Conformational properties and thermodynamics of the RNA duplex r(CGCAAAUUUGCG)₂: comparison with the DNA analogue d(CGCAAATTTGCG)₂

Maria R. Conte[+], Graeme L. Conn[1,§], Tom Brown[1] and Andrew N. Lane*

Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [1]Department of Chemistry, University of Southampton, Southampton SO17 1BJ, UK

ABSTRACT

The thermodynamic stability of nine dodecamers (four DNA and five RNA) of the same base composition has been compared by UV-melting. The ΔG of stabilisation were in the order: r(GACUGAUCAGUC)₂ > r(CGCAA-ATTTGCG)₂ ≈ r(CGCAUAUAUGCG)₂ > d(CGCAAATT-TGCG)₂ ≈ r(CGCAAAUUUGCG)₂ > d(CGCATATATGCG)₂ ≈ d(GACTGATCAGTC)₂ > r(CGCUUUAAAGCG)₂ ≈ d(CG-CTTTAAAGCG)₂. Compared with the mixed sequences, both r(AAAUUU) and r(UUUAAA) are greatly destabilising in RNA, whereas in DNA, d(TTTAAA) is destabilising but d(AAATTT) is stabilising, which has been attributed to the formation of a special B′ structure involving large propeller twists of the A–T base pairs. The solution structure of the RNA dodecamer r(CGCAAAUUUGCG)₂ has been determined using NMR and restrained molecular dynamics calculations to assess the conformational reasons for its stability in comparison with d(CGCAAATTTGCG)₂. The structures refined to a mean pairwise r.m.s.d. of 0.89 ± 0.29 Å. The nucleotide conformations are typical of the A family of structures. However, although the helix axis displacement is ~4.6 Å into the major groove, the rise (3.0 Å) and base inclination (~6°) are different from standard A form RNA. The extensive base-stacking found in the AAATTT tract of the DNA homologue that is largely responsible for the higher thermodynamic stability of the DNA duplex is reduced in the RNA structure, which may account for its low relative stability.

INTRODUCTION

The conformation and solution properties of nucleic acids are strongly dependent on base-composition, sequence and chemical structure. In aqueous solution, DNA is usually in the B family of conformations, whereas under conditions of low water activity, it adopts the A form, which is also the preferred conformation of RNA in aqueous solution. However, extended tracts of adenines in DNA form a thermodynamically more stable structure that is stiffer than mixed-sequence DNA. This structure is characterised by high propeller twists of the A–T base pairs, leading to extensive base-stacking and bifurcated hydrogen bonds along the helix (1). It has been proposed that the stiffness of this structure is partly related to the spine of hydration in the minor groove (1,2).

In general, RNA is very much more stable than DNA (3–5), though the actual free energies of dissociation of duplexes vary greatly with sequence. Chemically, the important differences between DNA and RNA are the 2′-OH on the sugar in RNA, and the methyl group in dT (i.e. 5-methyl-dU). The difference in the chemistry of the sugars largely accounts for the quite different conformations of DNA and RNA in aqueous solution; the presence of the C2′-OH in RNA stabilises the C3′-endo sugar conformation whereas the 2′-deoxy sugars in DNA tend to be in the C2′-endo conformation. In contrast, the methyl group of dT seems to affect primarily the thermodynamic stability of DNA (6). The differences in geometry and chemistry between DNA and RNA also affect the hydration properties of the major and minor grooves, according to both X-ray crystallography (7–9) and NMR (10). These differences in chemistry and conformation presumably are responsible for the very different thermodynamic stability of DNA and RNA duplexes.

We are using a variety of techniques to understand the relationship between conformational properties and thermodynamic stability of nucleic acids. Whereas there are numerous X-ray structures of DNA in both the B and A forms, there are no high resolution solution structures of DNA in the A form, and few solution structures of RNA duplexes have been reported. We have chosen to study in detail the conformational properties of the RNA dodecamer r(CGCAAAUUUGCG)₂, for which the DNA analogue has been extensively studied both by X-ray diffraction and NMR d(CGCAAATTTGCG)₂ (11–13). We have determined the thermodynamic stability of the analogous DNA and RNA dodecamers, and of related dodecamers of identical composition.

MATERIALS AND METHODS

Materials

r(CGCAAAUUUGCG), r(GACUGAUCAGUC), r(CGCUU-UAAAGCG), r(CGCAUAUAUGCG), r(CGCAAATTTGCG),

d(CGCAAATTTGCG), d(GACTGATCAGTC), d(CGCTTTAA-AGCG) and d(CGCATATATGCG) were synthesised using phosphoramidite chemistry and purified by anion exchange HPLC on a Dionex Nucleopac Pa-100 column, followed by reverse phase HPLC as previously described (14). For NMR spectroscopy 112 A_{260} units of r(CGCAAAUUUGCG) were dissolved in 10 mM Na-phosphate, 100 mM KCl, pH 7 containing 0.2 mM EDTA and 0.1 mM DSS, annealed from 80°C and lyophilised. The sample was redissolved in 0.6 ml 90% H_2O:10% D_2O or 100% D_2O for NMR spectroscopy.

Methods

The duplex-to-strand transitions for the eight duplexes were measured in 1 M NaCl using the hyperchromicity as previously described (14,15). The thermodynamic parameters were then determined from the dependence of the melting temperature T_m on oligonucleotide concentration C_t according to the van't Hoff relation:

$$1/T_m = \Delta S/\Delta H - (R/\Delta H)\ln(C_t) \qquad \mathbf{1}$$

where ΔS and ΔH and R are the changes in entropy and enthalpy, respectively, and R is the gas constant. Errors on the parameters were determined by standard methods (16) and were typically ±5% for ΔH and ±6% for ΔS for an estimated standard deviation on T_m of ±0.5 K. Essentially the same results were obtained by non-linear regression to the untransformed equation:

$$T_m = \Delta H/[\Delta S - R\ln(C_t)] \qquad \mathbf{2}$$

^{1}H NMR spectra were recorded at 14.1 T on a Varian Unity NMR spectrometer and at 11.75 T on a Varian UnityPlus spectrometer. Phase-sensitive 2D NMR spectra were recorded using the hypercomplex method (17). Spectra in H_2O were recorded using the Watergate pulsed gradient method for solvent suppression (18) with acquisition times of 0.4 s in t_2 and 0.05 s in t_1. NOESY spectra were obtained using mixing times of 25, 50, 100 and 250 ms.

NOESY spectra in D_2O were recorded at 30°C with acquisition times of 0.7 s in t_2 and 0.06 s in t_1, with mixing times of 50, 100 and 250 ms. Two quantum filtered-COSY spectra were recorded at 30°C with acquisition times of 0.8 s in t_2 and 0.07 s in t_1. Data matrices were transformed as 16384 by 2048 complex points, using a Gaussian function for apodisation in both dimensions.

^{31}P NMR spectra were recorded at 9.4 T on a Bruker AM400 spectrometer as previously described (19).

Apparent rotational correlation times were determined from the driven truncated NOE experiments (20) using the Cyt and Uri H6–H5 with eight irradiation times from 30 to 800 ms as previously described (21,22). [31]P relaxation rate constants were determined at 30°C using standard methods as previously described (15,19). It has been shown that the CSA contribution in phosphodiesters dominates relaxation at field strengths of ≥9.4 T (19,23). The correlation time, τ, and the effective CSA, $\Delta\kappa_{app}$ were determined using a systematic search procedure as previously described (15).

Structure calculations

Cross-peak volumes in NOESY spectra were estimated by taking rows parallel to F2 and measuring the area of each peak in cross-section using the fitting routines within Felix 95.0 (Gaussian line-shape). The width in F1 was determined from one or more resolved cross-peak and the volume calculated as Area(F2) × width(F1). Volumes were then normalised to those of the Cyt and Uri H6–H5 cross-peaks. With well-resolved (10) and well digitised cross-peaks, the integrations are accurate and precise (estimated precision is ±10% of the volume). The normalised volumes for the base–sugar protons were then used to find glycosidic torsion angles with NUCFIT which takes into account spin diffusion, rotational anisotropy (24) and saturation effects (25). In the isolated spin-pair approximation, base–H1′ NOEs can only discriminate between *syn* and *anti* conformations about the glycosidic bond. Analysis of NOE time courses allows rather more precise determination of the glycosidic torsion angle. The value found can be considered as the median value of fluctuations of a magnitude typically observed in a free dynamics run (24). The torsion angles were then used as moderately tight restraints in the structure calculations (i.e. ±15°). This torsion is not well determined by base–H1′ distances alone, which can discriminate only between *anti* and *syn* conformations (24).

Other normalised volumes were extrapolated linearly back to zero mixing time, from which distances were calculated according to:

$$r = 2.44[v_0/v(cyt)]^{-1/6} \qquad \mathbf{3}$$

Upper and lower bounds were on the distances were set as ±0.3 Å for r < 2.7, ±0.5 Å for 2.7 < r < 3.6 and ±1 Å for r > 3.6. The short distances, which correspond to the strongest NOEs have relatively small error bounds as the size of the NOE requires a short distance. Note that between 2 and 2.6 Å, the NOE intensity would differ by as much as 5-fold, and as the calibration distance is 2.45 Å, this is far more than the experimental errors. Hence, this degree of tightness is in fact rather conservative. Lower bounds for the weaker NOEs were justified because no line broadening was observed for any of the peaks for which NOEs were measured, and the time-dependence showed no evidence of unusually rapid relaxation. The calculation of the NOEs for complete structures further justified the fairly loose bounds set for the weakest NOEs.

The angle δ was restrained based on estimates of coupling constants determined from the DQF-COSY spectrum (see below). Restraints on the backbone angle γ were obtained from qualitative assessment of the coupling between H4′ and H5′/H5″. In the g^+ rotamer, both couplings are small (~3 Hz), whereas in either of the other two rotamers, one coupling is large (~12 Hz) (26). We have estimated upper limits to the value $\Sigma_{4'} = {}^3J_{4'3'} + {}^3J_{4'P} + {}^3J_{4'5'} + {}^3J_{4'5''}$. In the C3′-endo conformation, ${}^3J_{4'3'} \approx$ 8–10 Hz (and see below). The width at half-height of the H4′ resonance, observed from H3′–H4′ and H1′–H4′ cross-peaks in NOESY spectra recorded with a digital resolution of 1.25 Hz/pt, provides an upper limit to $\Sigma_{4'}$ ($\Sigma_{4'}$ +H4′ line-width). Calculations (not shown) indicate that under our experimental conditions the natural width at half-height of H4′ should be ~2 Hz. A value of the width of <20 Hz is consistent only with the g^+ rotamer. In some cases it was possible to show that the H4′–H5′/H5″ correlations were very weak in either the DQF-COSY or DQ-COSY experiments, further confirming small coupling constants. With this information, it was possible to restrain γ to 60 ± 40° for 10 residues.

Table 1. Thermodynamic data for r(CGCAAAUUUGCG)$_2$ and related dodecamers

Oligomer	T_m (1 µM) K	$-\Delta G(298)$ kJ/mol	$-\Delta H$ kJ/mol	$-\Delta S$ kJ/mol/K
r(CGCAAAUUUGCG)$_2$	335(333)	81.6(76.9)	433(407.6)	1.179(1.11)
d(CGCAAATTTGCG)$_2$	334(330)	81.7(74.1)	442(406.7)	1.209(1.117)
r(CGCAAATTTGCG)$_2$	342(–)	101.2(–)	522.2(–)	1.412(–)
r(CGCUUUAAAGCG)$_2$	331(333)	67.1(75.2)	332(393)	0.889(1.07)
d(CGCTTTAAAGCG)$_2$	325(329)	60.8(70.9)	321(395)	0.873(1.087)
r(CGCAUAUAUGCG)$_2$	338(334)	100(78.8)	555(412.6)	1.527(1.12)
d(CGCATATATGCG)$_2$	328(326)	74.6(66.4)	442(373.3)	1.233(1.03)
r(GACUGAUCAGUC)$_2$	342(336)	106(89.1)	558(482.8)	1.516(1.321)
d(GACTGATCAGTC)$_2$	326(328)	73.3(61.6)	422(340.7)	1.170(0.937)

Thermodynamic parameters were measured from concentration-dependent UV melting curves in 1 M NaCl as described in the text. Values in parentheses were calculated using the nearest neighbour interaction model as described in the text. Errors estimated from the data were ~±3–5% for ΔH, ±4–6% for ΔS and ±2–4% for ΔG.

Other restraints were derived from spectra in H$_2$O. Thus, as all base pairs showed evidence of hydrogen bonding, the heavy atoms involved in the Watson–Crick pairs were restrained in the range 2.8–3.25 Å. Tight restraints were also used that corresponded to the strong NOEs observed between AC2H and the N3H of the paired U (2.7 ± 0.3 Å), and the analogous GN1H and CN4H(2) (2.7 ± 0.3 Å). Other distances involving exchangeable protons were restrained much more weakly, with typical limits of ±1 Å. This allows for leakage processes by exchange with solvent. The upper limit is the more important one in these instances, as the sequential NOEs are limited at the lower end by the van der Waals contacts between neighbouring base pairs.

Restrained MD calculations were carried out on Silicon Graphics Indigo workstations using DISCOVER (Molecular Simulations, San Diego) with the Amber force field, with a dielectric constant $\varepsilon = 4r_{ij}$ to simulate the effects of electrostatics, with no cutoffs on the non-bonded interactions. Additional calculations with $\varepsilon = r$ were also used. Although the structures differed slightly (but less than the pairwise r.m.s.d. for $\varepsilon = 4r$), they satisfied the experimental restraints equally well. However, the energies are dominated by the electrostatic component of the forcefield, and gave somewhat poorer van der Waals energies than for $\varepsilon = 4r$, where the different contributions are more evenly balanced. We consider the calculations with $\varepsilon = 4r$ to provide more reliable overall results. Calculations were started from A-RNA (10 times) and 30 structures generated by randomising the co-ordinates with a short free dynamics run at 1000 K starting from A-RNA (with different random number seeds). This latter procedure produced a wide range of initial structures, which barely resembled a double-stranded duplex. The initial structures were then refined as follows: (i) 1000 steps conjugate gradient restrained energy minimisation (ii) 30 ps rMD equilibration at 300 K (iii) 200 ps rMD sampling at 300 K and (iv) 1000 steps conjugate gradient energy minimisation. Force constants of 40 kcal/mol/Å^2 and 40 kcal/mol/rad^2 were used for distances and torsion restraints, respectively.

The criteria for accepting structures were: a large, negative potential energy comparable to that of energy minimised RNA (without restraints), good stereochemistry (no significant van der Waals violations and bond length and angle energies as low as energy minimised RNA) and a low (<1.5 kcal/mol) residual restraint energy, with no individual violations in excess of 0.1 Å or 1°.

Refined structures were analysed using InsightII (Molecular Simulations). Helical parameters were calculated using Curves version 5.1 (27).

RESULTS

Thermodynamic stability

In general, RNA duplexes are considerably more stable than their DNA counterparts (3,4). However, d(A)$_n$.d(T)$_n$ tracts tend to stabilise DNA compared with mixed sequences of the same composition (3). This has been attributed to the improved stacking possible in such sequences, and a stabilising spine of hydration in the minor groove (28). We have compared the stability of DNA and RNA duplexes of the same composition, but different sequences. For all of these oligonucleotides, the melting curves were monophasic, and the T_m increased with increasing concentration in the range from 2 to >100 µM. Figure 1 shows van't Hoff plots, which were analysed according to equation **1**. The thermodynamic parameters are collected in Table 1. The most stable duplex is the mixed-sequence RNA dodecamer, and the least stable is d(CGCTTTAAAGCG)$_2$. The overall ranking by ΔG is: r(GACU-GAUCAGUC)$_2$ > r(CGCAAATTTGCG)$_2$ ≈ r(CGCAUAU-AUGCG)$_2$ > d(CGCAAATTTGCG)$_2$ ≈ r(CGCAAAUUUGCG)$_2$ > d(CGCATATATGCG)$_2$ ≈ d(GACTGATCAGTC)$_2$ > r(CGCUU-UAAAGCG)$_2$ ≈ d(CGCTTTAAAGCG)$_2$. The mixed-sequence DNA dodecamer has a lower T_m than the d(AAATTT)-containing sequence, whereas the mixed-sequence RNA dodecamer is much more stable than the r(AAAUUU)-containing duplex. Hence, although a short d(A$_n$T$_n$) tract stabilises the DNA duplex, an analogous r(A$_n$U$_n$) tract substantially destabilises the RNA duplex. The net result is that the r(CGCAAAUUUGCG)$_2$ and d(CGCAA-ATTTGCG)$_2$ duplexes have similar thermodynamic stability. Furthermore, the r(CGCUUUAAAGCG)$_2$ and d(CGCTTTAAA-GCG)$_2$ are even less stable than the mixed RNA and DNA duplexes, respectively. However, the influence of the methyl groups is substantial, as the dodecamer r(CGCAAATTTGCG)$_2$ is 20 kJ/mol more stable than the uridine analogue. This is consistent with other results (6).

The thermodynamic parameters can be calculated using the nearest neighbour model (4,5). All of the calculated T_m values gave reasonable agreement with the observed T_m values (within
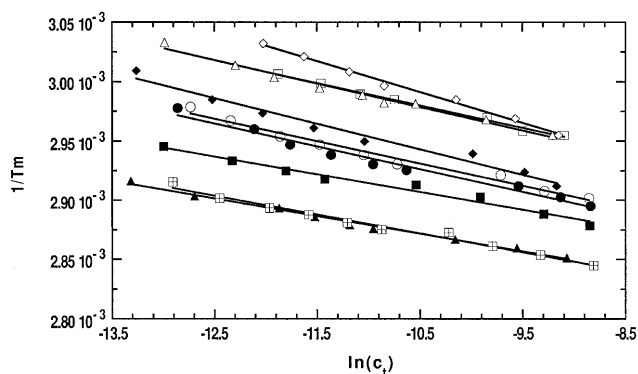
Figure 1. van't Hoff plots for the dodecamers.

the experimental limits). Although the calculated and experimental values of ΔG and ΔH showed similar trends, the agreement between the two was poorer. Significant differences were found for the r(UUUAAA) and d(TTTAAA)-containing sequences (Table 1). For these two sequences, the values of ΔG and ΔH were both much less negative than the predicted values, in contrast to the other sequence where ΔG and ΔH were more negative than the predicted values (Table 1). Nevertheless, the d(AAATTT) dodecamer is predicted to be more stable than the mixed DNA sequence, whereas the r(AAAUUU) dodecamer is predicted to be much less stable than the mixed sequence RNA dodecamer. The differences in stability are both predicted to be due to large changes in the enthalpy component, with an increase in enthalpy for the DNA and a decrease for the RNA. This would be consistent with substantial differences in base-stacking interactions (29). It is also clear from the data in Table 1 that runs of A_nU_n, but not alternating $(AU)_n$, are destabilising, regardless of the order of the A tract [i.e. $r(A_nU_n)$ or $r(U_nA_n)$]. This is in contrast to the DNA, where d(AAATTT) is stabilising, d(TTTAAA) is destabilising and d(ATATAT) is neutral compared with the mixed sequence. This correlates well with the known A-tract structure and stability in DNA. It seems probable that comparable rigid structures involving large propeller twists and bifurcated hydrogen bonds (1,28) are not formed in RNA (and see below).

NMR spectroscopy on r(CGCAAAUUUGCG)$_2$

To investigate the influence of conformation on thermodynamic stability, we have used NMR to determine the solution conformation and properties of r(CGCAAAUUUGCG)$_2$, which can be compared with the structures of the analogous DNA sequence determined both by X-ray crystallography (11) and NMR spectroscopy (12,13). Essentially all of the exchangeable protons and most of the non-exchangeable protons have been assigned, and in addition, C2′-OH resonances have been identified for the r(AAAUUU) dodecamer (10).

^{31}P NMR is useful for characterising the phosphodiester backbone, and determining the rotational correlation time from relaxation measurements. The ^{31}P NMR spectrum of r(CGCAAAUUUGCG)$_2$ showed a chemical shift dispersion of 0.69 p.p.m. For comparison, the shift dispersion in the DNA analogue of this sequence was $\Delta\delta = 0.6$ p.p.m. This range of shifts is typical of phosphodiester torsion angles in standard ranges (30,31). We have also measured the ^{31}P relaxation rate constants R_1 and R_2 at 30°C; the heteronuclear NOE was small (<1.05). The
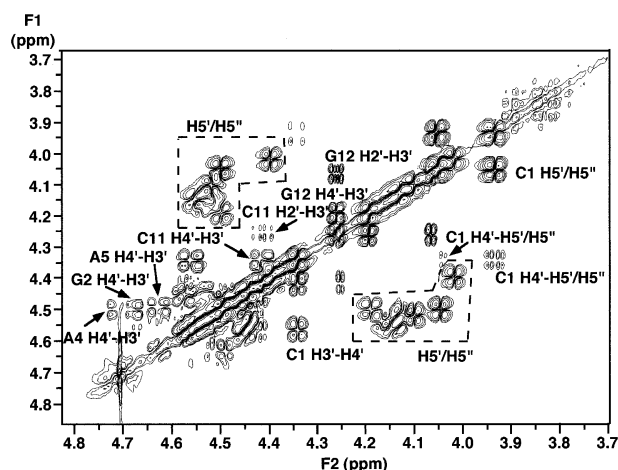

Figure 2. DQF-COSY spectrum of r(CGCAAAUUUGCG)$_2$. The spectrum was recorded at 30°C as described in Materials and Methods. The sugar region is shown.
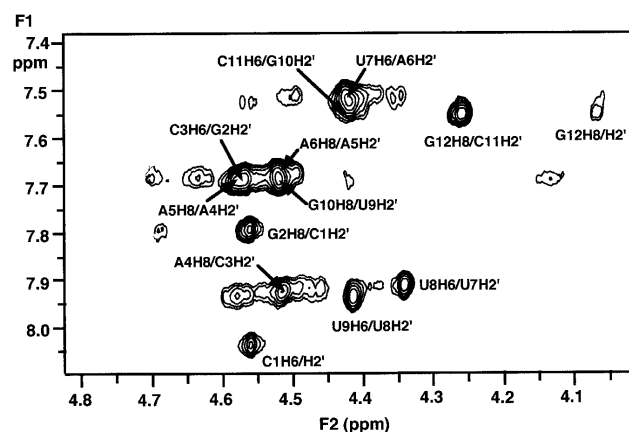

Figure 3. NOESY spectrum of r(CGCAAAUUUGCG)$_2$. The spectrum was recorded at 14.1 T and 30°C with a mixing time of 250 ms, showing the sequential base to H2′ NOEs.

mean value of τ determined from the relaxation measurements was 3.4 ± 0.2 ns. We have determined a similar value of 3.4 ± 0.1 ns at 30°C from the cross-relaxation rate constant of the Cyt and Uri H6–H5 vectors. The measured correlation times are as expected for a molecule of this size (15,22). The effective CSA, $\Delta\kappa_{app}$ was also reasonably well-determined at 158 ± 4 p.p.m., which is slightly larger than values reported for DNA duplexes determined in a similar fashion (~147 ± 7 p.p.m.), but is similar to another RNA duplex, 154 ± 6 p.p.m. (15). The slightly larger CSA found for RNA parallels the slightly greater chemical shift dispersion seen in RNA duplexes than DNA duplexes.

Solution conformation

The resolved H1′ resonances appear as relatively sharp singlets in 1D spectra (linewidth <2.5 Hz), and also in the NOESY spectra recorded with a digital resolution of 1.2 Hz per point. This places an upper limit of $^3J_{1'2'}$ of ~2 Hz for the non-terminal residues, indicating that the sugar conformations are in the N domain (i.e. near C3′-endo). Only the terminal residues showed H1′–H2′
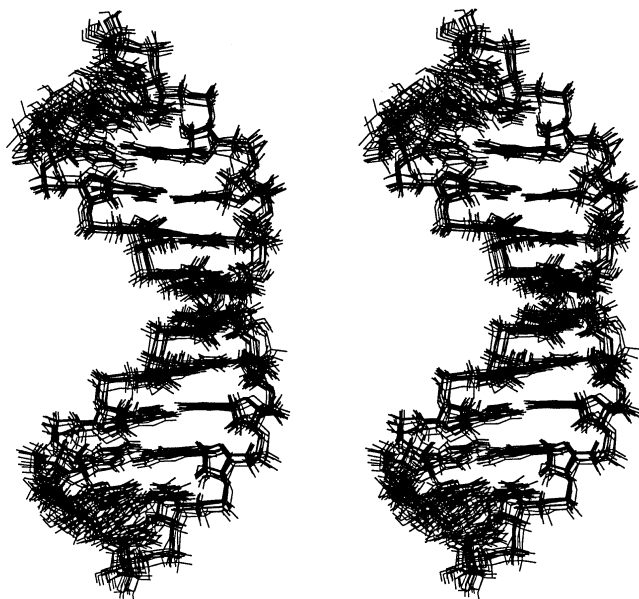
Figure 4. Structures of r(CGCAAAUUUGCG)$_2$. An overlay of the 10 best structures determined as described in the text are shown as stereo pairs.

cross-peaks in the DQF-COSY spectrum (Fig. 2). Further, the DQF-COSY spectra showed weak H2′–H3′ and strong H3′–H4′ cross-peaks, indicating that $^3J_{3'4'}$ is substantially larger than $^3J_{2'3'}$. Moreover, it was possible to measure $^3J_{3'4'} \approx 8$ Hz for some cross-peaks from the antiphase splitting in the DQF-COSY spectrum. These results confirm that the sugar conformations are near to C3′-endo. The range of P was estimated using the Karplus equation and the parameterisation for riboses (26). Thus, a value of $^3J_{1'2'} < 2$ Hz is consistent with $243° < P < 45°$, whereas $^3J_{2'3'} < 3J_{3'4'} \approx 8$ Hz implies $-18° < P < 54°$. Further, the H1′–H2′ NOE is more intense for N sugars than for S sugars, and the H1′–H4′ NOE reaches a maximum intensity of the O4′ conformation (P = 90°). Except for the terminal residues, the ratio of the H1′–H2′ to H1′–H4′ NOEs was consistent with N-type sugars. The combination of the measured coupling constants and NOE information allows ranges of $85 \pm 5°$ to be placed on the backbone angle δ, assuming $\phi_m = 36 \pm 6°$ (32). No additional endocyclic torsion angle constraints were used, so one degree of freedom remains to be determined by the structure calculations.

The intraresidue NOE intensities H8/H6 to H3′ > H2′ indicate that the glycosidic torsion angles are near –160°. NOE build-up curves were analysed using NUCFIT (24) as described in Materials and Methods, from which glycosidic torsion angles of $-160 \pm 15°$ were determined. The sugar pucker and glycosidic torsion angles are characteristic of nucleotides in the A conformation. This was confirmed by the very strong H2′(i) – H8/H6(i+1) cross-peaks in the NOESY spectra (Fig. 3) and weak H2′(i) – H8/H6(i) cross-peaks, which are characteristic of an overall A conformation. Furthermore, the CD spectrum was non-conservative and typical of the A form (not shown). Hence, the structure has characteristics of the A family.

Distance restraints were obtained by analysing the NOE time courses. We were able to determine 98 intraresidue distances and 132 sequential interresidue plus cross-strand distances (AH2–H1′, UN3H–AC2H, GN1H–CN4H1 and UN3H–

AN6H1). Based on NMR spectra in H$_2$O, we have maintained hydrogen bonding using restraints for the Watson–Crick base pairs, giving 30 distance constraints for the heavy atoms with a range of ± 0.3 Å. This accounts for a total of 260 distance constraints. With the constraints on δ from the coupling constants and analysis of the nucleotide conformations using NUCFIT(24), δ and χ could be restrained to fairly narrow ranges (80–90° and –145 to –175°, respectively). The dihedral angle γ of 10 residues could be restrained to the g$^+$ conformer ($60 \pm 40°$) (see Materials and Methods). Given the relatively wide spectral dispersion of the ^{31}P NMR spectrum, we have not applied any restraints on the backbone angles α, β, ε or ζ. In total, 324 conformationally sensitive experimental restraints (13.5 per residue) were used in the calculations, which is approximately twice the number of degrees of freedom in the system ignoring the finite sizes of the atoms. A further 140 NOEs were identified, but not used in the constraint list as they were already used in defining the torsion δ, are fixed distances (e.g. H6–H5 of U and C, H5′–H5′) or have no restraining power at the level of precision of the distance determinations (e.g. H2′–H3′, H3′–H4′, H4′–H5′/H5″).

Restrained MD calculations were run using the protocol described in Materials and Methods, starting from standard A-RNA and numerous A-RNA duplexes with different randomised torsion angles. Convergence was verified by examining the constraint energy and violations list, the total potential energy and the rms gradient of the energy, as shown in Table 2. Convergence was obtained to a pairwise r.m.s.d. of 0.89 ± 0.29 Å (Table 2) and an r.m.s.d. to the average of 0.6 ± 0.2 Å. The energy associated with bond length and angle deviations was small (<8 and <86 kcal/mol, respectively) and comparable to those found from energy minimised A-RNA (8.96 and 90.7 kcal/mol, respectively), indicating that good stereochemistry was maintained in these structures. Selected torsion angles from the best structures are given in Table 3, with the corresponding values for standard A-RNA and energy minimised A-RNA for comparison. As expected, the structures are all in the A family of conformations. The statistics given in Table 2 and 3 show that the structures are notably different from the standard A structure. The r.m.s.d. values to standard A and energy-minimised A-RNA were 2.2 ± 0.29 Å and 0.95 ± 0.31 Å, respectively. For comparison, the r.m.s.d. between standard and energy minimised A-RNA was 1.69 Å. The pairwise r.m.s.d. values for the structures are as good as one would expect for this density of constraints. We note also that the convergence was much improved with the inclusion of the nucleotide torsion angles γ. The 10 best structures are shown superimposed in Figure 4.

Table 2. Statistis of structure calculations for r(CGCAAAUUUGCG)$_2$

Structure	r.m.s.d. Å	U$_{pot}$ kcal/mol	U$_f$ kcal/mol
A3U3(A,ini)	2.21 ± 0.29	674.4	684.0
A3U3(A,min)	0.95 ± 0.31	–199.8	25.9
A3U3(fin)	0.89 ± 0.29	-200.8 ± 1.9	0.92 ± 0.19

Structures were calculated as described in Materials and Methods. r.m.s.d. values were calculated pairwise using the best 10 structures. The r.m.s.d. for A3U3(fin) is the value among all refined structures, the r.m.s.d. A3U3(A,ini) is between the standard A structure and the refined structures, and A3U3(A,min) is between energy-minimised A-RNA and the refined structures. U$_{pot}$ is the potential energy and U$_f$ is the residual constraint energy.

Table 3. Torsion angles in the refined dodecamer

Base	Torsion angles (°)			
	α	γ	δ	χ
C1	–	61 ± 0.6	78.4 ± 0.3	–154.2 ± 1.0
G2	83.9 ± 2.5	63.9 ± 1.3	78.2 ± 0.1	–160.6 ± 1.2
C3	–83.4 ± 2.2	65.3 ± 1.1	78.3 ± 0.3	–158.6 ± 1.3
A4	–81.6 ± 2.1	64.6 ± 1.4	78.3 ± 0.3	–162.9 ± 1.4
A5	–79.9 ± 2.3	68.8 ± 2.7	78.5 ± 0.2	–161.4 ± 2.6
A6	–79.3 ± 2.7	71.1 ± 1.5	83.0 ± 4.1	–169.3 ± 4.9
	155 ± 1.7	173.5 ± 4		
U7	–77.8 ± 1.9	64.8 ± 3	79.9 ± 1.6	–162.9 ± 1.5
U8	–78.3 ± 2	65.6 ± 2.7	81.3 ± 0.1	–167.2 ± 4.2
	152.1 ± 1.8	178.1 ± 1		
U9	–78.4 ± 2.7	63.5 ± 2.7	78.5 ± 2	–159.3 ± 2.7
G10	–79.4 ± 1.5	65.2 ± 2.1	78.8 ± 0.2	–161.9 ± 1.4
C11	–81.9 ± 2.3	64.0 ± 1	78.5 ± 0.2	–155.7 ± 2
G12	–81.9 ± 2	60.3 ± 1	79.5 ± 0.1	–156.2 ± 1
mean	–80.5 ± 2	64.8 ± 3	79.3 ± 1.4	–160.9 ± 4
A	–67	54	80	–160
A(EM)	–86	72	78	–164

Means calculated for nucleotides averaged over both strands.

The positions of the bases in the central core of the molecule (base pairs 3–10) are well determined (Fig. 4); much of the residual r.m.s.d. arises from poor definition of the terminal base pairs, where the density of constraints is lower than in the core. The glycosidic torsion angles and sugar puckers are in the range –150–170° and C3′-endo, respectively, which are typical of the A structure (Table 3). Because we have no direct restraints on ε, β or ζ, their values are determined largely by the force-field, and we do not consider them further. We have included the angle α because it is strongly correlated with γ (33). This is shown further by the results for the two nucleotides (A6 and U8) in which γ was not restrained. Two families of conformations were obtained for these residues. For example, for A6, γ, α = +71, –79 or +173/+155, whereas other torsion angles for the same phosphodiesters changed by <10°. It is clear that the backbone conformation at these two residues is not specified by the data. It also shows that a wide range of conformational space was sampled by the randomisation process (see Materials and Methods), and that the parameters for the other residues are determined largely by the experimental data. The variances of the δ and χ are comparable to the estimates on the experimental data. The refined structures gave values of γ and α that are different on

average from either the canonical A structure or the energy minimised A conformations (Table 3). Hence although the low variance on γ may in part arise from the forcefield, especially from the interplay of the experimental data and the Lennard–Jones energy, the experimental restraints must play a significant role in the determination of these correlated parameters. Quite small variations in torsion angles cause substantial variations in the positions of the nucleotides, especially the ribose moieties (Fig. 4). Nevertheless, the means are typical of the A family of conformations.

An alternative way to describe the structure is by helical parameters. We have calculated the helical twist, axial rise, base-pair inclination, displacement of the helix axis and the propeller twists (Table 4). The axis displacement of 4.6 Å into the major groove and the helical twist angles of ~31° are both characteristic of the A conformation. However, the axial rise (3.0 ± 0.3 Å) and base-pair inclinations (6 ± 2°) are quite different from either standard A-RNA or the energy minimised conformation (Table 3). The axial rise and base-pair inclination, of course, are not independent, as the separation between stacked base is 3.4 Å (31). The low base-pair inclination makes the observed axial rise approach that of the base–base separation. There are no obvious sequence-dependent variations within the present structures, with the exception that the inclination tends to approach the value expected for the A structures toward the ends of the duplex. The propeller twists are large, but are smaller on average than found in energy minimised A-RNA. Although the precision of the helical parameters in general is not high, it is clear that there are trends, and that the structure in the central r(AAAUUU) region of the duplex departs further from the standard A form more than the ends. Unfortunately, it is not clear how far into the duplex 'end effects' extend in RNA, and certainly the terminal base-pairs must be affected relatively more by the force-field as the density of constraints there is lower than in the core of the molecule.

The width of the minor groove is also characteristic of helix type. Because the positions of the phosphorus atoms are not specified in these structures, we have used the distance between C4′(i) and the cross-strand C5′(i+4) (minus 3Å) across the minor groove. In standard RNA this gives a minor groove width of 12 Å, which compares favourably with that determined from P–P separations (34). In the r(CGCAAAUUUGCG)₂ dodecamer, the minor groove width varies between 10 and 11 Å, with the narrowest sections in the centre of the molecule (i.e. the rAAAUUU tract). Although the groove is narrower than the standard RNA, it is more similar to that of energy-minimised RNA (9.9–10.8 Å). We note that this is much wider than in the d(AAATTT) tract of the DNA analogue, where the minor groove width is ~5 Å (Table 3).

Table 4. Helical parameters for r(CGCAAAUUUGCG)₂

Structure	Twist (°)	Rise (Å)	Inclination (°)	Pr. twist (°)	x-displacement (Å)	mgw (Å)
A-RNA	32.7	2.81	15.9	–14.4	–5.3	12
A3U3min	32.1	2.66	10.5	–22.5	–4.76	11.2 ± 0.6
A3U3f	30.8 ± 1.6	2.97 ± 0.3	6 ± 2	–20 ± 4	–4.58 ± 0.2	11.0 ± 0.6
dA3T3	36 ± 3	3.35 ± 0.015	0.4 ± 1.7	–13 ± 5	–0.51 ± 0.2	5.1 ± 0.5

Helical parameters were calculated over the central octamer using Curves v. 5.1. A-RNA is standard, unminimised A-RNA, A3U3min is the structure after energy minimisation without constraints, A3U3f is the refined structure with constraints and dA3T3 is the DNA analogue. mgw is the minor groove width.
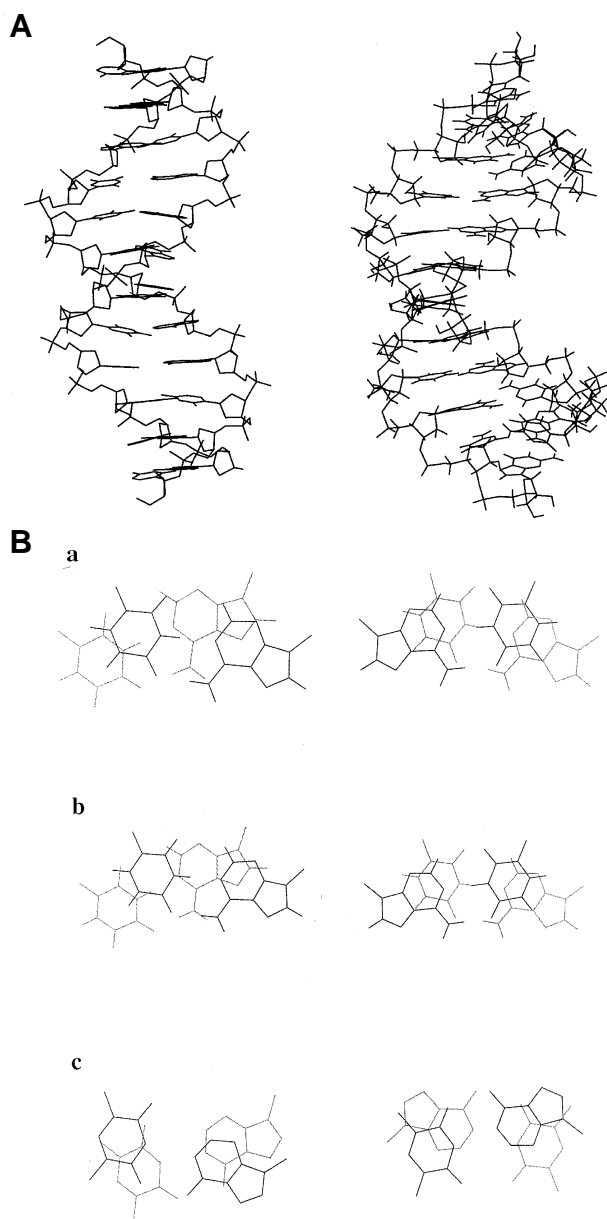
A

B   a

b

c

Figure 5. Base-pair overlaps: comparison of A3U3 with A3T3. (**A**) View into the major groove. (**B**) Base-stacking: (a) standard A: right A5.U8/A6.U7; left A6.U7/U7.A6; (b) A3U3 refined: right A5.U8/A6.U7; left A6.U7/U7.A6; (c) A3T3: right A5.T8/A6.T7; left A6.T7/T7.A6.

Figure 5 shows the base stacking in the central $r(A_3U_3)$ region, and in comparison with the homologous DNA structure, which has been examined in detail both by crystallography (11) and by NMR (12,13). A significant feature of the DNA duplex is the high propeller twist of the A–T base pairs, and the narrowed minor groove (Table 4). The high propeller twist gives rise to improved stacking of the bases, and the possibility of three-centre hydrogen bonding between adjacent base pairs (1). This has been associated with unusually slow exchange of the TN3H with solvent (35). In contrast, although the RNA dodecamer shows significant propeller twisting of the rA.rU base pairs, the low base-pair inclination decreases base stacking (34), and makes any possible bifurcated hydrogen bonds unstable (see below). The imino

protons of the RNA dodecamer showed exchange rates with water in the order G12N1H >> U7N3H ≳ U8N3H > U9N3H ≈ G2N1H > G10N1H (10), i.e. apart from the terminal residue, the exchange rates are faster in the centre of the A.U tract than toward the ends. In contrast, the imino protons in the DNA analogue showed exchange rates in the order: G12N1H >> G10N1H ≈ T9N3H > T8N3H > T7N3H (13). In addition to being generally smaller than the RNA rates under similar conditions, the exchange rates are slower in the centre of the A.T tract than toward the ends. This difference in behaviour would be consistent with a lack of bifurcated H-bonds in the RNA structure. No such H-bonds were found in these structures according to the distance and angle criteria used to describe the X-ray structure of the related dodecamer $r(CGCGAAUUGCGC)_2$ containing G.A mismatches (8). This structure is characterised by C3′-endo sugar puckers, glycosidic torsion angles in the range –150 to –170°, a small axial rise (~2.5 Å) and a large base-pair inclination (~18°), which is similar to the canonical A-RNA values (Table 3). Although the r(AAAUUU) structure in solution is similar at the level of the nucleotide conformations, the global structure is significantly different, most notably in the rise and inclination. However the NMR data do not agree with the canonical structure (Table 2). Whether the differences we observe can be attributed to the solution conditions or to the sequence differences cannot be determined at present.

DISCUSSION

In general, RNA is much more stable than the corresponding DNA species, and this has been shown to be largely the enthalpic contribution to ΔG (29), which is confirmed by the present results. As base-stacking contributes a major fraction of the enthalpy of melting of nucleic acids (29,34) it seems likely that the low stability of the r(AAAUUU) RNA duplex arises from a decreased enthalpy from poor base-stacking. The diminished base-stacking appears to arise primarily from the unusually low base-pair inclinations in the r(AAAUUU) tract (see above). Based on the data in Tables 1 and 3, we would predict that the alternating AU RNA sequence would have large positive base-pair inclinations, leading to extensive base stacking, and that the r(UUUAAA) sequence, which is even less stable than the r(AAAUUU) sequence, would also have low inclinations and a large axial rise in the r(UUUAAA) tract, leading to even poorer base stacking. The alternating r(AUAUAU) dodecamers have thermal stability similar to the randomised RNA duplex, and higher than either the r(AAAUUU) or r(UUUAAA) sequences (Table 1). The X-ray structures of two RNA duplexes containing such alternating sequences have been published which showed smaller rise and larger base-pair tilts than the present duplex, though the agreement between these two structures was not high, possibly because of the intermolecular interactions present in the crystal state (36,37). This is in agreement with the conformational model of the relative stability of such sequences.

It is notable that ΔG and ΔH for r(UUUAAA) and d(TTTAAA) are much lower than the predicted values, in contrast with the other sequences. This signals a likely failure of the nearest neighbour model, possibly because of long-range co-operative effects or unusual flexibility at the TpA junction in sequences of this kind (38).

Factors other than base-stacking must contribute to the stability of nucleic acid duplexes, such as electrostatics and ion condensation,

though this should contribute mainly via the entropic term (39,40), and hydration. The greater hydration of RNA compared with DNA has recently been proposed to account largely for the difference in the stabilisation enthalpy of DNA and RNA duplexes (41). However, the pattern of hydration in RNA seems to be quite similar in different RNA duplexes (8,37,41), which suggests that, in contrast to DNA (2), the hydration of RNA is relatively unaffected by the sequence and composition. We have recently shown that the r(AAAUUU) sequence is hydrated in the minor groove in solution, especially in around the C2′-OH, and that the pattern of hydration is quite different from that observed in the analogous DNA sequence (10,13). This suggests that hydration may contribute to the different stability of DNA and RNA. However, the observed differences in conformation must account at least in part for the observed difference in stabilisation enthalpy. This model of conformational differences correlating with thermodynamic stability is similar to that used for the A-tract DNA structures, and therefore represents a consistent framework for discussing the relationship between stability and conformation of nucleic acids.

We have shown that base-stacking is less favourable in r(AAAUUU) or r(UUUAAA) tracts than for mixed sequence RNA, leading to a lower enthalpy and therefore decreased net thermodynamic stability. This is despite the extensively hydrated minor groove.

ACKNOWLEDGEMENTS

REFERENCES

1 Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) *Nature* **330,** 221–226.
2 Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.-S., Kopka,M.L. and Dickerson,R.E. (1987) *Science* **238,** 498–504.
3 Lesnik,E.A. and Freier,S.M. (1995) *Biochemistry* **34,** 10807–10815.
4 Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
5 SantaLucia,J., Allawi,H.T. and Seneviratne,P.A. (1996) *Biochemistry* **35,** 3555–3562.
6 Wang,S. and Kool,E.T. (1995) *Biochemistry* **32,** 4125–4132.
7 Berman,H.M. (1994) *Curr. Opin. Struct. Biol.* **4,** 345–350.
8 Leonard,G.A., McAuley-Hecht,K.E., Ebel,S., Lough,D.M., Brown,T. and Hunter,W.N. (1994) *Structure* **2,** 483–494.
9 Portmann,S., Usman,N. and Egli,M. (1995) *Biochemistry* **34,** 7569–7575.
10 Conte,M.R., Conn,G.L., Brown,T. and Lane,A.N. (1996) *Nucleic Acids Res.* **24,** 3693–3699.
11 Edwards,K.J., Brown,D.G., Spink,N., Skelly,J.V. and Neidle,S. (1992) *J. Mol. Biol.* **226,** 1161–1173.
12 Jenkins,T.C., Brown,D.G., Neidle,S. and Lane,A.N. (1993) *Eur. J. Biochem.* **213,** 1175–1184.
13 Lane,A.N., Frenkiel,T.A. and Jenkins,T.C. (1997) *Biochim. Biophys. Acta.* **1350,** 189–204.
14 Ebel,S., Brown,T. and Lane,A.N. (1994) *Eur. J. Biochem.* **220,** 703–715.
15 Gyi,J.I., Conn,G.L., Lane,A.N. and Brown,T. (1996) *Biochemistry* **35,** 12538–12548.
16 Press,W.H., Flannery,B.P., Teukolsky,S.A. and Vetterling,W.T. (1986) *Numerical Recipes. The Art of Scientific Computing.* Cambridge University Press, ch 14.
17 States,D.J., Haberkorn,R.A. and Ruben,D.J. (1982) *J. Magn. Reson.* **48,** 286–292.
18 Piotto,M., Saudek,V. and Sklenar,V. (1992) *J. Biomol. Str.* **2,** 661–665.
19 Forster,M.J. and Lane,A.N. (1990) *Eur. Biophys. J.* **18,** 347–355.
20 Wagner,G. and Wüthrich,K. (1979) *J. Magn. Reson.* **33,** 675–680.
21 Lane,A.N., Lefèvre,J-F. and Jardetzky,O. (1986) *J. Magn. Reson.* **66,** 201–218.
22 Birchall,A.J. and Lane,A.N. (1990) *Eur. Biophys. J.* **19,** 73–78.
23 Williamson,J.R. and Boxer,S.G. (1989) *Biochemistry* **28,** 2819–2831.
24 Lane,A.N. (1990) *Biochim. Biophys. Acta* **1049,** 189–204.
25 Lane,A.N. and Fulcher,T. (1995) *J. Magn. Reson.B* **107,** 32–42.
26 Wijmenga,S.S., Mooren,M.M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed.), *NMR of Macromolecules. A Practical Approach.* IRL Press, Oxford, Ch. 8.
27 Lavery,R. and Sklenar,H. (1988) *J. Biomol. Struct. Dyn.* **6,** 63–91.
28 Dickerson,R.E. and Drew,H.R. (1981) *J. Mol. Biol.* **149,** 761–786.
29 Searle,M.S. and Williams,D.H. (1993) *Nucleic Acids Res.* **21,** 2051–2056.
30 Gorenstein,D.G. (1992) *Methods Enzymol.* **211,** 254–286.
31 Legault,P. and Pardi,A. (1994) *J. Magn. Reson.* **103B,** 82–86.
32 van Wijk,J., Huckriede,B.D., Ippel,J.H. and Altona,C. (1992) *Methods Enzymol.* **211,** 286–306.
33 Schmitz,U., Ulyanov,N.B., Kumar,A. and James,T.L. (1993) *J. Mol. Biol.* **234,** 373–389.
34 Saenger,W. (1984) *Principles of Nucleic Acid Structure.* Springer Verlag, New York, ch.17.
35 Leroy,J.–L. and Guéron,M. (1996) *Methods Enzymol.* **261,** 383–413.
36 Dock-Bregson,A.C., Chevrier,B., Podjarny,A., Moras,D., deBaer,J.S., Gough,G.R., Gilham,P.T. and Johnson,J.E. (1988) *Nature* **335,** 375–378.
37 Wahl,M.C., Ban,C., Sekharudu,C., Ramakrishnan,B. and Sundaralingam,M. (1996) *Acta Cryst.* **D52,** 655–667.
38 Lane,A.N. and Lefévre,J.-F. (1994) *Methods Enzymol.* 596–619.
39 Record,M.T.,Jr, Anderson,C.F. and Lohman,T.M. (1978) *Q. Rev. Biophys.* **11,** 103–178.
40 Manning,G.S. (1978) *Q. Rev. Biophys.* **11,** 179–246.
41 Egli,M., Portmann,S. and Usman,N. (1996) *Biochemistry* **35,** 8489–8484.