Journal of Biomolecular NMR, **22:** 265–280, 2002.
KLUWER/ESCOM

High-resolution NMR structure of an AT-rich DNA sequence

Nikolai B. Ulyanov[a], William R. Bauer[b] & Thomas L. James[a,*]
[a]*Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, U.S.A.;* [b]*Department of Molecular Genetics and Microbiology, Health Sciences Center, State University of New York, Stony Brook, NY 11794-5222, U.S.A.*

Received 23 November 2001; Accepted 14 January 2002

Key words: AT-rich sequence, DNA, GAA loop, NMR, Pribnow box, stem-and-loop structure

Abstract

We have determined, by proton NMR and complete relaxation matrix methods, the high-resolution structure of a DNA oligonucleotide in solution with nine contiguous AT base pairs. The stretch of AT pairs, TAATTATAA·TTATAATTA, is imbedded in a 27-nucleotide stem-and-loop construct, which is stabilized by terminal GC base pairs and an extraordinarily stable DNA loop GAA (Hirao et al., 1994, *Nucleic Acids Res.* **22**, 576–582). The AT-rich sequence has three repeated TAA·TTA motifs, one in the reverse orientation. Comparison of the local conformations of the three motifs shows that the sequence context has a minor effect here: atomic RMSD between the three TAA·TTA fragments is 0.4–0.5 Å, while each fragment is defined within the RMSD of 0.3–0.4 Å. The AT-rich stem also contains a consensus sequence for the Pribnow box, TATAAT. The TpA, ApT, and TpT·ApA steps have characteristic local conformations, a combination of which determines a unique sequence-dependent pattern of minor groove width variation. All three TpA steps are locally bent in the direction compressing the major groove of DNA. These bends, however, compensate each other, because of their relative position in the sequence, so that the overall helical axis is essentially straight.

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; DQFCOSY, double quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; fid, free induction decay; EDTA, ethylenediamine tetraacetic acid; TSP, sodium tetrasilyl propionate; PDB, protein data base or protein data base format; RMSD, root mean square deviation.

Introduction

AT-rich DNA sequences play a number of structural and functional roles in the cell cycle. Promoter sequences for transcription initiation often contain conserved AT-rich sequences specifically recognized by proteins. Examples include the Pribnow box in prokaryotes (McClure, 1985; Pribnow, 1975), and TATA box in eukaryotes (reviewed in Hampsey, 1998). So-called initiator elements, which surround transcription start and control initiation of transcription in TATA-less promoters, have very diverse AT-rich sequences in *Giardia lamblia* genes (Elmendorf et al., 2001). In addition, these genes have large upstream AT-rich elements, which modulate levels of gene expression in *G. lamblia* (Elmendorf et al., 2001). Many yeast promoters contain homopolymeric $A_n \cdot T_n$ sequences, which were shown to stimulate transcription via its intrinsic structure, possibly by proper positioning of nucleosomes in the upstream regions (Iyer and Struhl, 1995). Sequences of consecutive adenines and thymines, so-called A-tracts, constitute an important class of naturally curved DNA (Crothers and Shakked, 1999; Marini et al., 1982); intrinsically curved DNA plays a role both in prokaryotic transcription (Carmona and Magasanik, 1996) and positioning of nucleosomes in eukaryotic chromatin (reviewed in Olson and Zhurkin, 1996). AT-rich sequences, including A-

*To whom correspondence should be addressed. E-mail: james@picasso.ucsf.edu

tracts, are also strongly over-represented in origins of replication in many organisms (Boulikas, 1996).

Another example of extremely AT-rich sequences is the terminal region of vaccinia virus, a double-stranded DNA poxvirus with terminal hairpin loops. The ultimate 106 nucleotides at each end of the viral DNA form incompletely base paired hairpins with more than 90% AT content. The left and right terminal sequences are related by complementary (flip-flop) symmetry (Baroudy et al., 1982). During DNA replication, the hairpins are converted into an imperfect inverted repeat, having its pseudosymmetry axis at the terminus of the original hairpin (Baroudy et al., 1983). Portions of the replicative intermediate inverted repeat have been cloned into plasmid vectors, in which the imperfect palindromes can be reversibly extruded into cruciforms containing bulged bases (Delange and McFadden, 1990). The resulting bulged bases in this AT-rich telomere are especially interesting because they strongly influence the energetics of the cruciform extrusion (Benham et al., 2001). Currently, we are studying the structure of a fragment of vaccinia virus telomeric region containing G and C bulges, d(TAATTATAA)·d(TTAGTACATTA), using NMR. Here, we report our studies on the control sequence, which lacks the G and C bulges, because the control sequence is interesting on its own. The most interesting feature of this sequence is nine contiguous AT base pairs in the middle of the stem. The AT-rich stretch consists of three repeated motifs TAA·TTA, one of them in the reverse orientation. In addition, this sequence contains the TATAAT motif (in the reverse orientation), a consensus sequence for the Pribnow box. The Pribnow box sequence has been studied in our laboratory before (Schmitz et al., 1992), but in a context of a short oligonucleotide, DNA octamer d(GTATAATG)·d(CATTATAC); solving this structure in the context of a longer construct allowed us to estimate the influence of the duplex ends in the shorter sequence.

One common problem with studying AT-rich sequences in solution is that they have low thermal stability when designed as small constructs. A strategy used in this work was to make a monomolecular construct (hereafter, the 27mer) with an extraordinarily stable DNA hairpin loop, GAA:

$$\begin{array}{cccccccccccccc} C_1 & -C_2 & -T_3 & -A_4 & -A_5 & -T_6 & -T_7 & -A_8 & -T_9 & -A_{10} & -A_{11} & -C_{12} & -G_{13} & \\ | & | & | & | & | & | & | & | & | & | & | & | & & A_{14} \\ G_{27} & -G_{26} & -A_{25} & -T_{24} & -T_{23} & -A_{22} & -A_{21} & -T_{20} & -A_{19} & -T_{18} & -T_{17} & -G_{16} & -A_{15} & \end{array}$$

This hairpin loop has a compact structure stabilized by stacking interactions and a sheared G·A pair with two hydrogen bonds GNH_2–AN7 and ANH_2–GN3 (Hirao et al., 1994).

GC base pairs at the terminus additionally increase the stability of this construct. Making the construct mono- rather than bi-molecular makes the stability independent of DNA concentration. We report here a high-resolution structure for the 27mer, solved by homonuclear NMR and complete relaxation matrix methods.

Materials and methods

DNA synthesis

The 27mer oligonucleotide was synthesized at the SUNY at Stony Brook facility on an Applied Biosystems DNA synthesizer model 395 using the manufacturer's standard procedures. Lyophilized samples were deprotected by resuspension in NH_4OH followed by incubation at 55° for 8 h. The deprotection solvent was removed by 3–4 cycles of evaporation with a Speedvac at 0° for 12 h, followed by resuspension in H_2O. Samples were then purified by filtration through a Filtron 1K filter assembly by centrifugation in a Sorvall SS34 rotor for 90–120 min at 0°. Filtrates were resuspended and refiltered until the absorbance ratios were $A_{232}/A_{260} > 2.8$ and $A_{280}/A_{260} > 1.95$ (generally 4–5 cycles). Final adjustment of the solvent for the NMR experiments was made by 4–5 cycles of evaporation to dryness in the Speedvac (0°, 16 h) and resuspension in either H_2O or D_2O and buffer.

NMR experiments

NMR samples were prepared by dissolving the lyophilized DNA powder in potassium phosphate buffer, either in D_2O or in a 90/10% mixture of H_2O and D_2O. The concentration of potassium phosphate in the stock solutions was confirmed by atomic absorption spectroscopy (Sequoia Analytical, precision $\pm$ 0.5%). D_2O samples were additionally lyophilized several times from increasingly higher isotopically enriched D_2O. Unless stated otherwise, the buffer contained 30 mM K^+ at pH 8.0, 1 mM EDTA, and a trace amount of TSP. The oligonucleotide concentration was approximately 2 mM. All NMR experiments were run on a Varian Inova 600 MHz spectrometer; and unless stated otherwise, experiments were carried out at

10 °C. 1D spectra in water were acquired with jump-and-return water suppression (Gueron et al., 1991), with the maximum of excitation placed in the center of the imino proton region. All 2D spectra were acquired with States-Haberkorn phase cycling (States et al., 1982). 2D NOESY spectra in water were acquired with the SS-NOESY program (Smallcombe, 1993) with mixing time 125 ms, spectral width 13000 Hz, 2K complex data points, 490 fids, 32 transients, and 2 s relaxation delay. Water suppression was achieved with a sine-shaped last ninety-degree pulse 'S1558g', which creates an excitation maximum at 4300 Hz from the carrier frequency (i.e., 12.1 ppm at 10 °C). To minimize baseline distortions, one data point was linearly back-predicted in each fid during processing. Three 2D NOESY data sets were collected in D_2O, one with a mixing time of 75 ms and two with mixing times of 150 ms. In each case, the data were acquired with spectral width 6500 Hz, 2K complex data points, 512 fids, 16 transients, and relaxation delay 4 s. The 75-ms data set and one of the 150-ms data sets were acquired without any suppression of the residual HDO signal; in this case, HDO signal was subtracted in the time domain during processing. Another 150-ms data set was acquired with a weak presaturation of the water frequency during 1 s of the four-second relaxation delay period. A DQFCOSY spectrum (Rance et al., 1983) and two TOCSY spectra (Bax and Davis, 1985) (with and without presaturation of the residual HDO signal) were acquired in D_2O. Typically, all 2D data sets were zero-filled to a 2K × 2K size and multiplied with a Gaussian window function prior to Fourier transform. Processed data were baseline-corrected in both dimensions. 1D data were processed and analyzed with the Varian VNMR software. 2D data sets were processed with the NMRPipe software (Delaglio et al., 1995), transferred to the 'UCSF format', and then analyzed, assigned and annotated using the Sparky program (Goddard and Kneller, 1998) run on a O2 SGI workstation. Proton resonances were assigned using standard homonuclear methods. NOE cross-peaks were integrated using Gaussian line fitting and deconvolution of moderately overlapped peaks with Sparky.

Structure calculation

The NMR structures for the 27mer were refined using programs DYANA (Güntert et al., 1997) and miniCarlo (Ulyanov et al., 1993; Zhurkin et al., 1991), using NOE-derived distance restraints. The miniCarlo program is based on a set of specialized internal coordinates, called helical parameters. These parameters define the relative positions of idealized aromatic bases and sugars in a nucleic acid molecule, as well as internal conformation of sugar moieties. The sugar-phosphate atoms are then calculated by a chain closure algorithm (Zhurkin et al., 1978). After the atomic Cartesian coordinates are defined for the molecule, its energy is calculated using the full-atom Zhurkin–Poltev–Florentiev force field (Zhurkin et al., 1980). DYANA is also an internal coordinate-based program but, in contrast to miniCarlo, it uses generic internal coordinates – torsion angles. Both programs can carry out restrained energy minimization; miniCarlo also performs Metropolis Monte Carlo simulations, and DYANA performs molecular dynamics in the torsion angle space. We previously showed that using miniCarlo for refinement yields structures within 0.5 Å RMSD of those determined by restrained molecular dynamics (Ulyanov et al., 1993). However, in our experience, convergence is achieved more readily with miniCarlo – probably due to its use of idealized bases and sugars.

We used three 2D NOESY data sets in D_2O at 10 °C to determine quantitative distance restraints using the full relaxation matrix analysis program MARDIGRAS (Borgias and James, 1990) with the random error analysis procedure RANDMARDI (Liu et al., 1995). The MARDIGRAS program employs an iterative procedure to determine interproton distances from NOE data, and it takes into account a complete network of relaxation pathways in a molecule. Quantitative distance restraints for nonexchangeable protons were supplemented with the qualitatively categorized distance restraints for exchangeable protons based on the NOE cross-peak intensities in the water 2D NOESY data set. In addition, idealized hydrogen bonds restraints were used for all base pairs during the simulated annealing protocols (torsion angle dynamics with DYANA and simulated annealing Metropolis Monte Carlo with miniCarlo). However, at the final stages of refinement (restrained energy minimization with miniCarlo), hydrogen bond restraints were kept only for the Watson-Crick GC pairs: $C_1 \cdot G_{27}$, $C_2 \cdot G_{26}$ and $C_{12} \cdot G_{16}$. The DYANA program was used to carry out simulated annealing with the torsion angle dynamics (10 000 steps with the macro 'calc_all') starting with randomized initial structures. For the DYANA calculations, interproton distance restraints were used with the weight 1.0, and hydrogen bond restraints were used with the weight 5.0. In addition, bond length

C4′-O4′ was restrained in the range 1.41–1.42 Å with the weight 100.0, and the distance O4′-H4′ was restrained in the range 2.10–2.20 Å with the weight 20.0. These restraints ensured proper closure of the five-member furanose cycles and the correct stereochemistry for the prochiral centers. The δ torsion angle (C5′-C4′-C3′-O3′) was constrained to be in the range 79–150° in the DYANA calculations; this range covers all sugar conformations between C3′-*endo* and C2′-*endo*. The stereochemistry of the prochiral centers was checked with an in-house developed program CHIRANO. NOE intensities for the refined structures were calculated and compared with the experimental data, using the program CORMA (Keepers and James, 1984). Helical parameters of the structures were calculated with the Fitparam program (Ulyanov et al., 1992). Molecular graphics representations of the structure were prepared with the MidasPlus and Chimera programs (Ferrin et al., 1988; Huang et al., 1996).

Results

1D spectra

Imino proton spectra of the 27mer in water were acquired for a number of different potassium ion concentrations and temperatures. Variation of the potassium concentration from 10 to 100 mM K^+ does not produce significant spectral changes at 10 °C and pH 8 (four bottom spectra in Figure 1A), except for the appearance of a broad peak between 10 and 11 ppm at 10 mM K^+. The position of this peak is consistent with unpaired T's partially protected from solvent; it may be due to transient openings of the AT-rich core of the 27mer. Absence of this peak at higher concentrations of potassium is consistent with general stabilization of DNA duplexes with ionic strength. However, this peak reappears even at the highest potassium concentration, 100 mM, when the exchange rate with solvent is reduced by lowering the pH to 5.0 (Figure 1A, top spectrum). We are not aware of other reports showing the imino resonances of transiently unpaired T's in fully base paired portion of duplex DNA; this may be a feature specific for AT-rich sequences. 30 mM K^+ at pH 8 and temperature 10 °C was used for the acquisition of most 2D spectra and for the structure determination.

The 27mer sample was stored at 4 °C between the acquisitions of NMR spectra. Before each 2D acquisition, the sample was heated inside the spectrometer to a temperature corresponding to the disappearance of imino proton signals; then, the sample was cooled down to 10 °C. A typical temperature profile during the heating is shown in Figure 1B (five bottom spectra). The top spectrum corresponds to the annealed sample; this is the condition used for the structure determination. This spectrum did not change after the 2D data acquisition. It is clear that the annealed spectrum lacks a minor peak at ca. 12.3 ppm (shown with an asterisk); at the same time, the shoulder at ca. 13.2 ppm appears to be stronger. It is possible that the two correspond to species of the same proton (a species at 12.3 ppm has maximum intensity at 20–30 °C, and a species at 13.2 ppm is stronger at 10 °C. We never assigned these two resonances, because they do not have any cross-peaks in NOESY spectra (except, exchange peaks with water). In the aromatic region, certain resonances (later assigned as adenine H2 protons in the TpA steps) also showed significant pre-melting temperature dependence (data not shown). The strongest dependence was observed for the A_{21}H2 proton in the middle of the 27mer: its chemical shift changed from 6.37 to 6.70 ppm with temperature increase from 10 to 50 °C.

Proton assignments

Proton resonances of the 27mer are reasonably well dispersed, both in the imino and fingerprint regions (Figure 2). Partial spin systems of GC pairs (imino proton of G; H5, H6 and amino protons of C residues) were assigned based on a set of strong cross-peaks: imino–amino, amino–amino, amino–H5, and H5–H6, and relatively weak imino–H5 and amino–H6 cross-peaks (not shown). H5–H6 cross-peaks were verified in the TOCSY spectrum. AT pairs were identified based on the TH3–AH2 cross-peaks. Base pairs were placed in sequential order using imino–imino connectivities (Figure 2A); this was later verified using assignments for the nonexchangeable protons. All 12 Watson Crick base pairs of the 27mer were readily identified from the water NOESY data set, even though the imino–amino cross-peaks were weak in the $C_1 \cdot G_{27}$ base pair due to end fraying. The imino proton of G_{13} was not observed in water spectra. The equivalent imino proton (G_3 H1) has been observed at ca. 10.5 ppm by Yoshizawa et al. for the mini-hairpin d(GCGAAGC) at 5 °C but not at higher temperatures (Yoshizawa et al., 1997).

Nonexchangeable protons were assigned by standard methods, using NOESY, TOCSY and DQF-

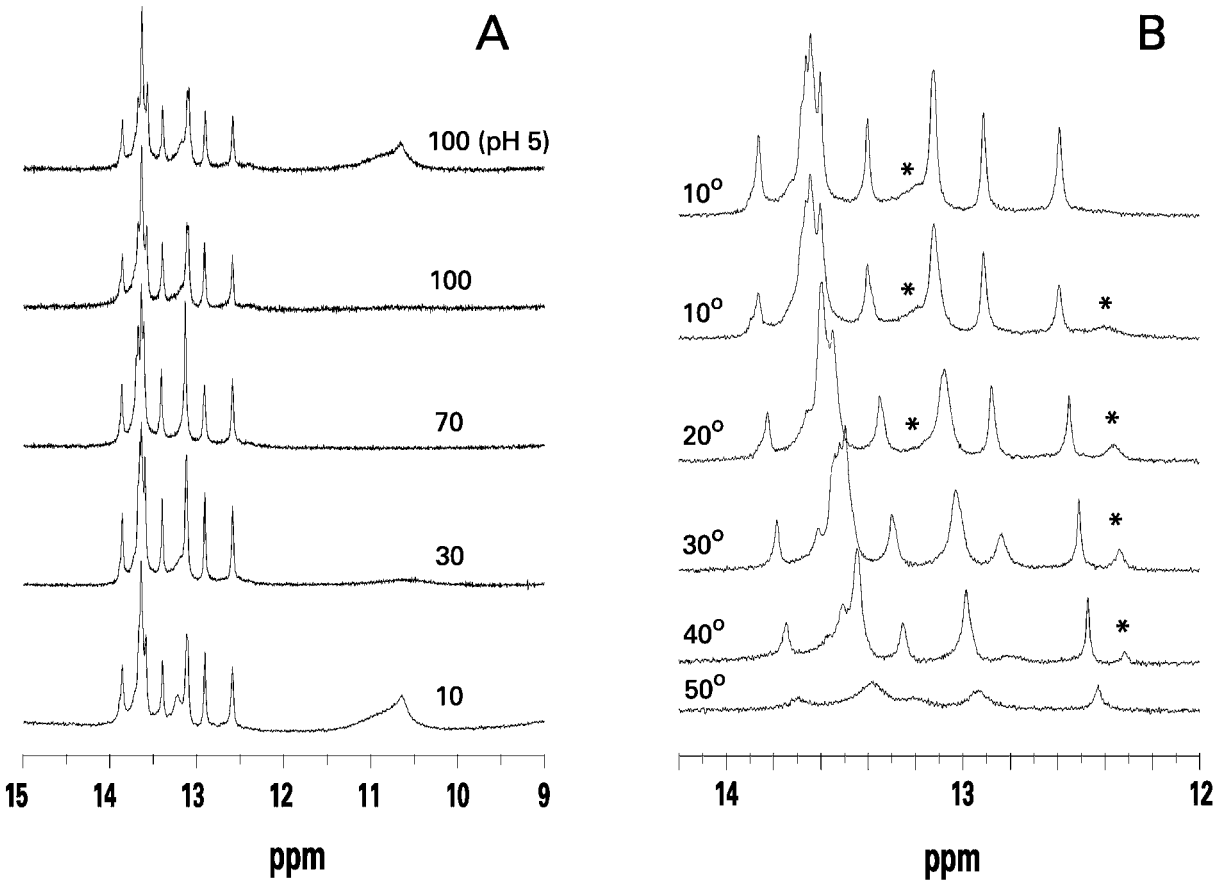

Figure 1. Imino proton spectra with jump-and-return water suppression. (A) Dependence of the spectrum upon the K^+ concentration at 10 °C. The labels indicate the [K^+] in mM. All spectra were collected at pH 8, except the top spectrum, which was acquired at pH 5. (B) Dependence of the spectrum upon the temperature at 30 mM K^+. The labels indicate the temperature. The top spectrum at 10 °C is for the annealed sample, and all other spectra are for the sample stored at 4 °C for several days. Asterisks denote unassigned minor species.

COSY spectra at 10 °C, and occasionally using data at 20 °C. Connectivities for the two strands of the 27mer stem were traced in the aromatic-to-anomeric portion of D_2O NOESY spectra (Figure 2B), and also in aromatic-to-H3′, aromatic-to-H4′, aromatic-to-H2′/H2″, and aromatic-to-methyl regions (not shown). The GAA loop has a special NMR signature (Hirao et al., 1994; Yoshizawa et al., 1997), which helped assign these residues. This signature includes readily identifiable upfield-shifted resonances of A_{14} H3′, H4′, H5′, H5″, A_{15} H5′, H5″ and G_{13} H1′. In agreement with findings of Hirao et al. (Hirao et al., 1994), some protons of the GAA loop have broad linewidth at low temperature, especially G_{13} H1′ and A_{15} H1′, H8 (Figure 2A). They become sharper at 20 °C but still remain broader than other protons (data not shown). In addition, protons H1′ and H6 of the stem-closing residue C_{12} are somewhat broadened. Protons of the loop residue A_{14} and stem-closing residue G_{16} are relatively sharp, even at 10 °C.

Proton assignments for the 27mer have been deposited with the BioMagResBank (entry 5167), and also they are included in the Supplementary Material. The assignments are 82% complete, which is typical for a DNA oligonucleotide assigned with homonuclear methods. Protons H5′ and H5″ were assigned for only a few residues. The assignments for resolved H5′/H5″ protons were model-based, and they were included only at late stages of refinement. Imino protons were assigned for all residues, except G_{13}; amino protons were assigned for all cytosines and for occasional adenines. Stem adenine H2 protons were assigned based on the water NOESY spectrum, and A_{15} H2 was assigned based on D_2O data sets. The A_{14} H2 proton does not have any cross-peaks in NOESY data sets; it

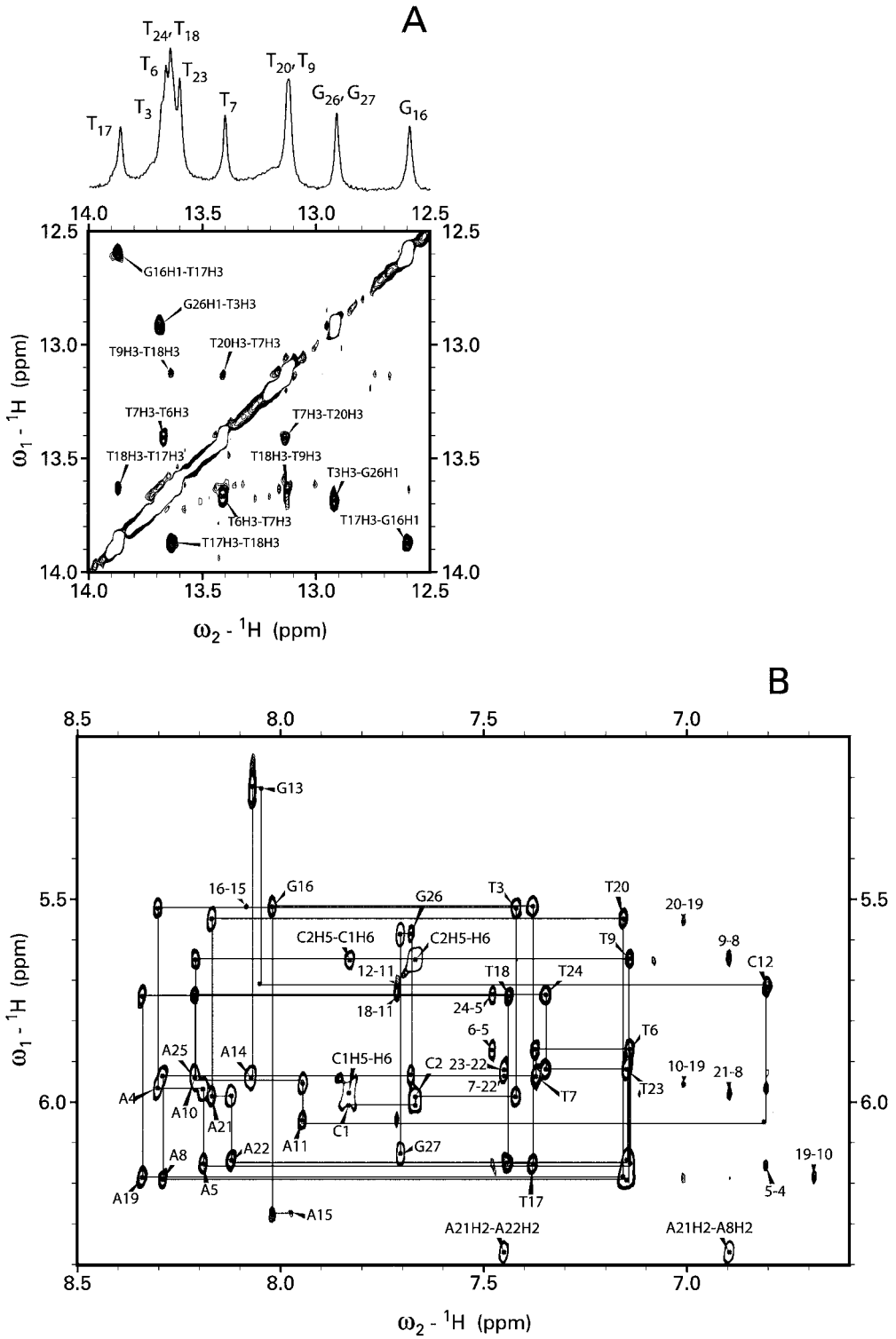

Figure 2. 2D NOESY spectra of the 27mer. (A) Water SS-NOESY spectrum, with the 1D jump-and-return spectrum shown on the top. (B) Fingerprint region of the D_2O NOESY spectrum. Lines show sequential connectivities, with the intraresidue H1′–H6/H8 peaks labeled. Peaks labeled with numbers only refer to interresidue H1′–H2 peaks. The first number refers to the residue with H1′, and the second number refers to adenine with H2 (for example, '20-19' refers to the T_{20}H1′–A_{19}H2 cross-peak).

was assigned based on its longer T_1 relaxation time in the T_1 inversion recovery experiment (not shown).

The connectivities in the fingerprint region are contiguous throughout the whole molecule, except for a single interruption at the A_{14}–A_{15} step (Figure 2A). It is possible, however, that the $A_{14}H1'$–$A_{15}H8$ cross-peak was not observed due to a large linewidth of $A_{15}H8$ at 10 °C. Indeed, this cross-peak, although a very weak one, was observed at 20 °C (not shown). In addition, relatively strong cross-peaks $A_{14}H4'$–$A_{15}H2$, $A_{14}H5'$–$A_{15}H2$ and $A_{14}H5''$–$A_{15}H2$ were observed for this step, which is indicative of an unusual conformation. Other features of the fingerprint region include a relatively intense intraresidue cross-peak $A_{14}H1'$–H8, which suggests that this residue may be in the *syn* conformation part of the time. The interresidue cross-peak $G_{13}H1'$–$A_{14}H8$ is also very strong; it is very prominent in NOESY spectra despite the broad $G_{13}H1'$ linewidth.

Structure calculation

We used the miniCarlo program to build and energy-minimize an initial model for the 27mer. This initial model was used as input for the MARDIGRAS program together with experimental NOE intensities from D_2O NOESY data sets. Only those intensities that could be reliably integrated were included in the calculations. MARDIGRAS was run separately for each of the D_2O NOESY data sets, each time utilizing a random error analysis procedure, RANDMARDI. According to this procedure, 50 random NOE intensity sets were created by perturbing the experimental intensity within the limits of its estimated relative integration error. The integration errors were estimated by comparing the intensities for the cross-peaks above and below the diagonal, but in no case were they set to less then 10%. MARDIGRAS was run for each of the random data sets, thus producing 50 distance estimates. For each distance, the lowest 10% and largest 10% of the estimates were discarded, and the data were pooled together for the three NOESY data sets, producing a set of distance restraints (lower and upper limits for each distance). Running the MARDIGRAS program also requires input of an overall correlation time; the whole procedure was run for a series of correlation times from 3 to 15 ns. Each time, the calculated restraints were compared with fixed interproton distances and distances with small allowed variation (mostly intra-sugar distances) in the 27mer. Such distances were reproduced best with correlation times of 8 and 9 ns; the two corresponding sets of distance restraints were combined and used for the structure calculation. The intra-sugar distances with low variation were excluded from this final set, because they do not carry any structural information.

The initial model was further restrained-energy minimized with miniCarlo using the set of distance restraints calculated as described above. The lower and upper bounds for each distance restraint were used to define the flat-well potential (Kerwood et al., 1991); this was added to the force-field energy during the restrained minimization. The resulting structure was then used as the input structure for another round of the RANDMARDI procedure. The reason for iterating this procedure is that we found some residual dependence of distances determined by MARDIGRAS upon the initial structure, especially at longer correlation times. In addition, at the end of each iteration, we re-analyzed NOE cross-peaks corresponding to violated distance restraints, checking for possible assignment and integration errors. The whole protocol was iterated four times. The course of refinement was monitored by conformational energy of the molecule and residual distance deviation (reported by miniCarlo) and by the NOE-based R^x factors (James, 1991) calculated with CORMA (Keepers and James, 1984). The fourth iteration produced little change in either the MARDIGRAS-calculated distances or the refined structure. This set of distance restraints was considered final; it contained 353 distances involving nonexchangeable protons with the average flat-well width (the difference between the upper and lower bounds) of 1.51 Å (Table 1). At this stage, restraints for nonexchangeable protons were supplemented with a set of semi-quantitative distance bounds for exchangeable protons, based on the water NOESY data set. Restraints are well distributed over the molecule, without gaps (Figure 3). However, the GAA loop has fewer interresidue restraints than the rest of the molecule – in part because fewer NOE peaks were observed in the loop, and in part because some peaks were not suitable for integration due to their large linewidths.

In order to determine how precisely the structure is defined by the experimental restraints, it is necessary to assess the convergence of the refinement protocol starting with various initial structures. To generate a pool of sufficiently different starting structures, we used DYANA; 100 starting conformations were generated with randomized torsion angles and then subjected to torsion angle dynamics simulated annealing. The purpose of this simulated annealing is

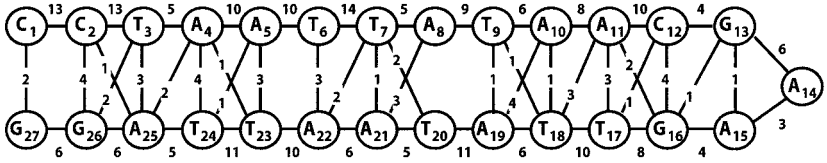

Figure 3. Distribution of interresidue distance restraints used for the refinement of the 27mer.

Table 1. Refinement statistics

Number of distance restraints	
Nonexchangeable	353
Flat-well width, Å	1.51
Exchangeable	63
Hydrogen bonds	18
Total	434
Per residue	16.1
Conformational energy, kcal/mol	-217.9 ± 2.3
Restraint energy, kcal/mol	51.5 ± 5.1
Residual distance deviation, Å	0.10 ± 0.01
NOE-based R^x factor ($\times 10^2$)	
75 ms NOESY data set	7.07 ± 0.19
150 ms data set	6.34 ± 0.22
150 ms data set	6.51 ± 0.18
Atomic root-mean square deviation, Å[a]	
All residues	1.04 ± 0.24
AT pairs	0.64 ± 0.11
Residues 12–16	0.83 ± 0.38

[a]Only heavy atoms are taken into account.

to fold the molecule into a helical conformation with a roughly correct pattern of stacking interactions. For this high-temperature simulation, we also used idealized restraints for hydrogen bonds in all AT and GC pairs and in the GA pair $G_{13}{\cdot}A_{15}$. We do not have direct NMR evidence of the $G_{13}{\cdot}A_{15}$ pair. However, sheared conformation of GA pair in the GAA loop was proved by Hirao et al. (1994) who studied 7-deaza substitutions in the extraordinarily stable mini-hairpin d(GCGAAGC).

Fifty structures with the best DYANA scores were selected for further refinement and stored as PDB files. Internal coordinates (helical parameters) were extracted from the PDB files with the Fitparam program and used as input for miniCarlo. A helical parameters-based Metropolis Monte Carlo simulated annealing was then carried out for each structure. We tried several different protocols for simulated annealing with miniCarlo. In the end, a quite simple protocol achieved low residual distance restraint deviations, low conformational energy, and convergence of refined structures. It started with 2000 Metropolis Monte Carlo steps at 1000 K, followed by exponential cooling to 107 K during next 2000 steps. The high-temperature simulation was carried out with the distance restraint force constant of 1 kcal/(mol Å^2); during cooling, it was exponentially increased to 4 kcal/(mol Å^2). One problem with using miniCarlo structures generated by other software is that backbone closure (Zhurkin et al., 1978) may fail if bond angles deviate too much from their ideal values in the input conformations. To alleviate this problem, the weight of the backbone closure term was gradually increased from zero to 100% during the high-temperature period. Monte Carlo simulated annealing was followed by the miniCarlo restrained energy minimization with the force constant of 2 kcal/(mol Å^2). Ten structures with lowest total energies were further restrained-minimized with the help of the grid search technique in the helical parameters-coordinate space. During this minimization, the hydrogen bond restraints were removed for all base pairs except three Watson–Crick GC pairs. The rationale for this is that NMR data indicate only the existence of those hydrogen bonds, but not their geometric parameters; using idealized restraints could bias the structures toward more regular conformations. Nevertheless, we kept hydrogen bond restraints for the GC pairs to minimize the amount of fraying at the termini. The ensemble of ten final structures is shown in Figure 4. The sixth-root weighted NOE-based R^x factor (James, 1991) was calculated for each structure using CORMA and reported along with other refinement statistics in Table 1.

Discussion

MARDIGRAS-generated distance restraints for nonexchangeable protons have a rather large average flat-well width of 1.5 Å (Table 1). Nevertheless, this set of distance restraints was sufficient to ensure the convergence of the refinement. Ten final structures shown in Figure 4 have pair-wise atomic RMSD values between 0.56 and 1.69 Å, 1.04 Å on average. The AT-rich core

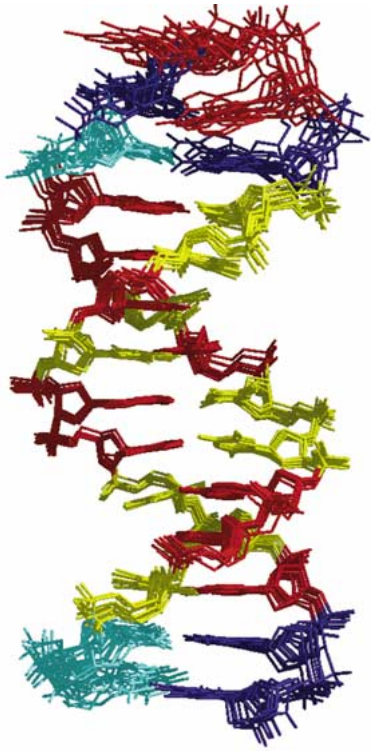

Figure 4. Ten final structures superimposed using the residues of the AT-rich core.

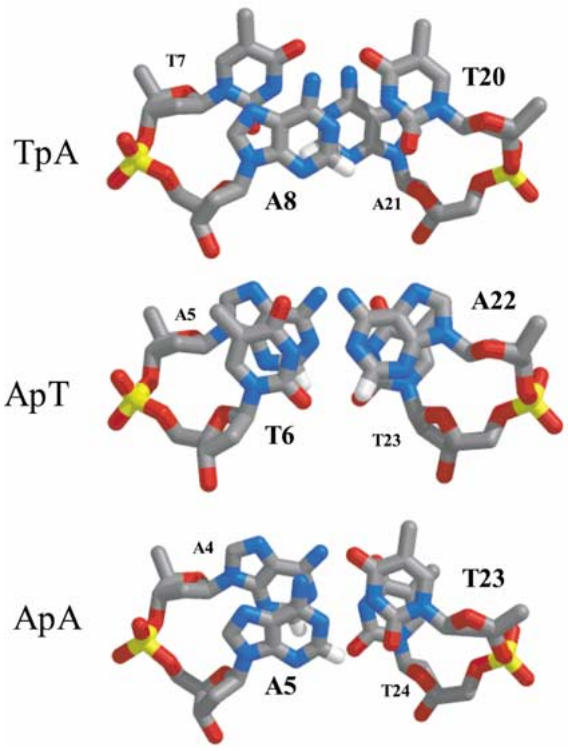

Figure 7. Stacking interaction patterns in representative TpA, ApT, and ApA·TpT steps. Oxygen atoms are shown in red, carbon atoms in gray, nitrogen atoms in blue, phosphorus atoms in yellow and adenine H2 protons are shown in white. Note the correlation of adenine H2 chemical shifts with their proximity with purine rings (see text).

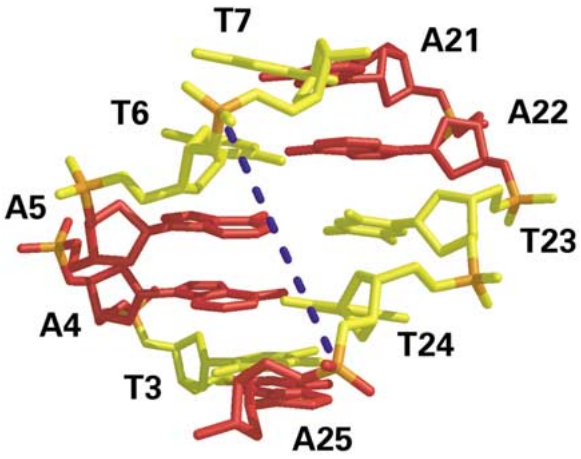

Figure 8. Illustration of correlation between the minor groove width and helical parameters (see text and Figure 5D). Adenines are shown in red and thymines in yellow.

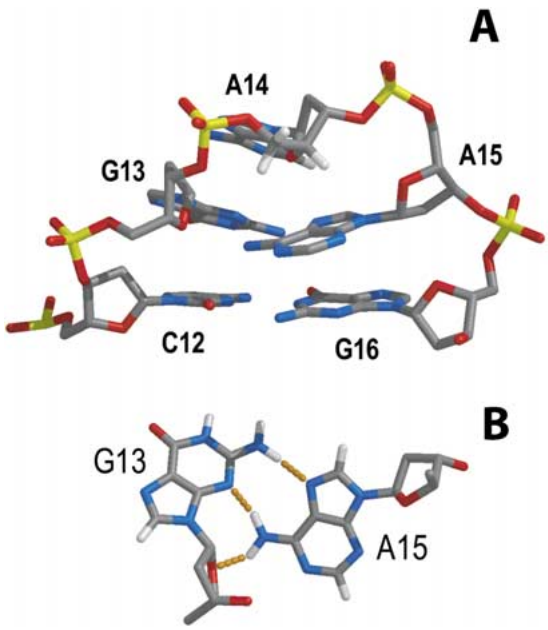

Figure 9. (A) Structure of the GAA loop and the closing GC base pair. (B) Structure of the sheared GA pair. Hydrogen bonds are shown in dotted lines.

of the molecule is defined even more precisely, with an average RMSD of 0.64 Å. The GAA loop and the closing GC pair are defined to a lesser extent, with RMSD of 0.83 Å. This is a relatively high value, taking into account that it is calculated for five residues only. To some extent, this is explained by fewer interresidue restraints in this region (Figure 3) and, in

addition, this may reflect the true dynamics of the loop (*vide infra*). These structures were refined using NOE data only. The precision of structure determination could be significantly improved for comparable systems, to RMSD of 0.33 Å, if NOE data are supplemented with residual dipolar coupling constants measured in partially oriented solutions (MacDonald et al., 2001; Tjandra et al., 2000). The force constant for the penalty term used in our refinement protocol with miniCarlo was rather low, 2 kcal/(mol Å^2), so the conformational energies of the refined structures were not significantly compromised. Yet the force constant was high enough to ensure convergence of the refinement, and low residual distance restraint deviation and NOE-based R^x-factors. These figures of merit (Table 1) are comparable to those for other DNA structures refined with similar methods (Mujeeb et al., 1993; Tonelli et al., 1998; Weisz et al., 1994). We must note, however, that agreement with the experimental data is less than perfect. For example, six interproton distances systematically deviate from experimental restraints by more than 1 Å; four of them involve terminal or loop residues, and the other two are intraresidue distances involving H3′ or H4′ protons. Possible reasons for these deviations are experimental errors, which tend to grow with the size of a molecule, and differential flexibility of parts of the molecule (*vide infra*).

Structure of the AT-rich core

Helical parameters describing the local geometry of refined structures were calculated with the Fitparam program and included as Supplementary Material. The stem portion of the 27mer has a conformation typical of other NMR structures of DNA oligonucleotides in solution (Ulyanov and James, 1995): it is, generally, B-DNA with some A-like features. The A-like features include positive roll parameter and decreased (compared to the crystalline B-DNA) slide parameter (Figure 5). Sugar conformations are predominantly C2′-endo and C1′-exo, although the sugar pucker of the terminal residue G_{27} is not well defined, and *average* sugar puckers of T_6, T_7, and T_{23} are intermediate between C1′-exo and O4′-endo (average pseudorotation phase angle 97–111°, see Supplementary Material). The latter conformation is energetically unfavorable; most likely, it is an averaging artifact due to a higher population of C3′-endo conformers for these residues. However, this could not be verified by analysis of scalar couplings $^3J_{HH}$ (Schmitz and James,

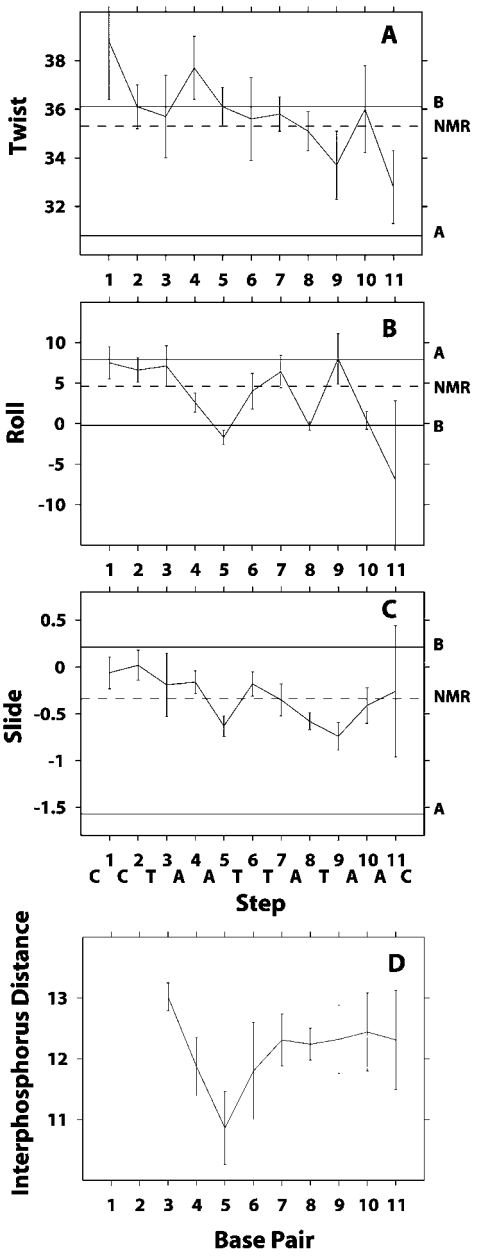

Figure 5. Selected local helical parameters for the 27mer calculated with the Fitparam program. The error bars show the standard deviation for the ten final structures. The horizontal lines labeled 'B' and 'A' show the average values observed in crystal structures of B-DNA and A-DNA, respectively (Gorin et al., 1995). The dashed horizontal line labeled 'NMR' shows the average values observed in solution NMR structures (Ulyanov and James, 1995). (A) Twist, degrees. (B) Roll, degrees. (C) Slide, Å. (D) Inter-phosphorus distance across the minor groove, Å. The line connecting phosphorus atoms across the minor groove spans three base pairs (see also Figure 8); the corresponding distance is assigned to the middle of these three base pairs. For example, inter-phosphorus distance assigned to base pair 5 spans base pairs $A_4{\cdot}T_{24}$, $A_5{\cdot}T_{23}$, $T_6{\cdot}A_{22}$, and it is measured between T_7P and A_{25}P. Generally, the n-th distance is measured between the phosphorus atoms in residues n+2 and 27-n+3.

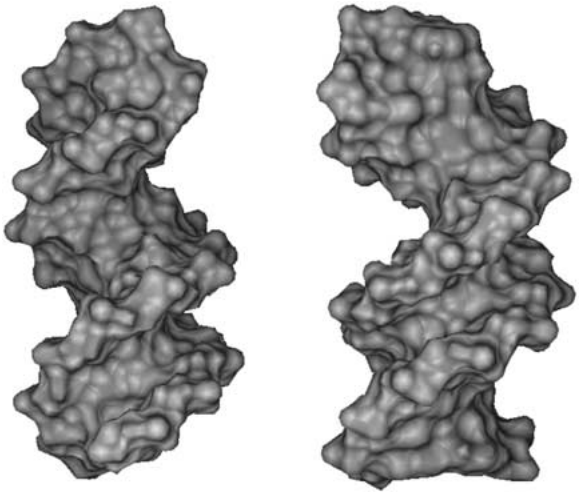

Figure 6. Surface representation of the final structure; the two views are related by a 90-deg rotation around the long axis. The surface was computed with the MSMS program (Sanner et al., 1996) and displayed with Chimera (Huang et al., 1996).

1995), because spectral overlap and large proton linewidth prevented us from quantitative measurement of such couplings. Similar to other AT-rich sequences, Watson–Crick AT pairs have high propeller twist parameters (Kerwood et al., 1991; Shatzky-Schwartz et al., 1997), from −10 to −22 °C (Supplementary material).

The local geometry of the 27mer stem is strongly sequence-dependent, and the variation is the most prominent for parameters roll, slide, twist (Figure 5) and buckle. Excluding two terminal GC pairs, the parameter roll has three maxima, at steps 3, 7 and 9 (Figure 5B), all of them TpA steps. Positive roll compresses the major groove of the double helix and opens its minor groove. Roll tends to be positive in pyrimidine–purine steps due to 'Calladine clashes' of purines in the minor groove (Calladine, 1982). Positive roll angle may lead to bending of DNA (Hagerman, 1984; Ulyanov and Zhurkin, 1984). However, for the 27mer, the TpA steps are spaced in such a way that their local bending cancels. As a result, the 27mer molecule is essentially straight (Figure 6). In contrast, roll angle has minima at the two ApT steps (Figure 5B). The opposite tendencies of the roll angle in ApT and TpA steps have been predicted by molecular mechanics calculations (Ulyanov and Zhurkin, 1984); these tendencies are also in agreement with other solution and crystal DNA structures (Gorin et al., 1995; Ulyanov and James, 1995).

The AT-rich core of the 27mer consists of nine contiguous AT base pairs. Its sequence includes three TpA·TpA, two ApT·ApT and three ApA·TpT steps. Each of the three classes of steps has its unique pattern of stacking interaction (Figure 7). ApT steps have the maximum amount of base overlap, and TpA steps have the least overlap. Chemical shifts of adenine H2 protons (filled circles in Figure 7) can be conveniently used to monitor the stacking interaction state in these steps. Proximity of H2 protons to an aromatic ring must cause an upfield shift of its resonances due to the ring current effect. According to the solution structure of the 27mer, the ring current effect must be strong for both adenines in TpA steps and for the first adenine in ApA steps, and it must be weak for both adenines in ApT steps and for the second adenine in ApA steps. These effects must add up for the middle adenine in TAA motifs. There are three TAA triplets in the 27mer sequence, $T_3A_4A_5{\cdot}T_{23}T_{24}A_{15}$, $T_9A_{10}A_{11}{\cdot}T_{17}T_{18}A_{19}$, and $T_{20}A_{21}A_{22}{\cdot}T_6T_7A_8$. Indeed, in all three cases, middle adenine H2 proton signals are shifted upfield to an unusual extent: 6.81, 6.69 and 6.37 ppm for A_4, A_{10} and A_{21}, respectively. Clearly, structural features of the 27mer correlate well with the observed proton chemical shifts; however, we did not attempt to calculate chemical shifts based on the refined structure (Case, 1998).

Chemical shifts are slightly different for the central AT pairs in each of the three TAA triplets, $A_4{\cdot}T_{24}$, $A_{10}{\cdot}T_{18}$, and $A_{21}{\cdot}T_7$ (see Supplementary Material). It is therefore clear that conformations of the three triplets must also be somewhat different. It is interesting to see if the current level of definition of NMR structures is sufficient to detect these differences. Indeed, at the level of helical parameters, there are differences between the three triplets. In particular, the $T_9A_{10}{\cdot}T_{18}A_{19}$ step has unusually low (for TpA steps) values of slide, -0.74 ± 0.15 Å, and twist, $33.7 \pm 1.4°$ (see Figure 5 and Supplementary material). It is not clear, if the local conformation of this TpA step is somewhat influenced by the proximity of the GAA loop. At the level of atomic RMSD, however, the structures of the three TAA·TTA triplets do not appear very different (Table 2). The difference between the conformations of the three triplets is 0.4–0.5 Å, but each individual triplet is defined to RMSD of 0.3–0.4 Å. Apparently, the current level of definition of the 27mer structure is only marginally sufficient to discern such small conformational differences reliably.

The 27mer sequence also contains a consensus sequence for the Pribnow box, TATAAT·ATTATA (residues 18–23 and 5–10). This fragment is defined within atomic RMSD of 0.50 ± 0.07 Å

Table 2. Atomic root-mean square deviations between three TAA·TTA triplets (Å)[a]

Triplet[b]	1	2	3
1	0.35 (0.10)	0.50 (0.13)	0.42 (0.12)
2		0.38 (0.11)	0.46 (0.10)
3			0.30 (0.09)

[a]Average RMSD values for the ten final structures; standard deviations are shown in parentheses. Only heavy atoms are taken into account.
[b]Triplet 1: $T_3A_4A_5{\cdot}T_{23}T_{24}A_{25}$, triplet 2: $T_9A_{10}A_{11}{\cdot}T_{17}T_{18}A_{20}$, triplet 3: $T_{20}A_{21}A_{22}{\cdot}T_6T_7A_8$.

(heavy atoms only). This sequence has been determined by NMR in this laboratory previously (Schmitz et al., 1992), in the context of a DNA octamer d(GTATAATG)·d(CATTATAC) at 16 °C. The RMSD value between the two independently determined Pribnow box structures is 0.70 ± 0.05 Å. This is a quite low value, taking into account different sequence context, minor differences in conditions and very different refinement protocols. This testifies to the general accuracy of the MARDIGRAS-based structure determination. At the same time, there are some differences between the two Pribnow box structures. Most notably, all TpA are bent in the direction compressing the major groove of DNA (roll parameter of 6–8°) in the 27mer; while in the octamer, the first TpA step has zero roll (see Appendix 1 in Ulyanov and James, 1995). Most likely, this discrepancy is due to the end proximity of this step in the octamer. Still, it is possible that there is a small bias in structure determination due to a particular force field and refinement protocol (see also Tonelli et al., 1998; Ulyanov et al., 1993).

Sequence-dependent variation of local helical parameters is connected to the minor groove width of the 27mer, which also strongly depends on sequence (Figure 6). The inter-phosphorus distance, measured across the minor groove, has a pronounced minimum around base pair 5 (Figure 5D). This minimum corresponds to the distance T_7P–$A_{25}P$ (Figure 8). In addition to high propeller twist, a number of helical parameters are connected to the formation of the narrow minor groove in AT-rich sequences; roll, twist and slide appear to be the most important. Low roll angle rotates the base pair about its long axis, compressing the minor groove. Low slide value shifts the top base pair to the left relative to the bottom pair (from the view shown in Figure 8), and this leads to a decreased span between cross-strand phosphorus atoms. High helical twist winds the double helix and it also decreases the inter-phosphorus distance. It is clear (Figure 8) that the geometries of three base pairs (4, 5 and 6) and two steps (4 and 5) are mainly responsible for the local dimensions of the minor groove in B-DNA. In this particular case, the following combination of helical parameters leads to the minimum of minor groove width around base pair 5 (see Figure 5). Roll has a minimum at step 5, and it is below average for step 4. Slide also has a minimum at step 5. Twist has a maximum at step 4, and it is above average at step 5. Understanding the rules governing the formation of the narrow minor groove may be important for designing ligands targeting specific DNA sequences (Bostock-Smith et al., 2001; Hawkins et al., 2001; Reddy et al., 2001). However, we must note that minor groove width is one of the less well-defined parameters in NMR structures based on NOE data (note the error bars in Figure 5D). In addition, the minor groove width may be sensitive to a particular force field or refinement protocol, because there are no direct experimental restraints constraining this parameter. This situation can be significantly improved when residual dipolar couplings are available as additional restraints (MacDonald et al., 2001; Tjandra et al., 2000).

Possible flexibility of the AT-rich core

Conventional refinement protocols, including the one used in this work, seek a structure with minimal weighted sum of a force-field energy and a penalty term based on experimentally observed parameters. Because of the time-average nature of NMR signals, such a structure corresponds to an average conformation in solution; the averaging could be very complex, however. An ensemble of structures typically calculated during NMR refinement (Figure 4) represents the precision of determination of such an average conformation. In other words, it shows the degree of indetermination of the average structure by the experimental data, and not the real conformational variations in solution, even though the two may sometimes coincide. One example of such a coincidence in the 27mer structure is the sugar pucker of the 3′-terminal residue G_{27}. Among ten final structures, 40% have this residue in the C2′-endo conformation, with the sugar pseudorotation parameter of 164 ± 7° (data not shown). In the remaining six structures, this residue is C3′-endo, with sugar pseudorotation of 17 ± 4°. The apparent average conformation, 75 ± 72°, was not actually populated at all. It is well documented that terminal residues in DNA duplexes are especially

flexible, being in dynamic equilibrium between C2′-endo and C3′-endo conformations (see, e.g., Lane et al., 1993; Schmitz et al., 1992). However, in our case, the sugar conformation of G_{27} is simply not defined sufficiently by NOE-derived distance restraints. Instead, the bimodal population of G_{27} sugar puckers was obtained largely due to the force field used during refinement, because the intermediate sugar conformations are unfavorable energetically (see, e.g., Olson, 1982).

If a molecule is flexible in solution, the calculated average conformation may actually be sparsely populated. For example, residues T6, T7 and T23 in the 27mer have average sugar pseudorotation parameters of 97 ± 27°, 111 ± 23° and 111 ± 25°, respectively (see Supplementary material). These sugar puckers, O4′-endo and low C1′-exo, have sub-optimal energies and, most probably, they also result from the dynamic C2′-endo/C3′-endo equilibrium with a higher population of C2′-endo. Such sugar puckers have been observed in other NMR structures as well. For example, a DNA duplex containing an A-tract has been solved with extra-high precision using residual dipolar coupling constants (MacDonald et al., 2001). Most pyrimidines in this structure have unfavorable O4′-endo sugar pucker (PDB entry 1FZX). Such a structure must be considered with caution, because the apparent sugar conformations may not be significantly populated in solution, but represent the average of C2′-endo and C3′-endo puckers (also see a discussion of this issue in Tjandra et al., 2000).

Another indication of flexibility of the AT-rich core of the 27mer comes from inspection of TpA steps. Adenine H2 protons have somewhat broadened linewidth in these steps, and they have a strong pre-melting temperature dependence (not shown). It has been shown that certain pyrimidine-purine sequences undergo a sub-millisecond motion in solution: TpA steps (Lane, 1989; McAteer and Kennedy, 2000; Schmitz et al., 1992) and CpA steps in the CAA context (Kojima et al., 2001). Similar to the situation with dynamic equilibria of sugar conformations, the helical structure of the 27mer must also be understood as an average conformation. While the total NOE-based R^x-factor has low values for the 27mer (Table 1), it has local maxima at the TpA steps when calculated for individual residues (not shown). Locally increased values of the R^x-factor may be due to internal inconsistencies in the experimental restraints, which are expected in flexible sites (Ulyanov et al., 1995).

The 1D imino proton spectra of the 27mer show a broad peak between 10 and 11 ppm at low ionic strength and/or low pH (Figure 1A). A possible interpretation of this peak is that it is due to thymine imino protons in AT pairs, which are transiently opened but still partially protected from solvent. In contrast to the previous examples, this transient opening must be slow on the NMR time scale, because open and closed states have separate NMR signals. While it is not possible to determine the nature of these open states based on our data, we note that according to molecular mechanics calculations, AT base pairs in DNA duplexes can form metastable structures stabilized by a single hydrogen bond, thymine O2–adenine NH_2 (Keepers et al., 1982). Also, according to quantum chemical calculations, partially opened AT pairs can be stabilized by interaction with water molecules (Kryachko and Volkov, 2001).

Structure and dynamics of the GAA loop

The structure of this loop in solution has been solved previously in the context of a DNA hairpin d(GCGAAGC) (Hirao et al., 1994; Yoshizawa et al., 1997), so we will give only a brief description here. In general, the loop is defined less precisely than the AT-rich core of the 27mer (Table 1). Nevertheless, many structural features of the loop are constant in all ten refined structures. In agreement with the published structure, the loop forms a very compact structure (Figure 9A). The loop is closed by a sheared pair $G_{13}{\cdot}A_{15}$, so that the loop per se consists of only one residue, A_{14}. The GA pair is stabilized by three hydrogen bonds, $G_{13}NH_2$–A_{15}N7, $A_{15}NH_2$–G_{13}N3, and $A_{15}NH_2$–G_{13}O4′ (Figure 9B). We do not have direct NMR evidence for any of these hydrogen bonds. The sheared configuration of the GA pair, however, has been demonstrated by Hirao et al. (1994) who studied the effects of 7-deaza substitutions in d(GCGAAGC). The GA pair is strongly buckled (Figure 9A) in all ten calculated structures, with the buckle angle of 36.1 ± 2.8°. This buckle has an effect on the neighboring GC pair: A15 pushes down the minor-groove edge of G16, which causes a very unusual positive propeller twist in $G_{16}{\cdot}C_{12}$, (Figure 9A). Adenine A_{14} is stacked on top of G_{13}, and sugar of A_{14} is stacked on top of A_{15}, which explains unusual upfield-shifted resonances of A_{14} H4′, H5′ and H5″. All residues in the loop have C2′-endo sugar puckers, and the backbone makes a sharp turn between A_{14} and A_{15}. However,

the backbone torsion angles in the loop are not defined well by the NOE-derived distance restraints.

In agreement with the previous report (Hirao et al., 1994), some resonances emanating from the loop residues have very broad linewidth, most notably G_{13} H8 and H1′ and A_{15} H8 and H1′. This is consistent with conformational exchange taking place for these residues, the nature of which is unknown presently. Interestingly, protons of A_{14} do not have linewidth as broad as in G_{13} and A_{15}, but there is independent evidence of motion for the A_{14} residue. The intensity of intraresidue cross-peak H1′–H8 is significantly higher for A_{14} than for other residues, which suggests that A_{14} may be part time in the syn conformation. A typical H1′–H8 distance is 3.9 Å and 2.5 Å for anti and syn conformations, respectively, and the MARDIGRAS-derived distance restraint for A_{14} H1′–H8 is 3.0–3.6 Å. This sort of internal inconsistency can be exploited using a multiple-copy refinement method (Görler et al., 2000), which aims to elucidate individual conformers and their probabilities. We are currently undertaking such refinement; preliminary results show that a 30–70% syn-anti equilibrium can explain the data, reducing the overall NOE-based R^x-factor by 10% for the ensemble of two structures. It is interesting that both anti and syn conformers have compact loop structures with fully stacked residue A_{14} (not shown). However, additional calculations are required to determine the uniqueness of such an ensemble.

Conclusions

We showed that stem-and-loop constructs with GAA loop could be successfully used to study DNA sequences with low thermal stability. Although properties of this tri-loop have been investigated quite some time ago (Hirao et al., 1994; Yoshizawa et al., 1997), this construct was not widely used since. Even though the structure of the loop itself is compromised by conformational averaging (the middle adenine *syn/anti* equilibrium is suspect), its main purpose is to stabilize the rest of the molecule. In addition, it has a characteristic NMR signature, which helps with proton assignments. The structure of the GAA loop perturbs the conformation of the closing CG base pair: strong buckle of the sheared G·A pair causes positive sign of propeller twist in CG pair. The perturbation appears to be only minor, if at all, beyond the closing pair.

We have determined a solution structure of DNA that includes nine consecutive AT pairs, the longest, to our knowledge, stretch of AT pairs solved to high resolution by NMR or X-ray crystallography. This structure displays a number of sequence-dependent features including a characteristic pattern of minor groove width variation, and local bending of each TpA step without the overall bending of the DNA helical axis. Cataloging such sequence–structure relationships is important for understanding such phenomena as sequence-directed bending of DNA, binding of drugs and proteins to DNA. AT rich sequences are especially underrepresented among DNA structures solved to high resolution (see, e.g., Ulyanov and James, 1995).

To solve this structure, we used very straightforward homonuclear experiments, but very elaborate computational procedures, which are described in detail above. Because of the extended shape of DNA duplexes, and of a relatively low proton density, a special care is required to determine accurate and high-resolution DNA structures based on NOE data. Two key elements are very accurate integration of NOE cross-peaks, which requires suitable software, such as SPARKY (Goddard and Kneller, 1998), and accurate determination of interproton distances using complete relaxation matrix method. The method routinely used in this laboratory includes the MARDIGRAS program combined with the RANDMARDI procedure accounting for the experimental and integration errors in NOE intensities (Borgias and James, 1990; Liu et al., 1995). Using internal coordinates based programs to refine DNA structures is also advantageous (see discussion of this issue elsewhere (Tonelli et al., 1998; Ulyanov et al., 1993)). Another approach proved successful recently is based on using residual dipolar coupling constants for structure determination (MacDonald et al., 2001; Tjandra et al., 2000). This approach produces significantly tighter structural ensembles, but it requires labeled DNA samples.

Lastly, the structure of the 27mer determined by NMR represents the average conformation of this molecule in solution, which may or may not be significantly populated. In particular, the GAA loop structure and some of the sugar conformations may be compromised by conformational averaging. The error bars obtained from the analysis of the ten refined structures represent the degree of indetermination of this average conformation, rather than the true flexibility in solution; this also relates to the structures refined with the residual dipolar coupling constants. Different methods are currently being developed to determine

structures of individual conformers (see, e.g., (Görler et al., 2000; Ulyanov et al., 1995)).

Acknowledgements

This work was supported in parts by Grants GM39247 and P41 RR01081 from the National Institutes of Health. The authors are grateful to Drs Anwer Mujeeb, Anne Savitt and Uli Schmitz for many helpful discussions, to Mr Peter Kissel for assistance in preparation of the oligonucleotides, to Dr Vladimir Basus for the help with experimental NMR and to the UCSF Computer Graphics Lab personnel for the help with computer graphics.

Supplementary material

One table with proton chemical shifts, one table with the mean values and standard deviations of the helical parameters for the ten refined structures, and one figure showing temperature dependence of resolved adenine H2 protons are available as supplementary material.

References

Baroudy, B.M., Venkatesan, S. and Moss, B. (1982) *Cell*, **28**, 315–324.

Baroudy, B.M., Venkatesan, S. and Moss, B. (1983) *Cold Spring Harb. Symp. Quant. Biol.*, **47**, 723–729.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–360.

Benham, C.J., Savitt, A.G. and Bauer, W.R. (2001), submitted.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Bostock-Smith, C.E., Harris, S.A., Laughton, C.A. and Searle, M.A. (2001) *Nucl. Acids Res.*, **29**, 693–702.

Boulikas, T. (1996) *J. Cell. Biochem.*, **60**, 297–316.

Calladine, C.R. (1982) *J. Mol. Biol.*, **161**, 343–352.

Carmona, M. and Magasanik, B. (1996) *J. Mol. Biol.*, **261**, 348–356.

Case, D.A. (1998) *Curr. Opin. Struct. Biol.*, **8**, 624–630.

Crothers, D.M. and Shakked, Z. (1999) In *Oxford Handbook of Nucleic Acids Structure*, Neidle, S., (Ed.), Oxford University Press, Oxford, pp. 455–470.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Delange, A.M. and McFadden, G. (1990) *Curr. Topics Microbiol. Immunol.*, **163**, 71–92.

Elmendorf, H.G., Singer, S.M., Pierce, J., Cowan, J. and Nash, T.E. (2001) *Mol. Biochem. Parasitol.*, **113**, 157–169.

Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R. (1988) *J. Mol. Graphics*, **6**, 13–27.

Goddard, T.D. and Kneller, D.G. (1998) *SPARKY*, Ver. 3, University of California, San Francisco.

Gorin, A.A., Zhurkin, V.B. and Olson, W.K. (1995) *J. Mol. Biol.*, **247**, 34–48.

Görler, A., Ulyanov, N.B. and James, T.L. (2000) *J. Biomol. NMR*, **16**, 147–64.

Gueron, M., Plateau, P. and Decorps, M. (1991) *Prog. Nucl. Magn. Reson. Spectrosc.* **23**, 135–209.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Hagerman, P.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4632–4636.

Hampsey, M. (1998) *Microbiol. Mol. Biol. Rev.*, **62**, 465–503.

Hawkins, C.A., Watson, C., Yan, Y., Gong, B. and Wemmer, D.E. (2001) *Nucl. Acids Res.*, **29**, 936–942.

Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. and Miura, K. (1994) *Nucl. Acids Res.*, **22**, 576–582.

Huang, C.C., Couch, G.S., Pettersen, E.F. and Ferrin, T.E. (1996) *Pacific Symp. Biocomputing*, **1**, 724.

Iyer, V. and Struhl, K. (1995) *EMBO J.*, **14**, 2570–2579.

James, T.L. (1991) *Curr. Opin. Struct. Biol.*, **1**, 1042–1053.

Keepers, J.W. and James, T.L. (1984) *J. Magn. Reson.*, **57**, 404–426.

Keepers, J.W., Kollman, P.A., Weiner, P.K. and James, T.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5537–5541.

Kerwood, D.J., Zon, G. and James, T.L. (1991) *Eur. J. Biochem.*, **197**, 583–595.

Kojima, C., Ulyanov, N.B., Kainosho, M. and James, T.L. (2001) *Biochemistry*, **40**, 7239–7246.

Kryachko, E.S. and Volkov, S.N. (2001) *Int. J. Quantum Chem.*, **82**, 193–204.

Lane, A.N. (1989) *Biochem. J.*, **259**, 715–24.

Lane, A.N., Bauer, C.J., Frenkiel, T.A. and Birchall, A.J. (1993) *Eur. Biophys. J.*, **22**, 135–43.

Liu, H., Spielmann, H.P., Ulyanov, N.B., Wemmer, D.E. and James, T.L. (1995) *J. Biomol. NMR*, **6**, 390–402.

MacDonald, D., Herbert, K., Zhang, X., Polgruto, T. and Lu, P. (2001) *J. Mol. Biol.*, **306**, 1081–1098.

Marini, J.C., Levene, S.D., Crothers, D.M. and Englund, P.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7664–7668.

McAteer, K. and Kennedy, M.A. (2000) *J. Biomol. Struct. Dyn.*, **17**, 1001–1009.

McClure, W.R. (1985) *Annu. Rev. Biochem.*, **54**, 171–204.

Mujeeb, A., Kerwin, S.M., Kenyon, G.L. and James, T.L. (1993) *Biochemistry*, **32**, 13419–13431.

Olson, W.K. (1982) *J. Am. Chem. Soc.*, **104**, 278–286.

Olson, W.K. and Zhurkin, V.B. (1996) In *Biological Structure and Dynamics.*, Sarma, R.H. and Sarma, M.H. (Eds.), Adenine Press, Schenectady, NY, pp. 341–370.

Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 784–788.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Reddy, B.S., Sharma, S.K. and Lown, J.W. (2001) *Curr. Med. Chem.*, **8**, 475–508.

Sanner, M.F., Olson, A.J. and Spehner, J.C. (1996) *Biopolymers*, **38**, 305–320.

Schmitz, U. and James, T.L. (1995) *Meth. Enzymol.*, **261**, 3–44.

Schmitz, U., Sethson, I., Egan, W.M. and James, T.L. (1992) *J. Mol. Biol.*, **227**, 510–531.

Shatzky-Schwartz, M., Arbuckle, N.D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T.E., Luisi, B.F. and Shakked, Z. (1997) *J. Mol. Biol.*, **267**, 595–623.

Smallcombe, S.H. (1993) *J. Am. Chem. Soc.*, **115**, 4776–4785.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.

Tjandra, N., Tate, S., Ono, A., Kainosho, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 6190–6200.

Tonelli, M., Ragg, E., Bianucci, A.M., Lesiak, K. and James, T.L. (1998) *Biochemistry*, **37**, 11745–11761.
Ulyanov, N.B. and James, T.L. (1995) *Meth. Enzymol.*, **261**, 90–120.
Ulyanov, N.B. and Zhurkin, V.B. (1984) *J. Biomol. Struct. Dyn.*, **2**, 361–385.
Ulyanov, N.B., Gorin, A.A., Zhurkin, V.B., Chen, B.C., Sarma, M.H. and Sarma, R.H. (1992) *Biochemistry*, **31**, 3918–3930.
Ulyanov, N.B., Schmitz, U. and James, T.L. (1993) *J. Biomol. NMR*, **3**, 547–568.
Ulyanov, N.B., Schmitz, U., Kumar, A. and James, T.L. (1995) *Biophys J.*, **68**, 13–24.
Weisz, K., Shafer, R.H., Egan, W. and James, T.L. (1994) *Biochemistry*, **33**, 354–366.
Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K. and Hirao, I. (1997) *Biochemistry*, **36**, 4761–4767.
Zhurkin, V.B., Lysov, Y.P. and Ivanov, V.I. (1978) *Biopolymers*, **17**, 377–412.
Zhurkin, V.B., Poltev, V.I. and Florentiev, V.L. (1980) *Mol. Biol. (Mosk)*, **14**, 882–895 (English translation).
Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. and Jernigan, R.L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7046–7050.