairs 3–10 and a B-DNA model is 2.62 Å for all heavy atoms and 1.81 Å for only the bases. If one considers only one strand, these values drop to about 1.4 and 1.25 Å, respectively. Thus, although the B-DNA form is well preserved, there are some significant differences between the present structure and an ideal B-DNA, which are discussed below.

DISCUSSION

UV melting experiments

In this work, we have performed a systematic study of the effect on DNA duplex structure when the fluorescent DNA base analogue tC is incorporated in place of C. We present UV melting data on a series of 10 DNA sequences where all 10 possible combinations of neighbouring bases to tC are investigated. From the results presented in Table 1, it is concluded that the average increase in duplex melting temperature when replacing a C with tC is ∼3°C. Similar increases in melting temperatures have also been reported before upon hybridization between tC-modified PNA single strands and their complementary PNA, DNA or RNA strands (between +3 and +4°C) (25) as well as upon hybridization between tC-modified DNA single strands and their complementary RNA strands (between +1 and +2.5°C) (24). For two of the sequences, we observe a slight decrease in melting temperature (GA −1°C and GC −3°C). Although this reveals a decrease in stability for these two tC-modified duplexes compared to their natural C-containing counterparts, the effect is very small compared with e.g. that reported for the fluorescent DNA base analogues 3-MI (17) and the benzoquinazolines (20–22). The average increase of 3°C in duplex melting temperature per incorporated tC might be explained by the increased π–π overlap between adjacent bases as a C is replaced by tC. It should also be noted that in all the cases where there is a pyrimidine (T or C) on the 5′ side of tC in the DNA strand there is a large increase in T_m. This might come as a result of a more favourable stacking interaction in TptC and CptC compared with those in TpC and CpC. In contrast, tC is not stabilizing when there is a guanine in the same strand immediately 5′ to the modified base. In this case, base stacking interactions between G and tC therefore appear to be slightly less favourable than those in GpC. On the whole, the previously reported ability of tC to discriminate well between G and A targets and the small discrepancy in melting temperatures between the modified and unmodified duplex observed in this study strongly support the fact that the perturbation introduced in the DNA duplex upon tC incorporation is minute and, thus, that tC works well as a cytosine analogue.

Circular dichroism

CD spectra of the series of 10 DNA sequences with different bases neighbouring tC and also of the CA(NMR) sequence show features characteristic of B-form DNA (Figures 2 and 3). As mentioned above, the difference in CD between the modified and unmodified CA(NMR) sequences appears at wavelengths at which tC has its electronic transitions (Figure 3) and, thus, most probably comes as a result of the difference between CD induced in the transitions of tC and CD induced in the transitions of C. Similar conclusions can be drawn from a comparison between the series of 10 modified and unmodified DNA sequences (Supplementary Material S2). Furthermore, it should be noted that the two CD bands appearing above 300 nm, which come exclusively as a result of CD induced in transitions of the tC chromophore, always have negative and positive signs at ∼315 and 375 nm, respectively. This result suggests that even finer details of the structure of tC and its immediate surroundings within the DNA duplex are practically the same independent of neighbouring bases. For some of the single strands containing tC, a CD with the reverse order of signs in the 300–475 nm area can be observed (data not shown), indicating an opposite helicity compared with B-form DNA—or even other, more complex intramolecular structures that may form under these conditions. A more detailed discussion of the structure of a duplex containing tC is presented below.

NMR structure

All NMR data, i.e. scalar couplings and assigned NOEs, are consistent with a structure of the B-DNA form. This is further confirmed by comparison of local conformational characteristics in the final structure such as torsion angles or comparison

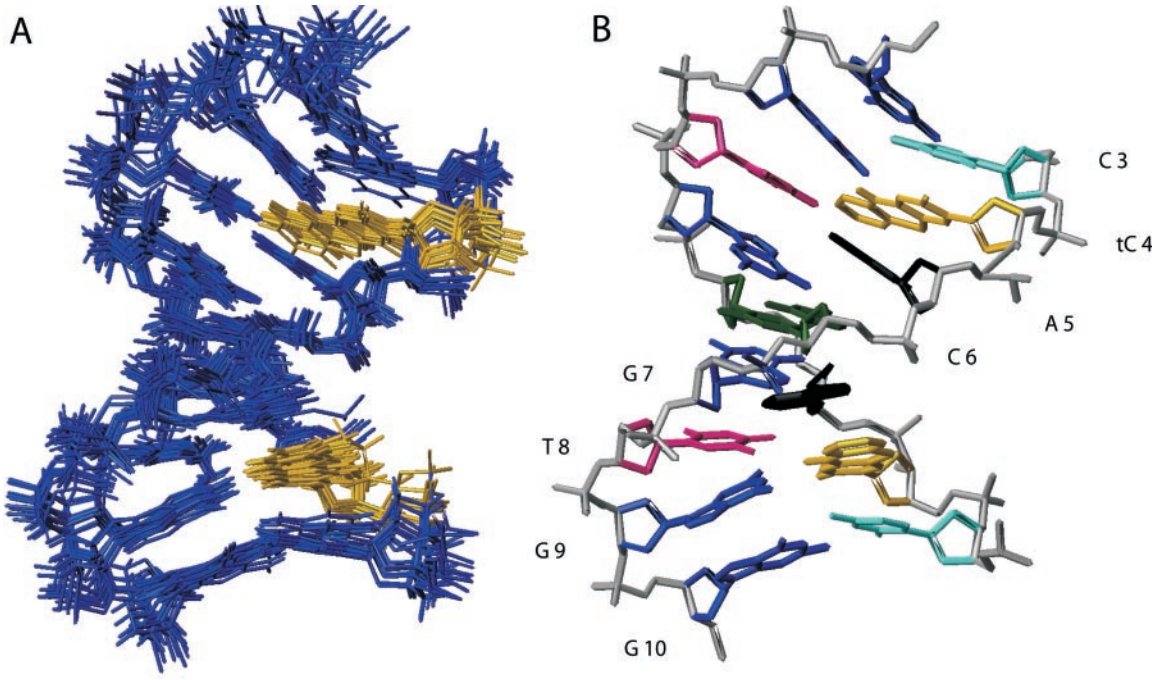

Figure 4. (**A**) The central eight base pairs of the CA(NMR) sequence for the 10 calculated structures with lowest CYANA target functions. The base analogue is shown in yellow. (**B**) One of the structures with nucleotides (except backbone) coloured according to type: adenines in black, cytosines in cyan, guanines in blue, thymines in pink and the base analogue in yellow.

of fragments with only two or three nucleotides (data not shown). Even the superposition of all eight inner nucleotides from one strand shows only limited structural deviation from an ideal B-DNA (Table 2). The calculated overall structure of the DNA duplex may be described as being slightly bent (Figure 4A) even though conclusions regarding macroscopic features are difficult to draw based on the short-range information provided by NOEs and torsion angle constraints. However, bending of DNA is frequently encountered at A-tracts and TG•CA steps (36–40); the two TG•CA steps in this particular sequence contain the modified tC:G base pair, and it may be reasonable to expect a certain degree of DNA-bending in this structure. Examination of direct experimental data shows that the local distortion around the tC-containing base pair in the calculations comes, among other things, as a result of NOEs from the H7, H8 and H9 protons of tC to the methyl group of T8 on the opposite strand in the neighbouring base pair. These NOEs and another one between the imino proton on tC and H2 on A5 are clearly visible in the spectrum but not consistent with an ideal B-DNA. In the calculation, these NOEs appear to pull the large tC base closer to the stack of bases. This positioning of tC closer to the stack of bases is not unexpected since such an arrangement of tC would favour hydrophobic interactions with the T8 methyl group and also improve base stacking interactions of the outer rings of tC and the neighbouring bases.

Implications for duplex stability and structure

CD and UV-Vis spectroscopy both indicate a B-DNA conformation for all sequences, whether tC-modified or not. NMR confirms this and gives valuable insight into the details of the tC base and its surroundings. The expected base-pairing pattern of tC (Figure 1) is confirmed by a number of contacts observed between the imino proton on tC and imino protons on adjacent bases as well as NOEs to the methyl group on T8 and the H2 proton on A5. This not only establishes specific and stable base pairing but also base stacking in accordance with what was expected from the designed structure of tC.

Besides confirmation of the B-DNA conformation, dynamics concerning the tC:G base pair opening and plausible flipping between *anti* and *syn* conformation are important. The base pair opening dynamics can be quantitatively evaluated from the relative intensities of diagonal peaks and cross peaks to the water signal in the NOESY spectrum. No indications can be found of increased base pair opening for the tC:G pair compared to the rest of the base pairs, indicating hydrogen bonds that are as stable as for the naturally occurring base pairs (data not shown). In terms of the *syn*/*anti* conformation of tC and the base pairing guanine, relative NOE intensities of the aromatic proton and the H1′/H2′/H2″ proton support a stable *anti* conformation for both bases, as well as for all other bases. In addition to these observations, detailed information can be gained from the NOE pattern between the aromatic protons H7, H8, H9 and H10 and the methyl group on T8 in the opposite strand. Here, the NOE intensity decreases in agreement with the expected structure of a tC base incorporated into DNA and recognizing the guanine on the opposite strand with its designed base pairing motif (data not shown). Increased dynamics, especially *syn*/*anti* flipping for the tC:G pair, would result in averaging of two NOE patterns.

Minor conformational exchange within tC is expected since the base can adopt two conformations in solution: the benzene ring can be 'flipped' up or down from the plane defined by the aromatic ring closest to the ribose ring. The thermodynamic stability of the conformations is essentially equal and

the energy barrier of interconversion is low; therefore, rapid flipping can be assumed. This is supported by the fact that free tC (KtC) has no CD signal and that a single set of chemical shifts is observed in the NMR spectrum. This also implies that the observed NOEs stem from both conformations and the observed deviations from the ideal B-DNA model for the tC base in the calculated structure mentioned above might be a reflection of these conformational changes. Since the positional variations of the affected atoms remain small, and since this part of tC is located in the major groove and, thus, can move fairly unhindered, these dynamics can take place without significant effect on the surroundings.

Even though the overall appearance of the calculated structure is that of a bent DNA duplex, locally the conformation is fully consistent with B-DNA, e.g. minor and major groove dimensions are normal. These are important observations for future studies involving DNA-binding molecules (e.g. proteins and DNA-binding drugs), which can take place undisturbed from the minor groove side of the duplex. The ability of a protein or other molecule to bind in the major groove might be reduced if the base analogue is placed directly in the binding site, but since the sterical effect of the base analogue is limited to only its immediate surroundings, tC can still be placed sufficiently close to a binding site in e.g. a FRET experiment.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES

1. Ha,T. (2001) Single-molecule fluorescence methods for the study of nucleic acids. *Curr. Opin. Struct. Biol.*, **11**, 287–292.
2. Rist,M.J. and Marino,J.P. (2002) Fluorescent nucleotide base analogs as probes of nucleic acid structure, dynamics and interactions. *Curr. Org. Chem.*, **6**, 775–793.
3. Ward,D.C., Reich,E. and Stryer,L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J. Biol. Chem.*, **244**, 1228–1237.
4. Holmén,A., Nordén,B. and Albinsson,B. (1997) Electronic transition moments of 2-aminopurine. *J. Am. Chem. Soc.*, **119**, 3114–3121.
5. Allan,B.W. and Reich,N.O. (1996) Targeted base stacking disruption by the EcoRI DNA methyltransferase. *Biochemistry*, **35**, 14757–14762.
6. Bloom,L.B., Otto,M.R., Beechem,J.M. and Goodman,M.F. (1993) Influence of 5′-nearest neighbors on the insertion kinetics of the fluorescent nucleotide analog 2-aminopurine by Klenow fragment. *Biochemistry*, **32**, 11247–11258.
7. Guest,C.R., Hochstrasser,R.A., Sowers,L.C. and Millar,D.P. (1991) Dynamics of mismatched base pairs in DNA. *Biochemistry*, **30**, 3271–3279.
8. Hochstrasser,R.A., Carver,T.E., Sowers,L.C. and Millar,D.P. (1994) Melting of a DNA helix terminus within the active site of a DNA polymerase. *Biochemistry*, **33**, 11971–11979.
9. Lycksell,P.O., Gräslund,A., Claesens,F., McLaughlin,L.W., Larsson,U. and Rigler,R. (1987) Base pair opening dynamics of a 2-aminopurine substituted Eco RI restriction sequence and its unsubstituted counterpart in oligonucleotides. *Nucleic Acids Res.*, **15**, 9011–9025.
10. Nordlund,T.M., Andersson,S., Nilsson,L., Rigler,R., Gräslund,A. and McLaughlin,L.W. (1989) Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies. *Biochemistry*, **28**, 9095–9103.
11. Rachofsky,E.L., Osman,R. and Ross,J.B.A. (2001) Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry*, **40**, 946–956.
12. Stivers,J.T. (1998) 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.*, **26**, 3837–3844.
13. Stivers,J.T., Pankiewicz,K.W. and Watanabe,K.A. (1999) Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry*, **38**, 952–963.
14. Freese,E. (1959) *J. Mol. Biol.*, 87–105.
15. Hawkins,M.E. (2001) Fluorescent pteridine nucleoside analogs: a window on DNA interactions. *Cell Biochem. Biophys.*, **34**, 257–281.
16. Hawkins,M.E., Pfleiderer,W., Mazumder,A., Pommier,Y.G. and Falls,F.M. (1995) Incorporation of a fluorescent guanosine analog into oligonucleotides and its application to a real time assay for the Hiv-1 integrase 3′-processing reaction. *Nucleic Acids Res.*, **23**, 2872–2880.
17. Hawkins,M.E., Pfleiderer,W., Balis,F.M., Porter,D. and Knutson,J.R. (1997) Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Anal. Biochem.*, **244**, 86–95.
18. Hawkins,M.E., Pfleiderer,W., Jungmann,O. and Balis,F.M. (2001) Synthesis and fluorescence characterization of pteridine adenosine nucleoside analogs for DNA incorporation. *Anal. Biochem.*, **298**, 231–240.
19. Driscoll,S.L., Hawkins,M.E., Balis,F.M., Pfleiderer,W. and Laws,W.R. (1997) Fluorescence properties of a new guanosine analog incorporated into small oligonucleotides. *Biophys. J.*, **73**, 3277–3286.
20. Godde,F., Toulmé,J.J. and Moreau,S. (1998) Benzoquinazoline derivatives as substitutes for thymine in nucleic acid complexes. Use of fluorescence emission of benzo[g]quinazoline-2,4-(1H,3H)-dione in probing duplex and triplex formation. *Biochemistry*, **37**, 13765–13775.
21. Godde,F., Aupeix,K., Moreau,S. and Toulmé,J.J. (1998) A fluorescent base analog for probing triple helix formation. *Antisense Nucleic Acid Drug Dev.*, **8**, 469–476.
22. Godde,F., Toulmé,J.J. and Moreau,S. (2000) 4-Amino-1H-benzo[g]quinazoline-2-one: a fluorescent analog of cytosine to probe protonation sites in triplex forming oligonucleotides. *Nucleic Acids Res.*, **28**, 2977–2985.
23. Wilhelmsson,L.M., Holmén,A., Lincoln,P., Nielsen,P.E. and Nordén,B. (2001) A highly fluorescent DNA base analogue that forms Watson–Crick base pairs with guanine. *J. Am. Chem. Soc.*, **123**, 2434–2435.
24. Lin,K., Jones,R.J. and Matteucci,M. (1995) Tricyclic 2′-deoxycytidine analogs: syntheses and incorporation into oligodeoxynucleotides which have enhanced binding to complementary RNA. *J. Am. Chem. Soc.*, **117**, 3873–3874.
25. Eldrup,A.B., Nielsen,B.B., Haaima,G., Rasmussen,H., Kastrup,J.S., Christensen,C. and Nielsen,P.E. (2001) 1,8-Naphthyridin-2(1H)-ones: novel bicyclic and tricyclic analogues of thymine in peptide nucleic acids (PNAs). *Eur. J. Org. Chem.*, 1781–1790.
26. Wilhelmsson,L.M., Sandin,P., Holmén,A., Albinsson,B., Lincoln,P. and Nordén,B. (2003) Photophysical characterization of fluorescent DNA base analogue, tC. *J. Phys. Chem. B*, **107**, 9094–9101.
27. Brown,T. and Brown,D.J.S. (1992) Purification of synthetic oligonucleotides. *Meth. Enzymol.*, **211**, 20–35.
28. Dawson,R.M.C., Elliot,D.C., Elliot,W.H. and Jones,K.M. (1986) Oxford University Press, NY.
29. 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.
30. Bartels,C., Xia,T.H., 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.
31. Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, Inc., NY.
32. Kim,S.G., Lin,L.J. and Reid,B.R. (1992) Determination of nucleic-acid backbone conformation by 1H NMR. *Biochemistry*, **31**, 3564–3574.
33. Güntert,P., Mumenthaler,C. and Wüthrich,K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
34. Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.), *NMR of Macromolecules: A Practical Approach*. Oxford University Press, NY, pp. 217–288.
35. Rodger,A. and Nordén,B. (1997) *Circular Dichroism and Linear Dichroism*. Oxford University Press, Oxford, NY.
36. Nilges,M., Clore,G.M., Gronenborn,A.M., Brunger,A.T., Karplus,M. and Nilsson,L. (1987) Refinement of the solution structure of the

DNA hexamer 5′d(Gcatgc)2: combined use of nuclear magnetic resonance and restrained molecular dynamics. *Biochemistry*, **26**, 3718–3733.
37. Nilges,M., Clore,G.M., Gronenborn,A.M., Piel,N. and McLaughlin,L.W. (1987) Refinement of the solution structure of the DNA decamer 5′d(ctggatccag)2: combined use of nuclear magnetic resonance and restrained molecular dynamics. *Biochemistry*, **26**, 3734–3744.
38. Baleja,J.D., Pon,R.T. and Sykes,B.D. (1990) Solution structure of phage lambda half-operator DNA by use of NMR, restrained molecular dynamics, and NOE-based refinement. *Biochemistry*, **29**, 4828–4839.
39. Mujeeb,A., Kerwin,S.M., Kenyon,G.L. and James,T.L. (1993) Solution structure of a conserved DNA sequence from the HIV-1 genome: restrained molecular dynamics simulation with distance and torsion angle restraints derived from two-dimensional NMR spectra. *Biochemistry*, **32**, 13419–13431.
40. Weisz,K., Shafer,R.H., Egan,W. and James,T.L. (1994) Solution structure of the octamer motif in immunoglobulin genes via restrained molecular-dynamics calculations. *Biochemistry*, **33**, 354–366.