Solution Structures of DNA·RNA Hybrids with Purine-Rich and Pyrimidine-Rich Strands: Comparison with the Homologous DNA and RNA Duplexes[†,‡]

Jeffrey I. Gyi,[§] Andrew N. Lane,*[,§] Graeme L. Conn,[||,⊥] and Tom Brown[||]

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

Received August 11, 1997; Revised Manuscript Received October 28, 1997[⊗]

ABSTRACT: The structures of d(GAAGAGAAGC)·d(GCTTCTCTTC), d(GAAGAGAAGC)·r(GCUUCUCUUC), r(GAAGAGAAGC)·d(GCTTCTCTTC), and r(GAAGAGAAGC)·r(GCUUCUCUUC) have been determined in solution from NMR data. Globally, the pure DNA and RNA duplexes were in the B and A forms, respectively. The two DNA·RNA hybrids were neither A nor B, but closer globally to the A than the B form. However, the thermodynamically less stable d(GAAGAGAAGC)·r(GCUUCUCUUC) duplex has a significantly different conformation from r(GAAGAGAAGC)·d(GCTTCTCTTC). Structures were calculated based on the NMR data, using restrained molecular dynamics. A new approach to the treatment of conformational averaging based on *a priori* probabilities has been used. The nucleotides were treated by fitting the scalar coupling data and NOE time courses to a two-state model comprising N and S sugar puckers each with a different glycosidic torsion angle, and the mole fraction of the S state. Restraint sets for different distributions of N and S states within molecules were constructed, such that each nucleotide was weighted in the ensemble according to the mole fractions (or *a priori* probabilities). The individual nucleotide conformations were strongly restrained, whereas the internucleotide restraints were set relatively loosely. Ensembles of conformations were generated and assessed by comparison of the NOEs calculated from ensemble-averaged relaxation matrices with the experimental NOEs. The ensemble averages accounted for the experimental data much better than any individual member, or for structures calculated assuming a single unique conformation. The two hybrids populated different degrees of conformational space. There was a general trend in minor and major groove widths in the order d(GAAGAGAAGC)·d(GCTTCTCTTC), d(GAAGAGAAGC)·r(GCUUCUCUUC), r(GAAGAGAAGC)·d(GCTTCTCTTC), r(GAAGAGAAGC)·r(GCUUCUCUUC) and a similar progression in global character from B-like to A-like structures. Furthermore, r(GAAGAGAAGC)·d(GCTTCTCTTC) showed a greater dispersion of conformations in the ensemble than d(GAAGAGAAGC)·r(GCUUCUCUUC), reflecting the greater flexibility of the sugars. If conformational averaging of the nucleotides is ignored, incorrect virtual structures are produced that nevertheless are able to satisfy a substantial fraction of the experimental data.

DNA·RNA hybrids are important intermediates in transcription (*1*), normal replication of double-stranded DNA (*2*), and reverse transcription by retroviruses (*3*). The RNA moiety is specifically degraded by the ubiquitous RNase H (*4*), which is also an integral activity of reverse transcriptase (*3*). There have been several reports that the thermodynamic stability of duplexes having the general composition rR·dY is substantially greater than that of dR·rY duplexes (*5*−*9*). We have recently shown that the global and local conformations of these hybrids are also substantially different (*9*). This has implications for antisense technology, as the activity of RNaseH potentiates the antisense activity of a probe DNA strand, by degrading the target mRNA, and releasing the DNA.

The solution conformations of several DNA·RNA hybrids of mixed sequence have been reported, and it is clear that the RNA strand is locally in the A conformation, whereas the DNA strand has nucleotide conformations more similar to that found in the B conformation (*9*,*10*−*16*). However, overall the hybrid duplexes are globally closer to the A form than the B form, albeit with a narrower minor groove than RNA. This latter feature has been proposed to be important for the activity of RNaseH (*10*,*11*), which will not degrade double-stranded RNA (*4*). However, it has been shown that the enzyme binds RNA more tightly than DNA (*17*).

We have recently completed the ^{1}H NMR assignments and analysis of the nucleotide conformations in d(GAAGAGAAGC)·d(GCTTCTCTTC) (denoted dR10·dY10), d(GAAGAGAAGC)·r(GCUUCUCUUC) (dR10·rY10), r(GAAGAGAAGC)·d(GCTTCTCTTC) (rR10·dY10), and r(GAAGAGAAGC)·r(GCUUCUCUUC) (rR10·rY10) (9). The DNA and RNA homoduplexes had high degrees of conformational purity in their sugar conformations, whereas the deoxyriboses in the two hybrids showed substantial

[†]This work was supported by the MRC of the UK and Oswel Research Products Ltd.

[‡] The following structural coordinates have been deposited in the Brookhaven Protein Data Bank: dR10·dY10 (filename 1axp, restraint rlaxpmr), rR10·rY10 (1rrr, rlrrrmr), rR10·dY10 (1rrd, r1rrdmr), and dR10·rY10 (1drr, r1drrmr).

[§] National Institute of Medical Research.

[||] University of Southampton.

[⊥] Present address: Department of Chemistry, Johns Hopkins University, East Charles St., Baltimore, MD.

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1997.

S0006-2960(97)01971-5 CCC: $15.00
Published on Web 01/06/1998

conformational averaging. The latter therefore require a more detailed analysis of the experimental data to be used for determining the structures (*15*,*18*−*21*). In this article we report the determination of the solution structures of the four duplexes based on NMR data. Because of the substantial conformational averaging in the nucleotides, we have developed a method that generates an ensemble of conformations that accounts for all of the experimental data and which reflects the populations of conformers known to be present in solution. This method of describing a conformation ensemble is compared with other methods in the literature.

EXPERIMENTAL SECTION

Oligonucleotides were prepared as described previously (*9*). NMR spectra were recorded at 11.75 T and 14.1 T on Varian UnityPlus and Unity spectrometers, respectively. NOESY spectra were recorded using the hypercomplex method (*22*) at several mixing times from 50 ms to 250 ms in D_2O at 25 to 35 °C. NOESY spectra in H_2O were recorded at 10 °C using Watergate (*23*) for solvent suppression. Cross-peak volumes were measured using Felix (v 95.0). Because of the significant spectral overlap for the pyrimidines, we have used the Gaussian line-fitting function on cross-sections through each cross-peak in both F1 and F2. The volume was taken as proportional to the product of the peak areas divided by the average of the two peak heights. Cross-peak volumes were then normalized to the cross-peak volumes of the average of resolved cytosine or uridine H6−H5 cross-peaks. This procedure was verified for resolved peaks using the volume integration with footprints within Felix. Generally the two methods agreed within 10% for the normalized peak intensities.

Structure Calculations. Nucleotide conformations were determined as described previously (*9*) assuming a two-state model in which there are two conformations in equilibrium described by a mole fraction, f_s, of the more populated state. One state is described by a nucleotide conformation where the pseudorotation phase angle *P* corresponds to the S range of sugar puckers, and the glycosidic torsion angle χ appropriate to an S sugar conformation. This state is denoted *P*(S) χ (S). The other state consists of an N-type sugar pucker and a different glycosidic torsion angle; this state is denoted *P*(N) χ (N). From the best fit to the NOE time-courses, precise distance constraints for both N- and S-type conformations were calculated for each nucleotide. Initial structures for the four oligomers were then constructed within InsightII (Molecular Simulations Inc., San Diego), using both A and B form for the DNA duplex and the two DNA·RNA hybrids, and standard A form for the RNA duplex. For the DNA duplex and the RNA strands of the hybrids and the RNA duplex, relatively tight distance constraints were used for inter-residue sequential interactions, as derived from the time-course of the NOEs (e.g. H1′(*i*), H2′(*i*)−H8/H6(*i*+1)) as these nucleotides showed little evidence of conformational averaging. For the DNA strands of the two hybrids, where a mixture of "N" and "S" state sugar puckers was found (see below), initially loose inter-residue constraints were applied. Standard distances (±0.3 Å) between the donor and acceptor atoms of the hydrogen bonds in the Watson−Crick base-pairs were used (i.e. three distances for GC base-pairs and two for AT(U) base-pairs). No additional planarity constraints were included. Additional cross-strand distance restraints were applied for T(U)N3H−AC2H and GN1H−CN4H(2) in the T(U)·A and G·C base-pairs, respectively, corresponding to observed NOEs. Weak distance constraints for observed NH(*i*)−NH(*i*+1) NOEs were also applied. Strong AC2H(*i*)−H1′(*i*+1) NOEs (sequential and cross-strand) were observed which were converted into distance constraints. Glycosidic torsion angles χ, sugar torsion angles δ, and where available γ restraints were also included. For the DNA strands in the hybrids, N-type sugar puckers were maintained where required by restraining the sugar ring torsions υ_0 and υ_1. For distances, the force constant was 50 kcal mol^{-1} Å^{-2} and for torsions it was 40 kcal mol^{-1} rad^{-2}. Typically, 11−13.5 nontrivial constraints per residue were used in the calculations. No artificial backbone constraints based on phosphorus chemical shifts were used, nor intra-sugar distance constraints that have no restraining power. The initial models were partially randomized during a brief (500 ps) period of unrestrained dynamics at 600 K. The constraints were then applied, and the system was subjected successively to restrained MD at 600 K (10 ps) followed by energy minimization, rMD[1] at 300 K (10−20 ps) and conjugate gradient energy minimization (2000 steps) using Discover (Molecular Simulations Inc., San Diego) with the AMBER forcefield with no cutoffs for nonbonded interactions and a distance-dependent dielectric constant ($\epsilon = 4r$). Slightly different protocols were employed to enable conformational space to be searched and to assess the influence of the forcefield on the converged structures. These included variation of the slope of the distance-dependent dielectric constant and scaling of the nonbonded interactions during the MD run from 10% to the full forcefield. Convergence was assumed to have been achieved when the residual restraint force fell to less than 1 kcal mol^{-1}, the rms gradient of the force was less than 0.0001 kcal mol^{-1}, and the total potential energy of the system was low (i.e. negative, with large negative contributions from the nonbonded interactions).

Ensembles of structures were calculated as follows. For each hybrid, ten structures were generated in which tight intranucleotide restraints were applied according to the analysis of the sugar puckers and nucleotide conformations. Individual nucleotides in each structure were randomly chosen so that restraints corresponding to the N or S domain were applied, with the requirement that in the ensemble of ten structures, each nucleotide on average would occupy the N or S conformation according to the derived populations of these states (see Supporting Information). Each structure was then refined using rMD, using 5 initial structures from each of the A and B family of conformations, for a total of 100 starting structures for each hybrid duplex. Rmsd values for coupling constants were calculated as

$$\mathrm{rmsd}(J) = [(1/N)\sum(J_{\mathrm{obs}} - J_{\mathrm{calc}})^2]^{1/2} \qquad (1)$$

where *N* is the number of observations.

It is implicitly assumed in these calculations that the nucleotides are adequately described as a two-state equilibrium, which is justified on the basis of population analysis of MD trajectories (*15*,*24*) and the available experimental

[1] Abbreviations: rMD, restrained molecular dynamics; rmsd, root mean squared deviation.

Table 1: Statistics of Structure Refinements

	structure	no. restraints per residue	⟨U⟩, kcal/mol	⟨Uf⟩, kcal/mol	⟨rmsdf⟩, Å	rmsdB	rmsd, Å
1	dR10·dY10	13.45	−152.2 ± 6.2	0.45 ± 0.02	0.82 ± 0.22	2.37 ± 0.15	4.1 ± 0.32
2	dR10·rY10[e]	11.5	−155.7 ± 2.1	0.86 ± 0.13	1.03 ± 0.19	nd	1.55 ± 0.24
	dR10·rY10[f]	12.0	−138 ± 1	7.5 ± 0.07	0.80 ± 0.1	3.18 ± 0.38	1.78 ± 0.22
	dR10·rY10[g]	12.5	−169.7 ± 8.4	1.13 ± 0.8	1.07 ± 0.3	3.40 ± 0.4	1.59 ± 0.13
3	rR10·dY10[f]	11.2	−154.3 ± 3.4	3.97 ± 0.046	0.78 ± 0.11	2.51 ± 0.38	2.38 ± 0.24
	rR10·dY10[g]	13.0	−160.9 ± 8.2	2.66 ± 1.6	1.07 ± 0.2	2.98 ± 0.27	1.81 ± 0.17
4	rR10·rY10	11.2	−162.0 ± 2.2	0.49 ± 0.14	0.89 ± 0.13	nd	1.81 ± 0.06

[a] ⟨U⟩ is the potential energy averaged over 10 structures. [b] ⟨Uf⟩ is the residual forcing potential. [c] ⟨rmsdf⟩ is the pairwise rmsd between 10 structures. [d] rmsdB, rmsdA are the rmsd between the refined structures and energy-minimized B and A structures, respectively. [e] No δ restraints. [f] δ restrained to 83−93°. [g] Ensemble averages.

data. It is further assumed that the nucleotide conformations are essentially independent of their nearest neighbors, i.e. they are only weakly coupled. This is justified by observation of MD trajectories which in general show little if any correlation between the sugar repuckering of neighboring nucleotides (*24*). Hence, exhaustive sampling of the 1024 conformers in the ensembles is not required, as in essence the averaging (within the context of the current modeling) is local.

The ten best structures were accepted using the criteria of a large negative total potential energy comparable to that of an energy-minimized standard structure, the stereochemistry was acceptable as shown by the energy of bond length and angle deviations being comparable to that of an energy-minimized structure, the residual restraint violation energy was small (<1.5 kcal mol^{-1}) and no violations larger than 0.l Å (distances) or >1° (torsions).

RESULTS

The nucleotide conformations of the four duplexes were described previously (*9*). Briefly, the nucleotide conformations for the DNA duplex are typical of the B conformation ($\chi \approx -110°$, $P \approx 144-162°$) whereas those of the RNA duplex were typical of the A conformation ($\chi \approx -160°$, $P \approx 9°$).The conformations of the ribonucleotides in the two hybrids were also typical of the A form, whereas the deoxyribonucleotides showed substantial conformational averaging at the nucleotide level, though the major conformer seemed to be close to the B form. We have calculated precise restraints for the major conformer of each nucleotide as described in the Experimental Section. Internucleotide NOEs were then converted into distance constraints, which initially were set loosely according to apparent strength: s(trong) (<2.6 Å), m(edium) (<3.2 Å), and w(eak) (3−5 Å). For a given duplex, typically 250−350 dipolar interactions were observed. However, given the precision of the experimental data, and uncertainties due to rapid dynamics and conformational averaging, many of the distances (approximately 30%) derived from these interactions have no restraining power and were therefore not included in the restraint list. Thus, distances which vary less than 0.3 Å between extremes of conformations were not directly restrained, e.g. intrasugar distances H1′−H2″, H2′−H3′, H3′−H4′ which are determined by the covalent structure of the sugar rings. Typically 200−250 independent distance restraints were obtained for each duplex (10−12.5 per residue) of which half were intranucleotide, and 25 H-bonds and 20 cross-strand distances. In addition, some γ torsion angles were obtained from estimates of $\sum_{4'}$ and NOE intensities between H5′/H5″ and H8/H6.

Table 2: Helical Parameters Averaged over Central Octamer. Helical parameters were determined using CURVES v.5.1 (*36*)

molecule	twist, deg	rise, Å	inclination, deg	axis displacement, Å
dR10·dY10	35 ± 0.93	3.04 ± 0.12	−2.2 ± 0.5	−2.5 ± 0.1
B-DNAmin[a]	34 ± 1.3	3.0 ± 0.12	0.02 ± 0.7	−2.7
dR10·rY10[b]	31.2 ± 1.0	2.9 ± 1.4	0.96 ± 0.5	−4.08 ± 0.04
dR10·rY10[c]	31.7 ± 2	2.82 ± 0.2	0.5 ± 0.9	−3.76 ± .05
rR10·dY10[b]	31.4 ± 1.2	2.97 ± 0.17	0.4 ± 0.9	−3.93 ± 0.14
rR10·dY10[c]	33.7 ± 1.9	2.9 ± 0.12	5.09 ± 1.0	−3.32 ± 0.06
rR10·rY10	31 ± 1.6	2.63 ± 0.2	8.1 ± 0.6	−5.16 ± .05
A-RNAmin[d]	30.5 ± .5	2.7 ± 0.13	7.1 ± 0.6	−4.66 ± .02

[a] BDNAmin is unrestrained energy-minimized B-DNA. [b] Ensemble average. [c] Single structure refinement. [d] A-RNAmin is unrestrained energy-minimized A-RNA.

dR10·dY10. 269 experimental restraints (13.45 per residue) were used for the rMD calculations. The DNA duplex refined to the expected B conformation starting from either standard B or standard A, with no violations, a negative energy, and a residual restraint energy of 0.45 kcal mol^{-1} (Table 1). There were no distance violations >0.1 Å and no angle violations of >1°. The rms deviations between the initial and final structures, and the two final structures are also given in Table 1. The final structures agree within about 0.8 Å for all atoms, indicating that the restraint data reasonably describe the conformation. The rmsd values from the starting structures and from energy-minimized A and B were substantial (2.37 and 4.1 Å, respectively), and further, these standard structures showed numerous restraint violations, as shown by the residual restraint force, indicating that the restraints drive the calculation rather than the potentials of the Amber forcefield. The overall structure is clearly within the B-family of conformations.

rR10·rY10. 221 restraints (11.2 per residue) were used in the rMD calculations. The initial A model gave a large number of restraint violations. In addition, A-RNA energy-minimized in the absence of restraints also gave poor agreement with the experimental data. The pure RNA duplex starting from standard A refined to a structure that is still A-like. However, there were significant differences from the canonical structure as shown by the derived parameters in Table 2. Based on the pairwise all-atom rmsd between structures refined using slightly different protocols and variations in the forcefield, the radius of convergence was about 0.9 Å. Nevertheless, the refined structures are on average closer to the energy-minimized A structure (rmsd = 1.81 Å) than the DNA duplex is to the energy-minimized B structure.

rR10·dY10 and dR10·rY10. We have applied the standard refinement protocol used for the DNA and RNA duplexes

Table 3: rmsd Values for Coupling Constants of the Deoxyribose Strand in the DNA·RNA Hybrids. rmsd Values were Calculated According to Equation 1

structure	rmsd(*J*)/Hz		
	$\Sigma_{1'}$	$\Sigma_{2'}$	$\Sigma_{2''}$
dR10·rY10[a]	0.39	1.16	2.91
dR10·rY10[b]	1.36	1.55	4.5
rR10·dY10[a]	0.43	0.71	2.68
rR10·dY10[b]	2.28	2.71	3.60

[a] Ensemble average. [b] Single structure.

on the two hybrids. In the absence of restraints on δ for the deoxyriboses, the dR10·rY10 hybrid refined to a reasonable geometry with a negative energy and small residual violations (ca. 0.9 kcal/mol for 11.6 restraints per residue), and showed a pairwise rmsd of 1.03 ± 0.19 Å. The rmsd to the unrestrained, energy-minimized hybrid was 1.55 ± 0.24 Å. Overall, the structures resemble the A form more than the B form. However, in these structures, the value of δ in the deoxyriboses was ca. 120°, which gives rise to large violations of the coupling constants. In contrast, when the deoxyriboses were restrained to give a single sugar conformation that gives the best compromise agreement with the coupling constants, the final structures gave a poorer net energy, and significant restraint violations (ca. 8 kcal/mol). The refined δ values were ca. 95°, which is unable to account fully for the scalar coupling data. For example, this value of δ implies a value of $\Sigma_{1'} > 15$ Hz, which exceeds the observed values of ca. 13−14 Hz (*9*), and also indicates that the distance between H1′ and H4′ should be at a minimum (ca. 2.6 Å), which is substantially less than observed. We conclude that although it is possible to build DNA·RNA hybrids with a variety of conformations in the DNA strand, within the context of a globally A-like structure, in these instances, unique intermediate conformations do not fully account for the available data (and see Table 3).

To account for the relatively high degree of conformational averaging of the deoxyriboses in the two DNA·RNA hybrids, we have considered a relatively simple model of conformational averaging that intrinsically accounts for the coupling constants and NOE data (see Methods) using a minimum number of assumptions about the number of conformational states that is considered. The nucleotide conformations for the deoxyribose strands were derived using the program NUCFIT (*25*) assuming a two-state model in which the glycosidic torsion angles and the S and N sugar pucker conformations are different. Two sets of tight constraints on the nucleotide conformations were used, one for the S state and the other for the N state, and internucleotide and cross-strand restraints set relatively loosely according to the classification strong, medium, and weak. Restraints for the RNA strands were set tightly as it is assumed that the ribonucleotides do not undergo conformational averaging (see above). Duplexes were then built in the A and B conformations, and refined using the cycle of energy minimization, rMD at high temperature, cooling, energy minimization, low temperature rMD (300 K) and final energy minimization. These final models were then examined for violations of the input constraints, and the NOESY intensities calculated. We have made use of the iterative refinement method based on cycles of structure calculations and NOE calculations (*26*), also called the "back-calculation" method (*27*). This process allowed conformational features not in agreement with the primary data to be identified, and adjustments to the input files to be made. It also allowed tighter restraints to be applied where several structures lead to the same local structures that also agreed with the NOEs.

The statistics of the refinements under these different protocols are given in Table 1. These data show that the restraints can be satisfied with reasonable energies and stereochemistries. More importantly, the ensemble averages account for the scalar coupling data, whereas the single-conformation rMD model does not (Table 3). The ensemble averaged NOEs also agree better with the NOE data than any single member of the ensemble, or the rMD single conformation model. In the refined structures, there were no distance violations >0.1 Å and no torsion violations >1°.

Figure 1 shows the refined structures. There is a clear trend from the B structure of the DNA duplex to the A structure of the RNA duplex, via the intermediate forms of the two hybrids. This is particularly manifest in the widths of the major and minor grooves. The minor groove becomes progressively wider and shallower, and the major groove deeper and narrower in the sequence dR10·dY10, dR10·rY10, rR10·dY10, rR10·rY10. This is accompanied by an increase in the displacement of the helix axis, leading to the formation of a hole in the center of the molecule.

Figure 2 shows an overlay of the 10 lowest energy/best fit structures of each conformer in the ensembles of the two hybrids and the 10 best structures assuming unique conformations for the DNA and RNA duplexes. The rR10·dY10 duplex shows a wider spread of structures than the dR10·rY10 duplex even though it was defined by more constraints. This presumably reflects the differences at the nucleotide level, in which the dR10·rY10 duplex showed less extensive averaging in the sugars (*9*). The two pure duplexes showed good convergence, and the family of conformations show that positions of the bases are well determined whereas the phosphate backbones are less well determined by the data. Further, the pyrimidine strand in the DNA duplex is less well defined than the purine strand, reflecting the relative density of constraints. The overall better definition of the DNA and the RNA duplexes compared with the two hybrids is because only a single conformation is assumed, and therefore tighter constraints for all of the data could be used than in the conformational ensemble required for the two hybrids.

The distribution of torsion angles is shown as a bars plot in Figure 3. In these plots, the torsion angle from each structure in the ensemble is shown as a bar. Thus, the grouping of the bars reflects the precision of the determination of each torsion angle. Given the nature of the constraints, the nucleotide parameters are quite well determined. Particularly noticeable in the DNA duplex is that most of the parameters are quite well determined. This is because we were able to constrain most of the γ torsion angles. Where they were not constrained, owing to peak overlap or uncertainty, two families of local conformational were obtained, with γ either near 60° or near 180°. The residues where this occurred (e.g. G11,C12,C15,T16,C17,T18,C20 in dR10·dY10) showed that the value of α also generally fell into two families, which corresponds to the well known crankshaft compensation, and other backbone

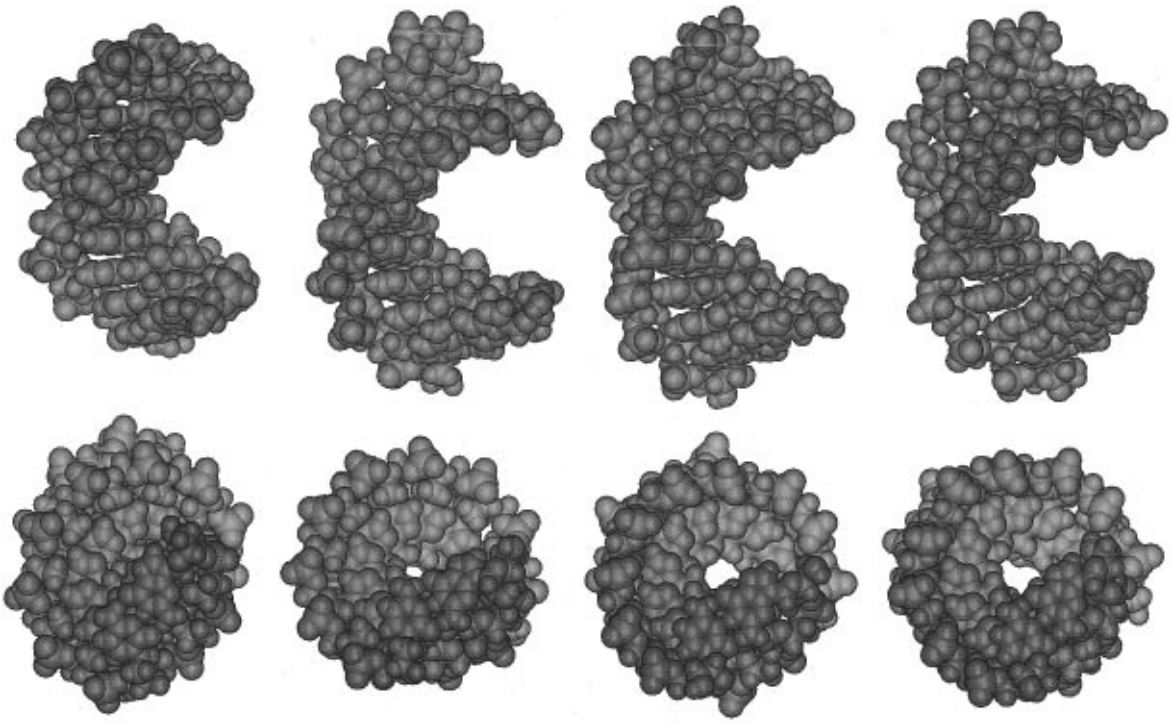

FIGURE 1: Structures of the decamers. CPK space-filling representation of the four duplexes. Depicted from left to right dR10·dY10, dR10·rY10, rR10·dY10, and rR10·rY10. Upper view is the side elevation highlighting differences between the major and minor groove widths (G1:C11 top). Lower is a view down the helix axis showing increasing base-pair displacement from the global axis (G1:C11 proximal).

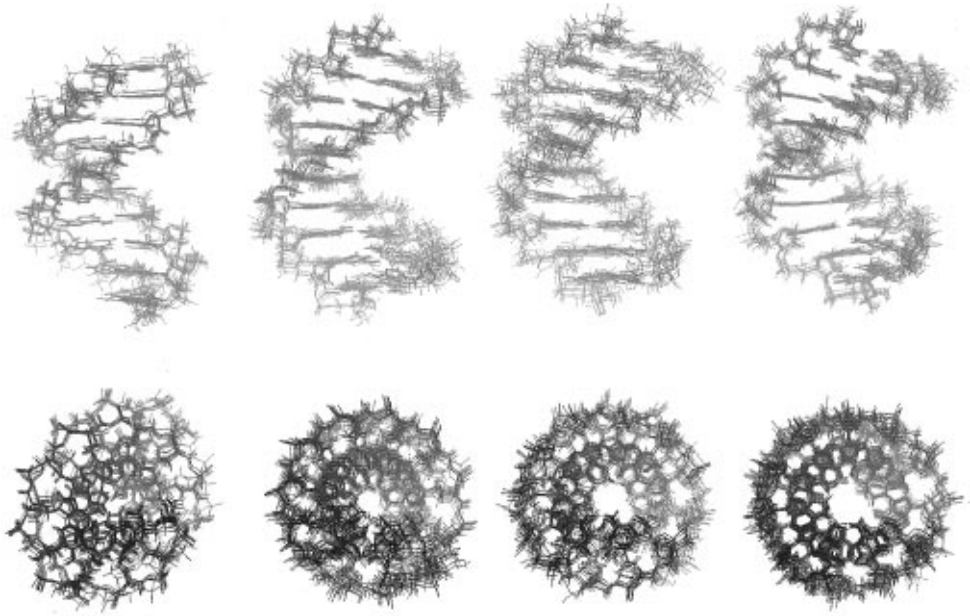

FIGURE 2: Comparison of the ensembles of the two DNA·RNA hybrids with the DNA and RNA duplexes. The 10 best structures were overlaid in the same orientation as in Figure 1. Left to right dR10·dY10, dR10·rY10, rR10·dY10, and rR10·rY10. For the two hybrids, each structure is one member of the ensemble, and therefore each molecule has a different conformation. In the families of pure DNA and RNA duplexes each structure was generated using the same restraints.

angles also adjust somewhat. This shows two things. First, the conformational sampling was sufficiently extensive to allow excursions into radically different rotameric domains. Second, the backbone torsion angles are quite well specified when χ, δ, and γ are all constrained, but poorly determined otherwise. In the two hybrids, relatively more of the multiple solutions for the backbone torsions is observed, especially in the DNA strands. This again correlates with the absence of γ constraints in these residues. In addition, multiple families are found because δ and χ adopt different values in different structures, which represents the ensemble, and accounts in part for the greater variation of torsion angles within the structure compared with the DNA duplex. In rare cases, some backbone torsion angles occupy positions that are stereochemically unfavorable, such as $\epsilon \approx 60°$ (g^+). Such conformations are unexpected, though they were observed occasionally in an rMD-tar calculation on a DNA·RNA hybrid (*15*). However, as there are no direct experimental constraints on these torsions, and the energies, bond angles, and bond lengths lie within the normal ranges for phos-

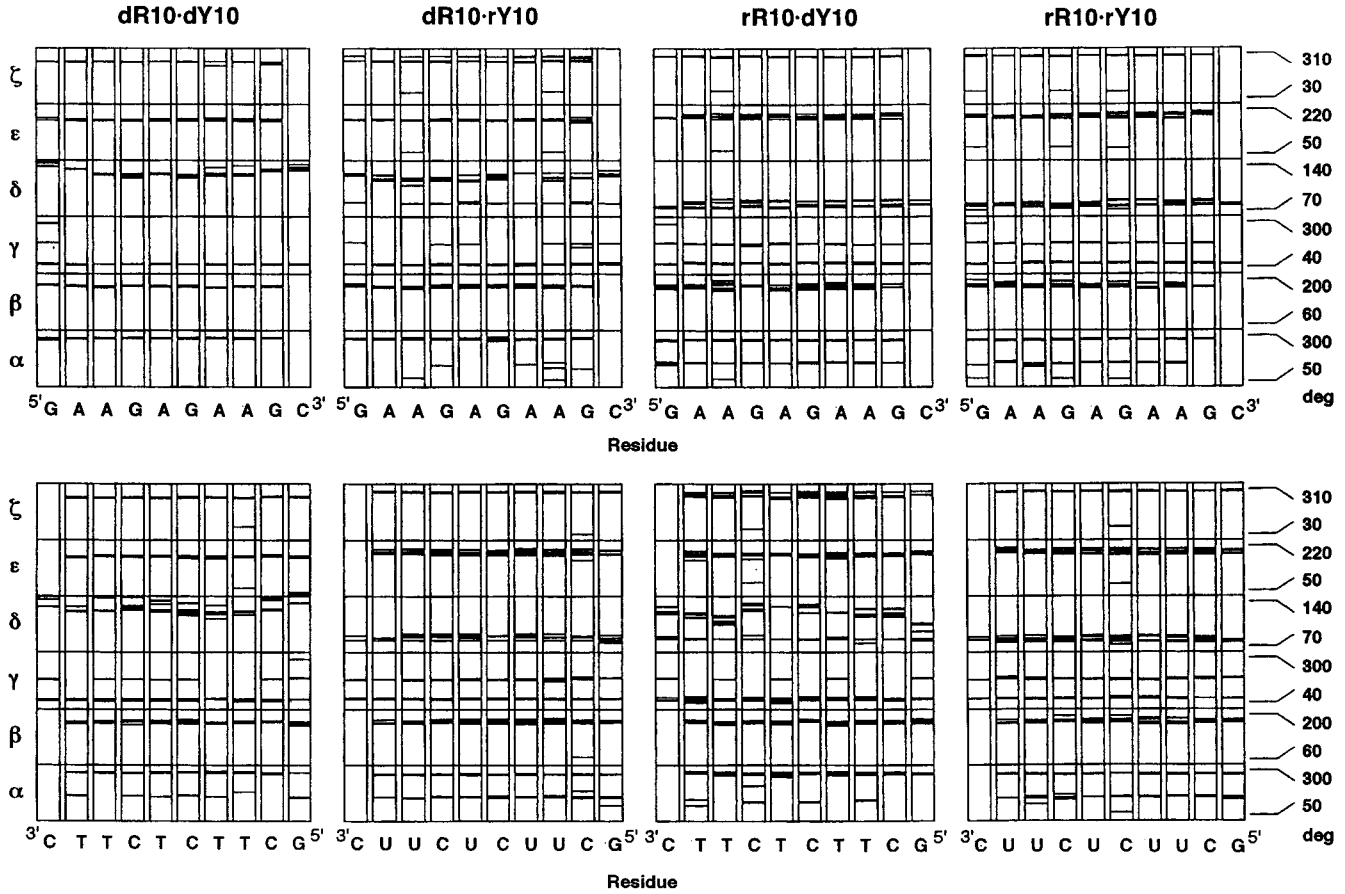

FIGURE 3: Bars plot of torsion angles. The torsion angles were calculated for individual structures and are each represented by a single bar. For both dR10·dY10 and rR10·rY10, the 10 bars represent equivalent conformations from the refinement, whereas for the two hybrids, each bar represents a different member of the conformational ensemble.

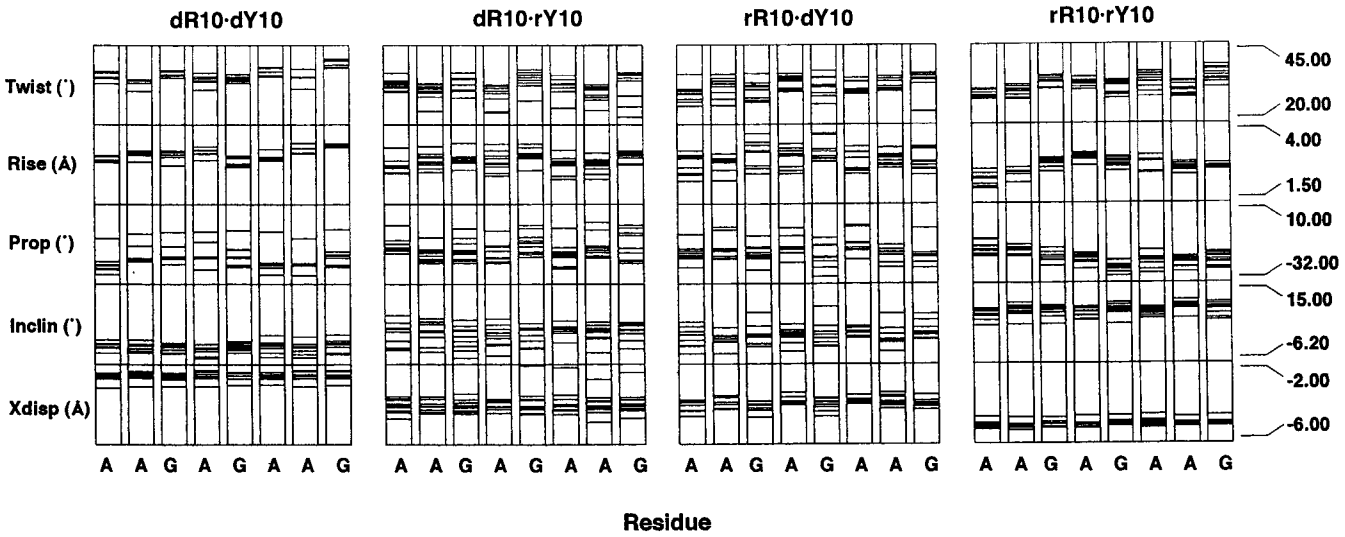

FIGURE 4: Bars figure of helicoid parameters. Helicoid parameters were obtained using CURVES 5.1. Each bar repesents a value from one conformation, as described in Figure 3.

phodiester, we have no experimental justification for disallowing these structures. Their rare occurrence suggests that if they were present in solution at these frequencies, they would go undetected.

Figure 4 shows the distribution of helicoid parameters for the ensembles across the series of structures. The helicoid parameters for the DNA and RNA duplexes are reasonably tight at each step, and the variation of values reflects the precision with which these parameters are determined. The variation of the helicoid parameters at each step for the two hybrids, however, shows a much wider dispersion on average, which demonstrates the influence of the nucleotide conformations. The presence of N and S sugars requires adjustments of backbone torsion angles which correlate also with helicoid parameters, especially the propeller twist, rise, and helical twist, and less so for the axis displacement and inclination (*21*). Although there is considerable scatter, there is a significant positive correlation between the helical rise

and the helical twist (see Supporting Information). The base-pair tilts tended to be small and generally showed no or a weak correlation with the rise. However, there is a correlation with the rise and the inclination when averaged over entire molecules (see below). In addition to the nucleotide conformations, the helicoid parameters, and torsion angles in the ensembles encompass a substantially wider range of the conformation space than the DNA and RNA duplexes.

Some of the molecule-averaged defining helicoid parameters for the structures are given in Table 2. The DNA duplex is characterized by a high twist (10.3 bp per turn) and rise, low base inclination, and a fairly small axis displacement typical of the B family of conformations. The RNA structure is characterized by a low twist (11.8 bp per turn), low rise and a positive base-pair inclination, and a large displacement of the helix axis into the minor groove. This is typical of the A family of conformations. Whereas the absolute accuracy of the parameters is probably not high, especially for the axis displacement, the observed trends are consistent with expectation for these types of duplex. The two hybrids are characterized by a low twist and a rise that is intermediate between that of the RNA and DNA duplexes. Furthermore, the base-pair inclinations are small and positive, which correlates with the higher rise than in the RNA duplex, and are intermediate between the A and B forms. This is also shown by the axis displacement and accounts for the size of the hole seen in the end on view of the duplexes (Figure 1). Thus, the global structures also show similar average trends as previously described for the nucleotides.

DISCUSSION

All of the data show that the DNA·RNA hybrids have local and global conformations that are somewhat different from either the DNA or RNA duplexes. The statistics of the refinements show that the DNA and RNA duplexes can be adequately represented as unique conformational states. This is not to say that conformational fluctuations do not occur in these molecules, only that such fluctuations do not affect the NMR observables sufficiently at the level of resolution attainable with these sequences. In contrast, the unique conformation model does not satisfy the experimental data for the hybrids as well as the conformational blend. Because of the paucity of data, the ensemble of structures generated cannot be unique. Thus, for a decamer duplex, there are approximately 130 degrees of freedom to be determined by relatively imprecise data at a ratio of about two per degree of freedom. Positing a mixture of only two conformational states immediately doubles the number of degrees of freedom and halves the ratio of data to parameters to be determined. However, if the nucleotides in the DNA strands exist as a mixture of conformations, then it is not possible to describe the hybrids as a mixture of only two conformations. In principle, up to 2^{10} conformations are present in solution, though this would imply that each nucleotide is independent of its nearest neighbors, i.e. there is no cooperativity. For $f_s = 0.7$ (*9*), the most probable distribution of conformations is where seven nucleotides are S and three are N (there are 120 of these conformations). An alternative viewpoint, whereby there is complete cooperativity in the DNA strands, would be that approximately 70% of the molecules have DNA strands that are all C2′-endo, and 30% are all C3′-endo. Most likely the situation lies between these extremes. Nevertheless, studies on other nucleic acid systems have shown that the DNA backbone is extremely flexible, and the large number of degrees of freedom can accommodate substantial local structural perturbations at the level of only 2−3 bp. For example, monointercalation causes a local unwinding and change in backbone torsion angles (*28*), but only at the site of the intercalation and next neighbors. Many mismatched structures show similar behavior in terms of conformation at sites distal to a G·A mismatch (*29*), in the strongly perturbed tandem G·A mismatched structure (*30*,*31*), and supported by the ready dynamic exchange between syn and anti conformations in G·A (*29*) and G·G mismatches (*32*). We have recently shown that incorporating N-sugar conformations into an otherwise B DNA structure causes only local rearrangements in the adjoining backbone (*20*). Sugar repuckering is a feature of detailed molecular dynamics calculations of DNA·RNA hybrids (*24*), supporting the present model for treating conformational averaging in these systems.

This treatment of conformational averaging is incomplete, as the ensemble generated represents only a sample of the whole ensemble of known conformers. It is also impossible with the current level of data to describe individual conformers in greater detail, because the ratio of independent experimental data to variables is too low. In this method of treating the ensemble, we are implicitly assuming that the nucleotides can be described in terms of a two-state equilibrium and that neighbors are only weakly coupled. This cannot be completely true in general, and is a fundamental limitation of this method, though this approximation can be justified according to our current experimental and theoretical knowledge of nucleic acids (see above). However, it does not rely strongly on the parameterization of the forcefield. Other methods of ensemble calculations have been described, including population analysis of MD-tar calculations (*15*) and *a posteriori* probability calculations on ensembles of conformers (*19*). While these approaches are fundamentally different in philosophy, in practice they appear to give a rather similar picture of the conformational ensemble. In our approach, we assume a subset of conformations in advance (with experimentally determined probabilities) and therefore are blind to other kinds of averaging that presumably occur, whereas in the alternative approach, the data are used to filter out those conformations that cannot be compatible with the data. Thus, in one method, too few conformations are probably included, and in the other, too many may be included. The use of both time and ensemble averaged direct NOE restraints has been shown to improve the fit to the experimental data (*33*). However, the determination of probabilities of the conformers was shown to be inaccurate for imprecise data even using cross-validation methods (*34*). Given the number of data available from NMR experiments, this is not surprising, as the problem is intrinsically underdetermined. Hence, modeling and the properties of the force-fields used in generating structures is unavoidable.

Nevertheless, the general features of the solution behavior of the two hybrids compared with the parent duplexes is clear. There is a progressive transformation of the B-like DNA duplex toward the A-RNA duplex as the strands are exchanged dY10→rY10, dR10→rR10. This is manifest

particularly as the average displacement of the base-pairs from the global helix axis, increasing base-pair incliniations and decreasing rises. However, the transformation of dR10•dY10 into dR10•rY10 leads to a very unstable molecule, that has the most intermediate structure. The variation in the helical rise from DNA to RNA can be correlated with the observed electrophoretic mobility (*9*). Although the hydrodynamic friction should be proportional to the length of the duplexes, i.e. DNA would be expected to migrate more slowly than RNA, the opposite behavior is observed, and we have shown that the rotational correlation time (*9*) and translational friction coefficients (*35*) for the DNA and RNA duplexes are very similar. However, in the electrophoresis experiment, the effective charge is extremely important, and this depends not only on the degree of ion condensation, but also the size of the Debye ionic atmosphere dragged along with the solute particle. These depend on the axial charge density, which will be greatest for the RNA duplex and least for the DNA duplex, with the hybrids intermediate. Hence, at least for short oligonucleotides, the electrophoretic mobility seems to depend on the mean axial rise, so that this method indirectly measures a global property of the duplexes.

As far as antisense technology is concerned, it may be that DNA•RNA hybrids of different strand compositions may be more or less resistant to attack by RNaseH. It would be interesting to compare the relative rates of cleavage of mixed-sequence DNA•RNA hybrids versus the all-pyrimidine and all-purine RNA strands.

ACKNOWLEDGMENT

NMR experiments were carried out in the MRC Biomedical NMR Centre, Mill Hill, London. GLC was supported by a scholarship from the Caledonian Trust for the Universities of Scotland.

SUPPORTING INFORMATION AVAILABLE

Table showing the distribution of sugar conformations in an ensemble and four graphs showing correlation of helical twist with helical rise and rise with base-pair tilts (6 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Hansen, U. M., and McClure, W. R. (1980) *J. Biol. Chem. 255,* 9564−9570.
2. Adams, R. L. P., Knowler, J. T., and Leader, D. P. *The Biochemistry of the Nucleic Acids*, 10th ed., Chapman and Hall, London, Ch. 6.
3. Varmus, H. (1988) *Science 240,* 1427−1435.
4. Stein, H., and Hausen, P. (1969) *Science 166,* 393−395.
5. Hung, S.-H., Yu, Q., Gray, D. M., and Ratliff, R. L. (1994) *Nucl. Acids. Res. 22,* 4326−4334.
6. Ratmeyer, L., Vinayak, R., Zhong, Y. Y., Zon, G., and Wilson, W. D. (1994) *Biochemistry 33,* 5298−5304.
7. Lesnik, E. A., and Freier, S. M. (1995) *Biochemistry 34,* 10807−10815.
8. Wang, S., and Kool, E. T. (1995) *Biochemistry 32,* 4125−4132.
9. Gyi, J. I., Conn, G. L., Brown, T., and Lane, A. N. (1996) *Biochemistry 35,* 12538−12548.
10. Fedoroff, O. Y., Salazar, M., and Reid, B. R. (1993) *J. Mol. Biol. 233,* 509−523.
11. Lane, A. N., Ebel, S., and Brown, T. (1993) *Eur. J. Biochem. 215,* 297−306.
12. Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., and Reid, B. R. (1993) *Biochemistry 32,* 4207−4215.
13. Gao, X., and Jeffs, P. W. (1994) *J. Biomol. NMR 4,* 367−384.
14. Gonzalez, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., and James, T. L. (1994) *Biochemistry 33,* 11062−11072.
15. Gonzalez, C., Stec, W., Reynolds, M., and James, T. L. (1995) *Biochemistry 34,* 4969−4982.
16. Cross, C. W., Rice, J. S., and Gao, X. (1997) *Biochemistry 36,* 4096−4107.
17. Lima, W. F., and Crooke, S. T. (1997) *Biochemistry 36,* 390−398.
18. Bonvin, A. M. J. J., and Brünger, A. T. (1995) *J. Mol. Biol. 250,* 80−93.
19. Ulyanov, N. B., Schmitz, U., Kumar, A., and James, T. L. (1995) *Biophys. J. 68,* 13−24.
20. Lane, A. N. (1996) *Magn. Reson. Chem. 34,* S3−S10.
21. Lane, A. N. (1997) *Molecular modeling and structural determination of nucleic acids,* ACS Symposium series, American Chemical Society, Washington, DC (in press).
22. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson. 48,* 286−292.
23. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. Struct. Dyn. 2,* 661−665.
24. Cheatham III, T. E., and Kollman, P. A. (1997) *J. Am. Chem. Soc.* 4805−4825.
25. Lane, A. N. (1990) *Biochim. Biophys. Acta 1049,* 189−204.
26. Gariépy, J., Lane, A. N., Schoolnik, G., and Jardetzky, O. (1986) *Biochemistry 25,* 7854−66.
27. Banks, K. M., Hare, D. R., and Reid, B. R. (1989) *Biochemistry 28,* 6996−7010.
28. Searle, M. S. (1993) *Progr. NMR Spectrosc. 25,* 403−480.
29. Lane, A. N., Jenkins, T. C., Brown, D., and Brown, T. (1991) *Biochem. J. 279,* 269−281.
30. Lane, A. N., Martin, S. R., Ebel, S., and Brown, T. (1992) *Biochemistry 31,* 12087−12095.
31. Chou, S.-H., Cheng, J.-W., and Reid, B. R. (1992) *J. Mol. Biol. 228,* 138−155.
32. Lane, A. N., and Peck, B. (1995) *Eur. J. Biochem. 230,* 1073−1087.
33. Bonvin, A. M. J. J., Boelens, R., and Kaptein, R. (1994) *J. Biomol. NMR 4,* 143−149.
34. Bonvin, A. M. J. J., and Brünger, A. T. (1996) *J. Biomol. NMR 7,* 72−76.
35. Bonifacio, G., Brown, T., Conn, G. L., and Lane, A. N. (1997) *Biophys J. 73,* 1532−1538.
36. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6,* 63−91.

BI9719713