Journal of Biomolecular NMR, **20:** 311–323, 2001.
KLUWER/ESCOM

Application of dipolar coupling data to the refinement of the solution structure of the Sarcin-Ricin loop RNA

J.J. Warren[a] & P.B. Moore[a,b,*]
Departments of [a]Molecular Biophysics and Biochemistry and [b]Chemistry, Yale University, New Haven, CT 06520-8107, U.S.A.

Received 6 March 2001; Accepted 18 May 2001

Key words: residual dipolar couplings, RNA, sarcin-ricin loop, solution structure determination

Abstract

Residual dipolar couplings can provide the long-range information that most NMR solution structures lack. The use of such data in protein structure determinations is now fairly routine, but even though these data should be much more useful for nucleic acids, their application to nucleic acid structure determination is still in its infancy. Here we present a method for producing accurate, dipolar-refined structures of nucleic acids that is more efficient than those used previously, and apply it to E73, a 29 nucleotide RNA that includes the sarcin-ricin loop from rat 28S rRNA. The results enable us to address the differences between the crystal structure of E73 and the solution structure proposed for it previously.

Introduction

NMR solution structures are determined primarily using restraints derived from nuclear Overhauser effects, which report proton-proton distances less than 5 Å, and scalar coupling constants, which provide information about local dihedral angles. The absence of long-range information from NMR data sets has little effect on the quality of NMR-determined structures of globular molecules, like the average protein domain, but it is a serious problem for elongated molecules, like helical nucleic acids. The reason is that local errors tend to add over the length of such molecules with the result that their overall shapes are poorly determined. This problem is aggravated in nucleic acids, relative to proteins, by two additional factors. First, nucleic acids provide less data for the NMR spectroscopist because they have fewer protons, weight for weight. Second, the chemical shift dispersion is worse for nucleic acids, making the relatively small number of proton resonances in their spectra harder to resolve.

*To whom correspondence should be addressed. E-mail: moore@neutron.chem.yale.edu

Residual dipolar coupling techniques recently pioneered by protein spectroscopists (Tjandra and Bax, 1997a) present nucleic acid spectroscopists with a significant opportunity. Dipolar couplings, which average to zero in molecules that are isotropically oriented and rapidly tumbling, can be observed if the distribution of molecular orientations sampled is made even slightly anisotropic. If the level of anisotropy is adjusted appropriately, coupling data can be obtained that are large enough to be accurately measured, and yet small enough so that spectra retain the simplicity typical of normal, isotropic solution spectroscopy. The balance required can be obtained using samples consisting of the molecule of interest dissolved in dilute solutions of liquid crystals composed of phospholipid mixtures (Bax and Tjandra, 1997; Tjandra and Bax, 1997a) or filamentous phage (Clore et al., 1998c; Hansen et al., 1998), both of which orient in magnetic fields.

The size of the dipolar coupling observed between two nuclei in an oriented sample is given by (Bothner-By, 1995; Clore et al., 1998a; Gayathri et al., 1982):

$$D_{PQ}(\theta, \phi) = D_a^{PQ}\left[\left(3\ \cos^2\ \theta - 1\right) + 1.5R\ \sin^2\ \theta\ \cos\ 2\phi\right], \quad (1)$$

where R is the rhombicity (which is related to the shape of the molecule) and D_a^{PQ} is a scale factor that subsumes the gyromagnetic ratios of the nuclei whose coupling is being observed, their order parameter, their separation, etc. The reason dipolar coupling data are useful for solution structure refinement is that the values of the couplings measured are determined by the polar angles (θ and ϕ in the above equation) describing the orientations of the vectors between coupled nuclei in a coordinate axis system that is the same throughout the molecule. Thus these couplings are a source of long-range structural information. The orientation of this universal axis system in the molecular frame of reference is determined by the orientation that the molecule tends to assume in the orienting medium. For example, for steric reasons, helical nucleic acids orient slightly more frequently with their long axes parallel to the long axes of a dilute, oriented array of filamentous phage than perpendicular to them, and thus the Z-axis of the coordinate system describing nucleic acid dipolar couplings will correspond to the long axis of the nucleic acid helix.

Methods have been developed for using dipolar coupling data in protein structure determinations, but their application to RNA structures is not straightforward. The problem is that before dipolar couplings can be used for structure determination (Brunger et al., 1998; Clore et al., 1998a, b; Tjandra et al., 1997), the values of D_a^{PQ} and R in Equation 1 must be determined, and this is more difficult for nucleic acids than for proteins for two reasons. First, the number of couplings that can be measured in nucleic acid spectra is smaller because the number of observable proton resonances is smaller. Second, because the orientation of internuclear vectors tends to be random in proteins, D_a^{PQ} and R can be determined fairly accurately from coupling data histograms (Clore et al., 1998b). This approach fails for helical or quasi-helical nucleic acids because the internuclear vector orientation distribution is far from random. So far, those who have used residual dipolar coupling data in nucleic acid structure determinations have resorted to a computationally expensive strategy that involves repeatedly refining structures with dipolar data using many different values of D_a^{PQ} and R, and choosing the combination that results in the lowest energy structures (Bayer et al., 1999; Clore et al., 1998a; Tjandra et al., 2000). Here we present an alternative strategy which uses a steric model for RNA-phage interaction (Zweckstetter and Bax, 2000) to determine bounds for reasonable values of D_a^{PQ} and R. Using a simple mapping strategy (Warren and Moore, 2001) nucleic acid structures can be refined using CNS taking the ranges so determined into account (Brunger et al., 1998; Clore et al., 1998a, b; Tjandra et al., 1997). Used iteratively, rounds of D_a^{PQ} and R determination and refinement generate families of structures that rapidly converge.

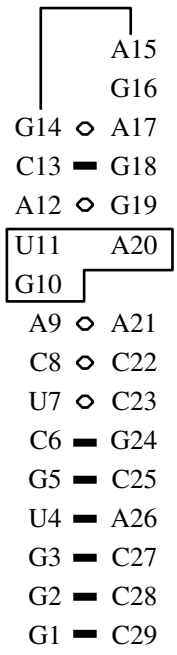

Figure 1. Secondary structure diagram of E73. Watson-Crick base-pairs are indicated by heavy lines and non-Watson-Crick pairs by open circles. G10, U11, and A20 (boxed) form a central base triple.

The molecule we have used as the 'test bed' for developing this approach is a 29-base RNA oligonucleotide called E73, Figure 1, which includes the sarcin-ricin loop (SRL) sequence from rat 28S rRNA. The SRL is a hairpin loop from the large ribosomal subunit rRNA that derives its name from the fact that is the target of the ribotoxins sarcin and ricin. The cleavage of specific bonds in the SRL by either toxin kills cells by inactivating their ribosomes (Endo and Wool, 1982; Hausner et al., 1987).

Because of its importance in translation, the SRL has been intensively studied. The structure of E73 has been solved both by NMR (Rife et al., 1999; Szewczak et al., 1993; Szewczak and Moore, 1995) and by X-ray crystallography (Correll et al., 1998), and the crystal structure of a similar RNA containing the *E. coli* SRL has also been determined (Correll et al., 1999). Although the crystal and NMR structures of E73 are quite similar, they differ in two respects: the

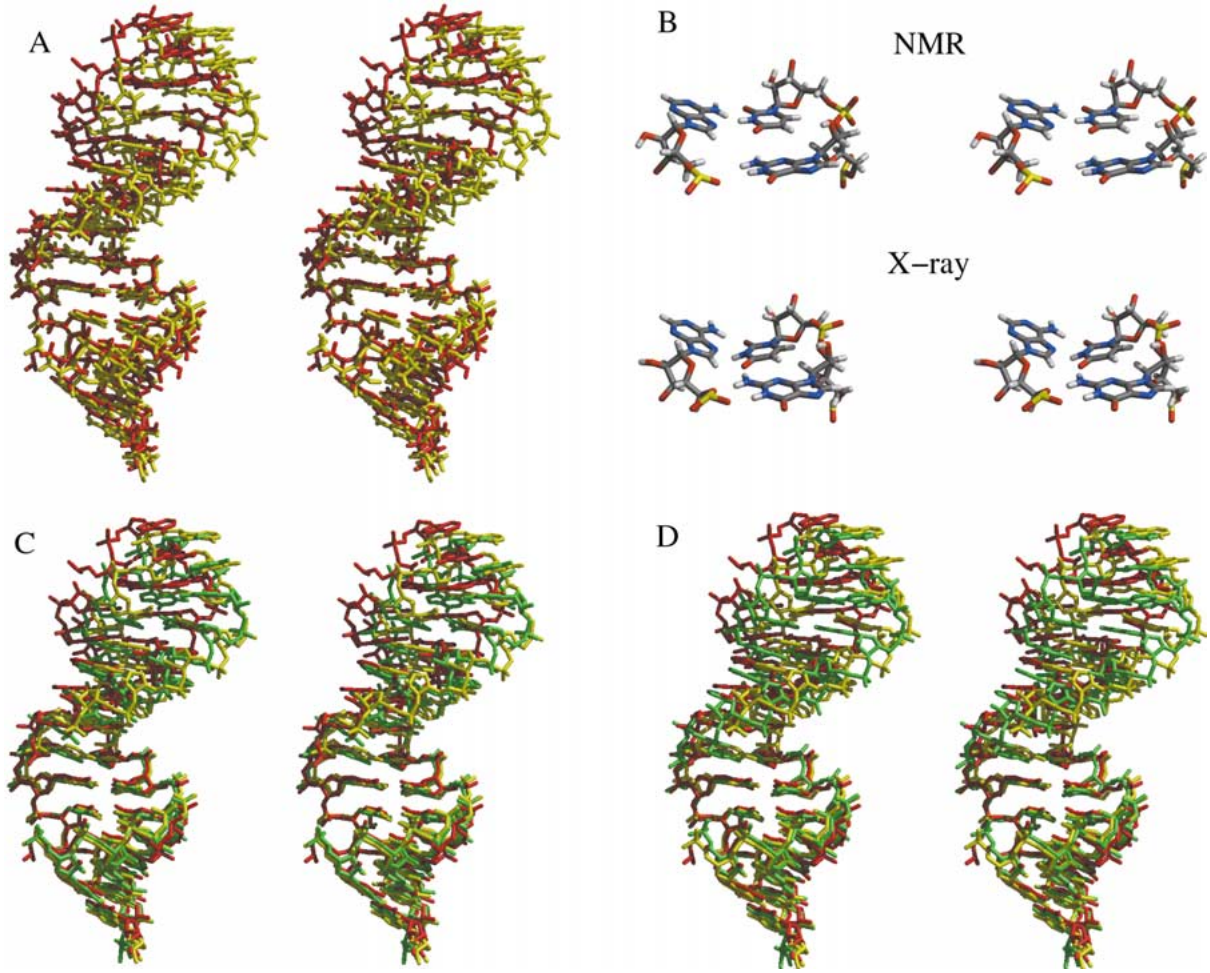

Figure 2. Differences between the solution and crystal structures of E73. A: Stereo pair showing superposition of the first 6 base pairs of the crystal (red) and NMR (yellow) structures. B: Stereo views of the G10-U11-A20 base triple from the Rife et al. (1999) NMR structure (top) and the Correll et al. (1998) crystal structure (bottom). Panels C and D: Comparison of average dipolar refined structures (green) with crystal (red) and NMR structure of Rife et al. (yellow). Panel C: stereo pair with NMR, crystal, and grid search dipolar structure. Panel D: stereo pair with NMR, crystal, and structure-based dipolar structure. In both panels the three structures are superimposed on the first six base pairs to emphasize differences in bend.

overall bend of the molecule, and the position of G10 in the G10, U11, A20 base triple (Figure 2). Two explanations could be offered for the difference in molecular shape. On the one hand, the NMR structure was deduced using only NOEs and coupling constants, and hence if the overall shape of the molecule is not in error, it is at least poorly determined. Figure 3A shows the result of superimposing a family of NMR-determined E73 structures on the first 6 base pairs of their terminal stem. The fanning out of the opposite end of the molecule demonstrates the problem. On the other hand, crystallization can affect nucleic acid structures (see, e.g., Holbrook et al., 1991; Baeyens et al., 1996), and it could be that the structure of E73 in crystals is different from the structure of E73 in solution. This hypothesis is supported by J coupling data, which shows that some of the sugar puckers observed in solution differ from those predicted by the crystal. Thus, E73 seems to be a reasonable molecule to use to test the application of dipolar data to RNA structure determination. The local structure of the RNA, which is well understood, can provide a test of the accuracy of resulting structures. At the same time, by generating a more precisely determined solution structure for E73 we may be able to address the origin of differences between its crystal and NMR structures.

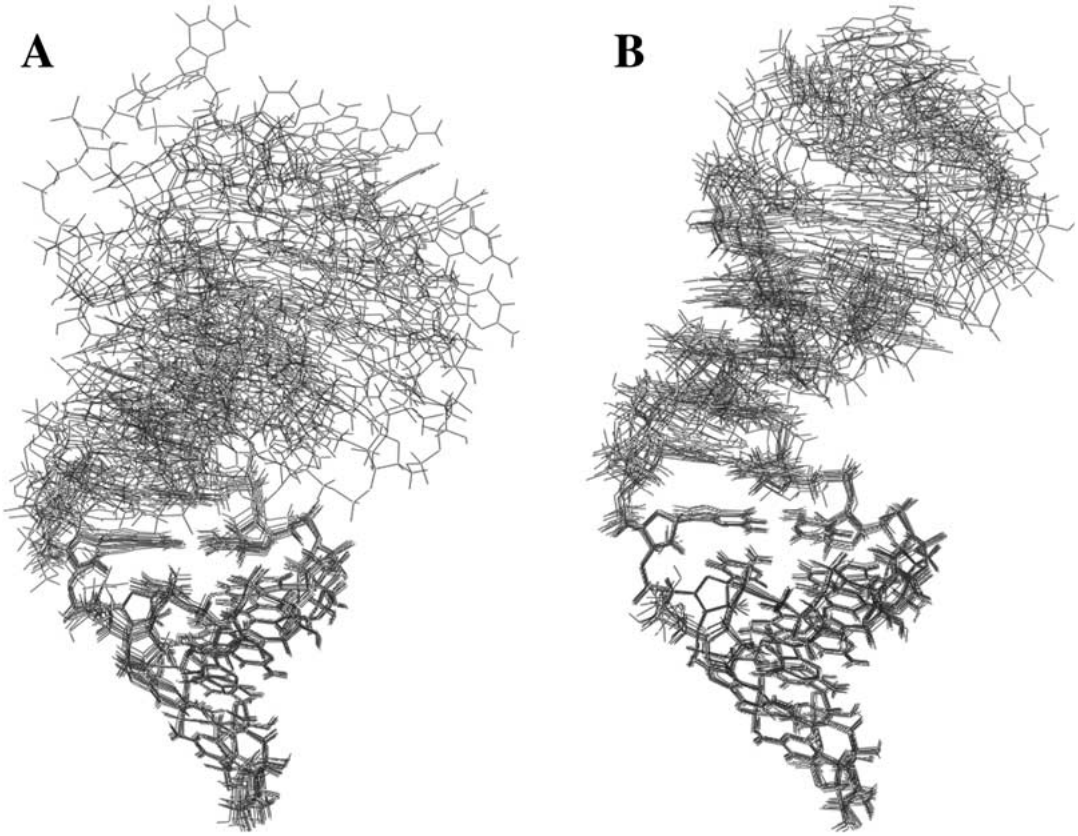

Figure 3. Incorporation of dipolar coupling data significantly improves the precision of structures of E73. A: The bend of E73 is underdetermined in the solution structure. Superposition of a family of 10 NMR structures of E73, refined using only NOE and dihedral restraints, on their first 6 base pairs. B: Increased precision of the structure-based dipolar refined NMR family (see text for details). As in panel A, the family of structures is superimposed on their first six base pairs.

Materials and methods

Samples

^{13}C and ^{15}N uniformly labeled nucleotides were prepared by growth of a nuclease-deficient strain of *E. coli* on minimal media with ^{13}C glucose and ^{15}N ammonium chloride as the sole carbon and nitrogen sources, respectively, as previously described (Batey et al., 1995). Both labeled and unlabelled samples of the E73 RNA (GGGUGCUCAG UACGAGAGGA ACCGCACCC) were synthesized by *in vitro* transcription of a semi-single stranded template with recombinant T7 RNA polymerase. Crude transcription mixtures were gel purified, and the product band was electro-eluted and dialyzed into NMR buffer (50 mM NaCl, 5 mM cacodylate, 0.1 mM EDTA, pH 6.8) before being concentrated to 1–1.7 mM. For nonexchangable NMR, samples were exchanged into D_2O by repeated lyophilization and resuspension in D_2O. Samples were transferred to reduced volume Shegemi NMR tubes for data collection.

Pf1 phage was prepared as described previously (Hansen et al., 2000). Oriented samples were prepared by first exchanging Pf1 into D_2O solution containing the desired NMR sample buffer by repeated pelleting in a tabletop ultracentrifuge and resuspension in the desired buffer. A small amount of the phage thus prepared was added to the desired NMR sample (which was already in D_2O and the appropriate buffer). Final phage concentrations were ~25 mg/ml.

Spectroscopy

All spectra in this study were collected either on a Varian Inova (500 MHz for ^1H) or a Varian UnityPlus (600 MHz for ^1H). Both magnets were *Z* gradient-capable. Data were processed on Silicon Graphics workstations using Felix 97 (MSI/Biosym, Inc.).

Although nearly complete proton assignments were obtained for E73 at the time it was first characterized (Szewczak et al., 1993; Szewczak and Moore, 1995), few carbon assignments were determined. ^{13}C assignments had to be obtained before one-bond C-H dipolar coupling constants could be determined. This was accomplished using information from two complementary experiments: an HcCH TOCSY and an hCCH TOCSY (Bax et al., 1990). The carbon and proton assignments obtained are available from the authors as supplementary material for this article.

One bond J_{CH} couplings were determined from J_{CH} modulated 2D HSQC spectra (Tjandra and Bax, 1997b). Generally sets of 10 such spectra were collected in an interleaved fashion, with J_{CH} evolution times ranging from 0.3 ms to 5.7 ms over the ten subspectra. The total data collection time was approximately 12 h. Couplings from aromatic and anomeric regions were collected separately both because of the large differences in chemical shift between these two regions and because the constant time periods required for these two classes of C-H bonds are different (15.3 and 25 ms for aromatic and anomeric, respectively). Spectra were phased and peaks were picked using Felix 97 (BIOSYM/Molecular Simulations). The time evolution of the volume of each peak was fit to the function $A\cos(2\pi \cdot J_{CH} \cdot t)$, where A is the amplitude, t is the dephasing time and J_{CH} is the one bond coupling of interest, using the minimization procedure in Kaleidagraph (Abelbeck Software). Each experiment (aromatic and anomeric regions) was repeated three times in the presence and absence of orienting phage Pf1. Dipolar couplings were estimated as the difference between the average J_{CH} couplings in the presence and absence of orienting material. The error associated with each dipolar coupling is the standard deviation determined using the three independent measurements made of it. Estimated errors generally fell between 0.75 and 2.5 Hz and tended to be higher near the residual water line and in regions of near chemical shift overlap. NOE buildup curves were calculated using a series of water NOESY spectra. These spectra were recorded at 278 K using a WATERGATE water suppression scheme (Lippens et al., 1995; Piotto et al., 1992) and mixing times of 100, 125, 175, 250 and 350 ms.

Structures were refined using CNS (Brunger et al., 1998) and the RNA-DNA Amber parameter set of Rife et al. (1999). The experimental NOE and dihedral restraints used for E73 were identical to those used earlier by Rife et al. (1999). All refinements began with a random, extended structure. An initial family of structures was generated using only NOE, torsion angle, and weak planarity restraints, following essentially the anneal.inp script that is part of the CNS package. Structures that met acceptance criteria (no NOE violations > 0.5 Å and no torsion angle violations $> 5°$) underwent a further refinement consisting of 20 000 ps of room temperature torsion angle molecular dynamics (Rice and Brunger, 1994; Stein et al., 1997). During this refinement stage square dipolar restraints were gradually introduced linearly from 0.1 to 20 kcal/Hz^2. Note that these values are approximately 80-fold greater than those which Clore and Garrett found optimal (Clore and Garrett, 1999). However, when we repeated these calculations with the force constant range lowered to 0.001 to 1 kcal/Hz^2, no significant benefits resulted. In either case, approximately 90% of the structures produced did not violate experimental data, but since the bond length and angle energies of the RNA-DNA Amber parameter set we used are weaker than those in the more commonly used Berman parameter set, distortions in molecular geometry were common. This necessitated the generation of larger families of structures, and the selection of the low energy members of that family. The final family of 10 grid search structures were taken from a pool of ~130 structures (from refinements at *Rs* of 0–0.15), with no structures contributed from higher *R* values. The final 10 structures from the structure-based refinement were drawn from a pool of 35 structures. Both the best 10 grid and the best ten structure-based structures were selected from their respective families using the criterion of lowest overall energy, and were averaged identically to produce final structures.

Molecular graphics

Molecular graphics images were produced using the MidasPlus (Ferrin et al., 1988; Huang et al., 1991) program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081).

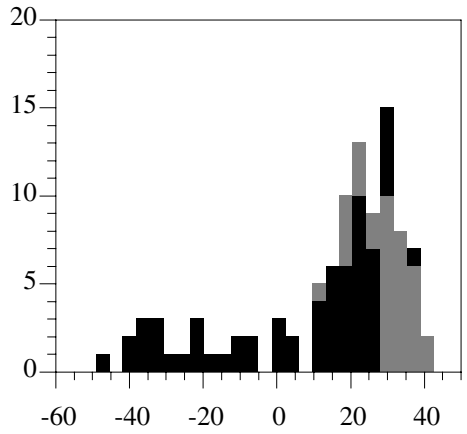

Figure 4. Distribution of the one hundred dipolar couplings observed for E73. Aromatic C-H couplings are shaded. Ribose couplings are shown in black.

Results

Measurement and evaluation of dipolar data

^{1}H-^{13}C dipolar couplings were determined for 100 of the 222 C-H bonds in E73 by comparing J couplings measured in the presence and absence of ∼25 mg/ml phage Pf1 (Hansen et al., 1998, 2000). The number of couplings collected was limited by overlap in the HSQC spectra of the molecule, especially in the H2′, H3′, and H4′ regions. In addition, no effort was made to measure H5′ or 5″-C5′ couplings to avoid averaging problems which could result from the strong 2-bond J-couplings between these geminal protons (Zhu et al., 1994, 1995).

Figure 4 shows the distribution of coupling values observed. Two of its properties are noteworthy. First, it has a characteristic powder pattern shape, which distributions of dipolar coupling data should. Second, the dipolar couplings observed for carbon-proton pairs from the aromatic bases cluster near the right hand side of the distribution, as one anticipates they should for a largely helical nucleic acid. The planes of the bases of nucleic acids are roughly perpendicular to the helix axis, and since the helix axis should orient roughly parallel to the orienting material, θ (Equation 1) will be approximately 90° for base dipolar couplings, yielding values near the positive extreme of the distribution.

Dipolar data in hand, two existing methodologies were used to determine their implications for the structure of E73: singular value decomposition (SVD) and direct refinement. SVD and direct refinement against dipolar data embody contrasting assumptions about a structure. SVD assumes that the structures of the domains of a molecule are known, and uses that information to extract values for D_a^{PQ}, R, and the orientation of the universal coordinate axis system to which they relate from the dipolar data, which leads to information about domain orientations and the dynamics of the molecule. In direct refinement, by contrast, the dynamical properties of the molecule are usually ignored. D_a^{PQ} and R are input from other sources and the orientation of the coordinate axis system and the structure are refined together to obtain the structure that is most consistent with the dipolar data, presumably improving the quality of the structural model.

Singular value decomposition

One application of Prestegard and co-workers' SVD method is determining the relative orientation of domains from dipolar data (Losonczi et al., 1999). Given a set of coupling data and errors, as well as a structural model for each domain, their algorithm uses singular value decomposition to determine whether any orientation of the coordinate axes and values of D_a^{PQ} and R can be found that are consistent with both the data and structure. If a multidomain protein is rigid, then by splitting it into its constituent domains, one can determine the relative orientations of the domains from the relative orientations SVD provides for their coordinate axes. In the case of E73, one could imagine determining its bend by applying SVD to its stem and terminal loop separately. Mollova and coworkers have successfully used a similar approach (Skrynnikov et al., 2000) to determine the relative orientation of the acceptor and anticodon arms of tRNA^{VAL} (Mollova et al., 2000).

Unfortunately, the results of applying SVD to E73 structure were ambiguous. As indicated in Table 1, it was necessary both to exaggerate errors associated with the data and to exclude ∼40% of them from the calculations in order to get any solutions at all. This was true regardless of whether the starting structure used was the crystal structure or the NMR structure (Table 1), and the problem persisted even if the structure was divided into smaller portions (data not shown). The only plausible explanation is that there are significant differences between the true solution structure of E73 and both the published NMR and crystal structures. Indeed, in the original publication on SVD, the authors indicated that the method will

Table 1. Results of application of singular value decomposition (SVD) to the structure of E73. Dipolar couplings and coordinates of either crystal structure or NMR structure of Rife et al. were input into SVD, and the average resulting D_a^{PQ} and R for the loop and helix of E73 are given in the first and second or fourth and fifth rows, respectively. The third and sixth rows give the number of couplings used for each calculation after couplings inconsistent with any solutions were eliminated. The number in parentheses is the total number of couplings collected for the region

		Structure used	
		Crystal	NMR
Loop (10–20)	D_a^{PQ}	−37.2	−36.7
	R	0.2	0.006
	# (42)	35	26
Helix (1–9, 21–29)	D_a^{PQ}	−44.9	−45.5
	R	0.09	0.10
	# (58)	26	28

not work unless the structure being analyzed is already very accurately known (Losonczi et al., 1999). Additionally, a smaller value for D_a^{PQ} emerged for the loop region of E73, regardless of starting structure used, suggesting that it is more dynamic than the stem.

CNS refinement with dipolar data

An alternative approach to SVD is direct refinement of the structure against dipolar data using CNS (Brunger et al., 1998a, b; Tjandra et al., 1997). CNS handles dipolar data in an interesting way. The universal coordinate axis is modeled explicitly as four atoms (representing the origin and the x, y, and z axes), which are included in the simulation. D_a^{PQ} and R are input directly and the dipolar energy term used in simulations to force alterations in molecular structure as a function of the difference between the observed dipolar coupling and the coupling calculated using Equation 1, above.

Determination of the values of D_a^{PQ} and R to input is fairly straightforward in the case of protein structures, since the values can be estimated directly from the distribution of couplings observed (Clore et al., 1998b; Warren and Moore, 2001). However, this approach requires that the orientations of the bonds between coupled nuclei be randomly distributed, which is likely to be true for many proteins, but unlikely to be so for helical nucleic acids. To examine this point we generated an artificial dipolar coupling data set for E73, using the SSIA program of Zweckstetter and Bax (2000). Their program uses a simple steric model for the interaction between macromolecules and liquid crystals to predict D_a^{PQ}, R, and the coordinate axis orientation. The input parameters are the macromolecule's structure, and the type and concentration of the orienting material. SSIA also uses the parameters found to compute the couplings between specified nuclei. Artificial data were generated in this manner for the same 100 C-H pairs for which we had obtained dipolar data experimentally, and the distribution of the computed data was then processed using an algorithm that returns estimates for D_a^{PQ} and R (given random vector orientations) (Warren and Moore, 2001). The values of R estimated from the computed data were consistently ∼0.1 lower than the value of R used to compute them in the first place (data not shown), an error much larger than that expected for a data set this large.

Grid search

Given the failure of structure-independent methodologies like those just discussed to extract accurate values of D_a^{PQ} and R from nucleic acid dipolar data, a grid search approach has been developed for refining nucleic acid structures (Bayer et al., 1999; Clore et al., 1998a; Tjandra et al., 2000). It involves refining structures over and over again using many different values of D_a^{PQ} and R. The rationale is that refinements done with the correct values of D_a^{PQ} and R should yield the lowest energy structures, since those values should be the ones most consistent with the observed NOE, dihedral, and dipolar coupling data as well as with the force fields restraining the geometry of the molecule. Pardi and co-workers have confirmed this hypothesis in refinements of DNA structures with artificial data sets (Vermeulen et al., 2000), which seems reasonable provided that the data lack systematic error. In order to reduce the formidable magnitude of the search to be performed, and given the helical structure of E73 and the abundance of aromatic data available, we assumed that the right extreme of our distribution was well-sampled by our data (see Figure 4 above). If this is so, the search may be restricted to one dimension (see Bayer et al., 1999; Clore et al., 1998a), because when the right extreme of a dipolar coupling distribution is known, there is only one possible value of D_a^{PQ} for each value of R.

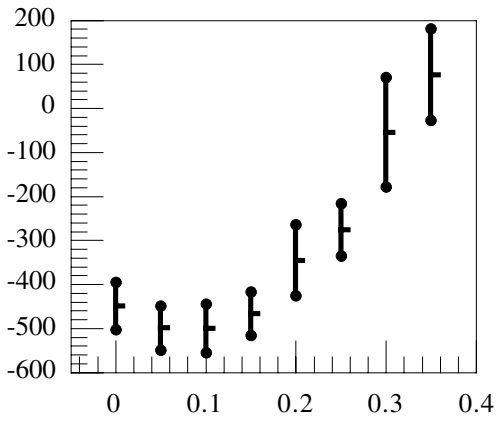

Figure 5. Average energies of accepted structures from grid search refinement of E73 using dipolar data. The center bar gives the average energy and error bars give the standard deviations of the energy for structures refined at each value of R.

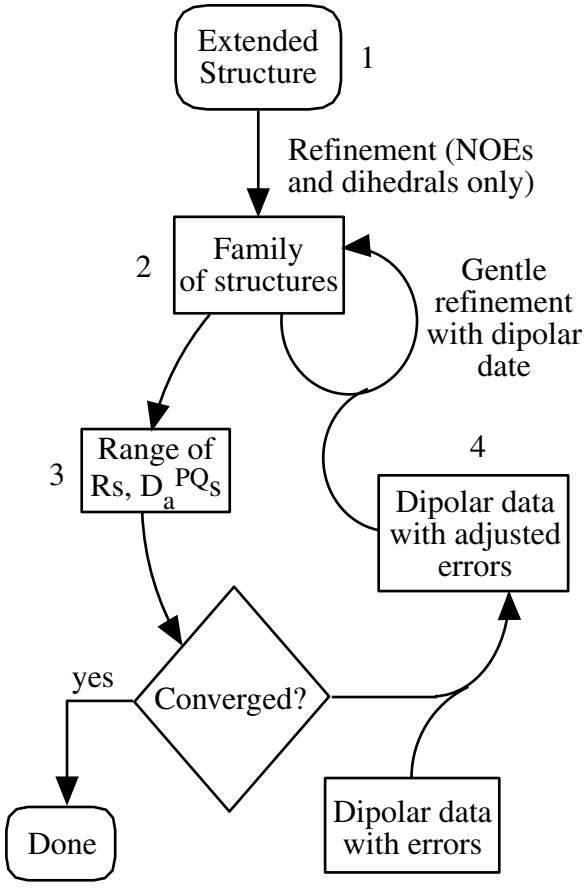

Figure 6. Outline of structure-based refinement procedure described in the text. Numbers indicate the order of each activity. In general one proceeds from 1 to 4 and then back to 2. One then cycles from 2 through 4 until structures converge.

Figure 5 shows the results of this one dimensional grid search. For each value of R (and its concomitant value of D_a^{PQ}) approximately 40 structures were generated, and the average energy of the accepted structures (those which had no NOE violations >5 Å and no dihedral violations > 5 °C), plus or minus the standard deviations, is plotted versus R. The minimum energy is found at an R value of 0.05 to 0.1, corresponding to a value for D_a^{PQ} of −37.2 to −34.8 Hz, but the energy well is quite shallow, and the value of R is not well determined. In fact, the $R = 0$, 0.05, 0.1, and 0.15 refinements all contributed at least one structure to the final family of the ten lowest energy structures.

This final family was averaged, and the average structure is compared to the crystal structure in Table 2. Column 1 in the table reports the statistics for the NOE/dihedral based solution structure reported earlier (Rife et al., 1999), and column 2 gives the statistics for the average structure that emerged from the grid search. Notably, while the all atom RMSD to the crystal structure is slightly reduced in the grid search structure, the torsion angle RMSD actually increases. The bottom left panel of Figure 2 compares the bends of the Correll et al. crystal structure, the Rife et al. NMR structure, and the grid search average structure. Note that the grid search produces a final structure that has a bend very similar to that of the NMR structure refined without dipolar data. It should also be noted that inclusion of dipolar data causes a significant increase in the precision of our family of structures, dropping the average all atom RMSD to the average structure by ∼0.6 Å.

Structure-based refinement

Although the grid search methodology appears to generate accurate structures of nucleic acids, we were uncomfortable using the energy of final structures to determine D_a^{PQ} and R. (We were also unhappy with the large amount of computer time necessary to implement grid searches.) The values assigned D_a^{PQ} and R have a systematic impact on the interpretation of ALL the dipolar data used in a structure calculation, and since the estimates used for D_a^{PQ} and R in a grid search approach derive from that same dipolar data, the covariance matrix for the bond orientation estimates must be very different from that associated with the estimates of torsion angles and interatomic distances that are also input into such a computation. The former will have conspicuous off-

Table 2. Statistics for solution structures of E73. The number and types of restraints used to refine the solution structure of Rife et al., our grid search structure, and our structure-based solution structure are listed in the first three rows. Rotamer errors, row four, refers to the number of backbone or chi rotamers which differed from the rotamers observed in the crystal structure of Correll et al. Torsion angle RMSDs were calculated comparing each structure to the crystal structure, using backbone torsion angles and the chi rotamer for bases 7–23. All atom RMSDs also were calculated between each solution structure and the crystal structure. The all atom RMSD to the average structure, in the final row, is the average RMSD of family members of each family of NMR structures to the average structure in the family

Structure	Rife et al.	Grid search	Structure based
# of Constraints (bases 7–23)			
NOEs	110	110	110
Dihedrals	61	61	61
Dipolar	0	100	100
Rotamer Errors (7–23)	19	19	11
Torsion Angle RMSD (7–23)	35.7	37.60	21.2
All atom RMSD (bases 2–28)	1.62	1.55	1.88
RMSD to average structure	1.99	1.42	1.35

diagonal terms, while the latter should be effectively diagonal. It seems unreasonable, therefore, to trade psuedo-energies that depend on bond orientations on an equal basis with pseudo-energies that depend on discrepancies between specific interatomic distances and individual NOEs, for example, which is what the grid search procedure does. Furthermore, given the complexity of the energy surface explored during refinement, and the incompleteness of the knowledge available about the error structure of the input data on which the shape of that surface in part depends, it is not obvious what the effect would be on final energy estimates of using values for D_a^{PQ} and R obtained independently in a refinement instead of allowing the refinement to determine them itself, and it is likely that this effect will vary from one structure determination to the next. The energy of structures determined could go up; it could go down. Thus we were interested in finding an approach that would simplify the problem by uncoupling structure refinement from D_a^{PQ} and R.

The SSIA algorithm (Zweckstetter and Bax, 2000) seemed to offer a better way of insuring that structures emerge that have global conformations consistent with D_a^{PQ} and R. The procedure we used is outlined in Figure 6. First, a family of structures was generated using only NOE and dihedral restraints, and it was assumed that the true shape of the macromolecule in solution lies within the continuum of shapes spanned by this initial family. Each member of the family was then used as an input for SSIA to predict its R value. Although SSIA also outputs a predicted value of D_a^{PQ}, this value depends on the concentration of orienting material used (to which R is insensitive). Since we had more confidence in our knowledge of the right extreme of the coupling distribution than in our estimate of the phage concentration, we used that information to calculate D_a^{PQ} from R. The ranges of D_a^{PQ} and R values thus determined were mapped into errors in the coupling values, as previously described (Warren and Moore, 2001) because CNS lacks the capability to deal with errors in D_a^{PQ} and R directly, but can take coupling errors into account. A round of refinement with dipolar data followed. The family of accepted structures from this round was run through the SSIA algorithm as outlined above, and if the range of R values tightened, another round of refinement followed, until the structures converged.

We find this methodology appealing, since it explicitly connects the D_a^{PQ} and R predicted from the macroscopic shape of the molecule with the microscopic D_a^{PQ} and R used to refine the structure. Thus, while in a grid search there is no control on whether the final structures are reasonable given the R and D_a^{PQ} values used to generate them, in the structure-based methodology this is explicitly accounted for. Fundamentally, however, this approach is limited by the accuracy of the SSIA algorithm, as well as by

the quality and number of restraints available. When we applied this method to the refinement of E73, our refinement converged in 3 rounds. The range of calculated Rs dropped from 0.086 ± 0.074 to 0.075 ± 0.04. The final range of R values agrees quite well with what one would predict from the grid search results shown in Figure 5, and as one can see in Figure 3B the scatter of structures within the final family was quite small. The structure-based average RMSD to the average structure was 1.35 Å, slightly lower than in the grid search refinement. Again, as with the grid search model, the bend of the structure refined with dipolar data resembles that of the Rife et al. structure much more than that of the Correll et al. crystal structure (Figure 2D). Interestingly, while the all atom RMSD to the crystal structure rises slightly compared to the grid search, the backbone torsion angle RMSD drops dramatically (Table 2).

Structure of the loop

Figure 7 illustrates another interesting result of our refinement with dipolar data. Panels A and B show the G10-U11-A20 base triple in the Rife et al. NMR structure and the Correll et al. crystal structure, respectively. As mentioned above, the two triples differ significantly; in the crystal structure the three nucleotides form a planar base triple, while in the NMR structure three NOEs (dashed lines in Figure 7) from the G10 imino proton to A20 and U11 force G10 to slide beneath the plane of A20 and U11. These three distance restraints would be violated in the crystal structure, Figure 7B. Interestingly, in the average structure of the dipolar refined structure (whether a grid search or our structure-based methodology was employed) the three nucleotides form a planar triple, Figure 7C, but the bases of G10 in the members of the family are seriously distorted. This can occur because the bond and angle energies in the RNA-DNA Amber parameter set used (Rife et al., 1999) are weak enough to allow the geometry of G10 to be distorted to make the base triple planar while satisfying the three NOEs shown. Upon averaging the family of accepted structures and minimizing the bond and angle energies the base 'snaps back' into its correct geometry, in which these NOEs are unsatisfied.

The distortions in these structures led us to reexamine the NOE data on which they depend. NOE buildup curves revealed that the G10-imino U11-imino NOE, which is the most difficult of these distance restraints to satisfy, is unambiguously relayed, probably through

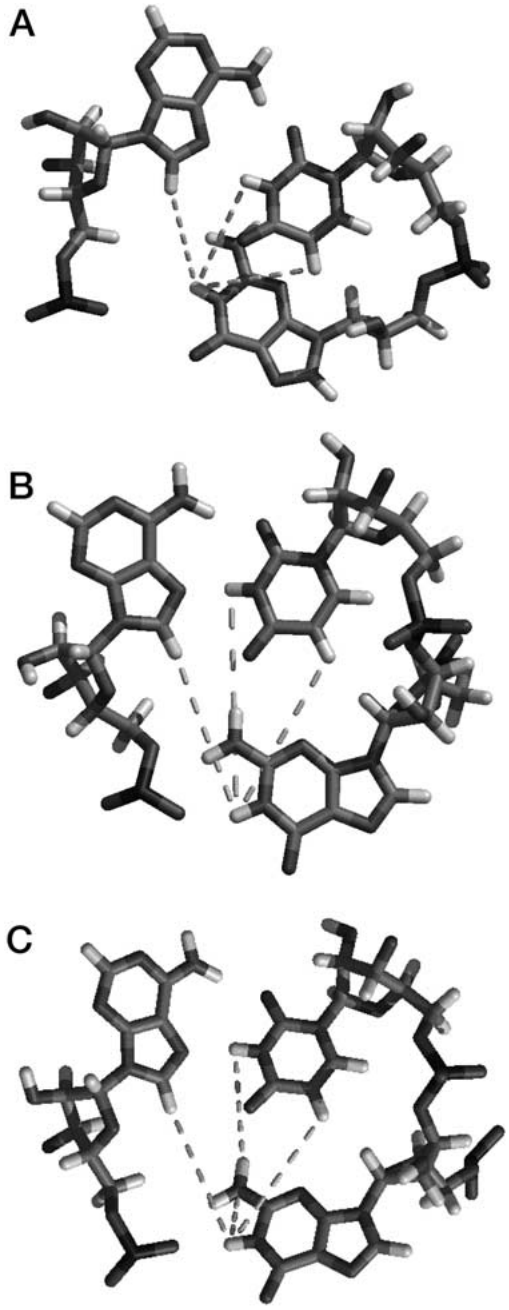

Figure 7. The structure of the central base triple changes when structures are refined with dipolar coupling data. A: Base triple from the Rife et al. NMR structure. Dashed lines indicate the three G10-imino NOEs that force G10 to slide under the plane of U-11 and A-20. B: In the Correll et al. crystal structure, the three bases are roughly planar, and the NMR distance restraints would be violated. C: Refinement with dipolar data produces average structures with planar base triples. NOE buildup curves reveal that the G10-imino-U11-imino NOE is relayed, probably through the G10-amino protons.

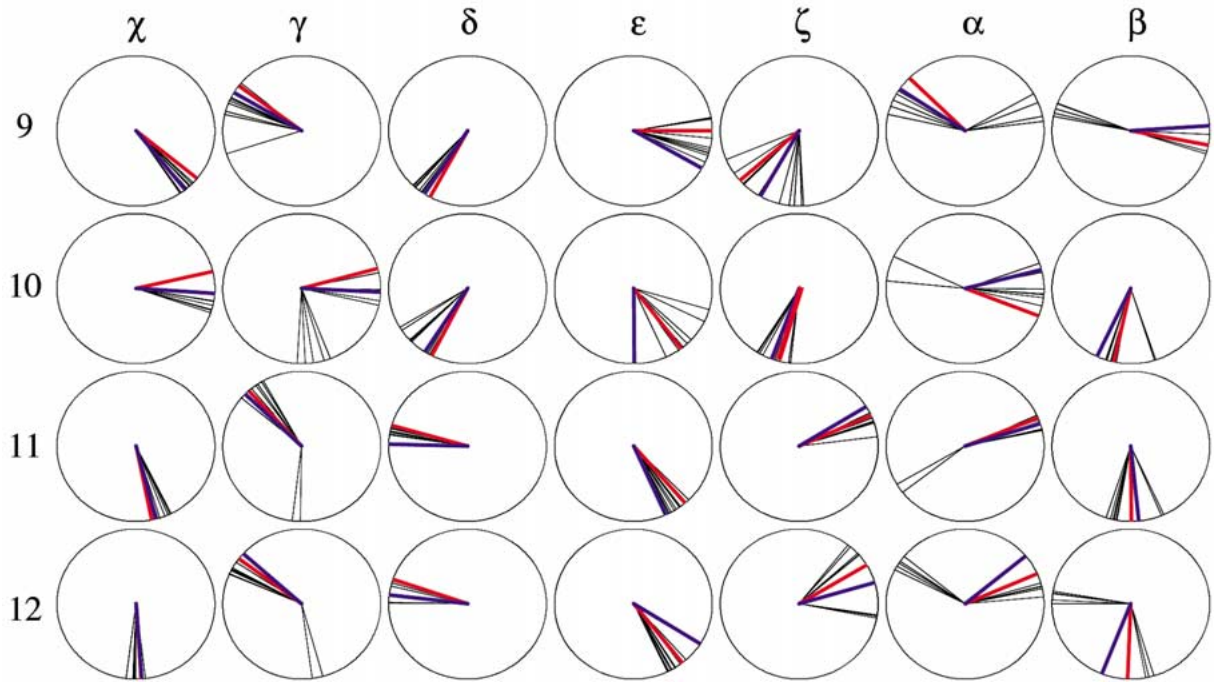

Figure 8. Representative backbone dihedral angles for E73. For each dihedral angle (indicated by Greek letters) and residue (numbers on the left), the torsion angle of each member of the family of structure-based structures (thin black lines), the crystal structure (heavier blue line) and average structure based structure (heavier red line) are shown. A dihedral angle of 0° would be indicated by a vertical line in this representation.

the G amino protons (no NOE is observed at 100 ms mixing time, but an NOE appears at longer mixing times, data not shown). This means that the distance between these two imino protons is significantly larger than originally believed. The G10 H1-A20 H8 and U11-H5 NOEs are more ambiguous due to overlap with the G10 amino protons, however both protons are within 6 Å of the G10 imino, and thus within the conservative distance bounds typically used for exchangeable NOEs today (as compared to the tighter restraints generally used when the first solution structure of E73 was generated (Szewczak and Moore, 1995)). When the structure was recalculated with these distance restraints relaxed and the G10 H1-U11 H3 NOE replaced by a G10 amino-U11 H3 the distortions in the geometry of G10 vanish and the base triple emerges.

Comparison of grid search and structure-based refinement

The average structures that emerge from grid-search and structure-based methodologies are quite similar. The all atom RMSD between the grid search and structure-based structures is only 1.03 Å, which is smaller than the average RMSD of either family to its average, Table 2. The torsion angle RMSD between the two structures, 32.5° for nucleotides 7–23, also compares favorably with the average torsion angle RMSDs for family members to the respective average structures, which are around 38° for both molecules.

Although the two methods produce final structures that are quite similar, there are two striking differences between the refinements. First, the total computation time necessary to generate the structure-based dipolar-refined NMR family was ∼35% of that required to perform the grid search. This difference would have been vastly greater had we undertaken a full two-dimensional grid search, as in (Tjandra et al., 2000, Vermeulen et al., 2000). Second, the torsion angle RMSD between the structure-based and crystal structures is strikingly lower than that between the crystal structure and the grid search average structure.

Discussion

The structure-based (SB) methodology for using dipolar data to refine NMR structures of nucleic acids presented above appears to arrive at final structures

more rapidly and efficiently than existing grid-search methodologies. Since the procedure involves the loosening of dipolar restraints, it is may seem surprising that our final family of structures is slightly more precise than the family generated by the grid search methods. Because of the shallowness of the energy well in Figure 5, however, the lowest 10 energy structures of the grid search were drawn from four separate refinements with R ranging from 0 to 0.15. By contrast, the final family from SB refinement was refined with R constrained to 0.075 ± 0.04.

The improvement in the precision of the solution structures obtained when dipolar coupling data is used in structure determination is obvious from a comparison of Figures 3A and 3B. The increase in precision of the final families from 1.99 Å RMSD to 1.35 Å, while not as large as that predicted for a similar sized DNA duplex (Vermeulen et al., 2000), is still quite significant. Although the accuracy of NMR structures is always difficult to assess, for three reasons we would argue that the accuracy of the E73 structures we have obtained has also improved. First, inclusion of dipolar data in our refinements led to the discovery of an error in local structure arising from spin diffusion in the G-10 U-11 A-20 base triple. Second, the inclusion of dipolar data, in SB refinement at least, greatly improved the agreement between solution and crystal structures, with respect to backbone torsion angles. Third, if the dipolar data were in error, or improperly interpreted, one would expect this to manifest itself both as long-range errors, which might not be obvious, and as errors in local structure, which certainly would be obvious in this case. We observe no such local errors.

The improvement obtained in torsion angle RMSDs using SB refinement requires some explanation. It should first be noted that the average torsion angle RMSD to the average structure is no smaller for the SB family than for the grid search family. Thus, the difference in torsion angle RMSDs relative to the crystal structure appears to result from shifts in average rotamer values, rather than increased precision in the SB family. Figure 8 shows a representative distribution of backbone rotamers from the SB family, taken from nucleotides 9–12. Thin black lines show the torsion angles for each member of the final family. The rotamers of the average SB NMR structure are shown by heavy red lines, and the crystal structure rotamers by heavy blue lines. Note the low precision of those torsion angles which are not experimentally restrained, especially α, β, and γ, and yet the high accuracy of the average structure. The grid searched structure shows comparable precision, but lower accuracy. The difference in accuracy is sufficient to generate a final grid search structure with more backbone torsion angles adopting different rotamers from the crystal structure, and these large differences in torsion angles dominate the measurement of the torsion angle RMSD.

There are two likely explanations for the lower accuracy of the grid search. First, it could be that the one-dimensional grid search did not include the optimal pair of parameters, D_a^{PQ} and R, perhaps because our value for the right extreme of the distribution was in error. A second explanation is that the grid-search method could be less accurate because E73 is dynamic. Both the results from SVD (above) and presence of intermediate sugar puckers in the loop (Szewczak and Moore, 1995) indicate that at least the closing tetraloop of E73 is more dynamic than its stem. If this is the case, then no single value of D_a^{PQ} (which subsumes the order parameter) will be consistent with all the observed coupling data and the true structure. In such a case one would expect that the lowest energy structures would be produced using values of D_a^{PQ} and R which were intermediate between those appropriate for the more and less dynamic portions of the molecule. This in turn would lead to slight inaccuracies in the positioning of bases and sugars which might be sufficient in some cases to push poorly defined torsion angles toward the incorrect rotamers. By contrast, in SB refinements D_a^{PQ} and R are allowed to take a range of values, perhaps allowing relatively more and less dynamic portions of the molecule to be refined with their respective appropriate values of D_a^{PQ} and R.

Variations in order parameters in different parts of the same molecule may be normal in nucleic acids due to their modular construction. The typical nucleic acid studied by NMR consists of relatively rigid helices (except for the terminal base pairs, which are often dynamic) that flank internal and terminal loops, which will generally be more dynamic. Barring direct measurement of the order parameters for the molecule, we expect the structure-based methodology to be more accurate for structures of this type.

Finally, it seems likely that the structure of E73 in solution really does differ from that assumed in crystals, and that E73 in solution is more bent than that in the crystal. This difference appears to be largely the result of small differences in backbone torsion angles along the length of the molecule, rather than

large, local differences such as changes to the central base triple. Note that these small differences in backbone torsion angles are overshadowed by the increased choice of correct rotamers, which yields an improvement in backbone torsion angle RMSD in structures with an increased difference in their bend. Interestingly, the stem loop containing the SRL in the intact large ribosomal subunit from *Haloarcula marismortui* has an even more pronounced bend relative to the crystal structure of E73 than the dipolar-refined solution structure of E73 does, and the bend is in the same direction (Ban et al., 2000).

Acknowledgements

This work was supported by NIH grant GM61258 to PBM and by an NSF predoctoral fellowship to J.J.W. The authors thank Luke Rice for assistance with CNS.

References

Baeyens, K.J., De Bondt, H.L., Pardi, A. and Holbrook, S.R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12851–12855.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) *Science*, **289**, 905–920.

Batey, R.T., Battiste, J.L. and Williamson, J.R. (1995) *Methods Enzymol.*, **261**, 300–322.

Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Magn. Reson.*, **88**, 425–431.

Bayer, P., Varani, L. and Varani, G. (1999) *J. Biomol. NMR*, **14**, 149–155.

Bothner-By, A.A. (1995) In *Encyclopedia of Nuclear Magnetic Resonance* (Eds, Grant, D.M. and Harris, R.K.), Wiley, Chichester, pp. 2932–2938.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

Clore, G.M. and Garrett, D.S. (1999) *J. Am. Chem. Soc.*, **121**, 9008–9012.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **131**, 159–162.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998b) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998c) *J. Am. Chem. Soc.*, **120**, 10571–10572.

Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G. and Steitz, T.A. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 13436–13441.

Correll, C.C., Wool, I.G. and Munishkin, A. (1999) *J. Mol. Biol.*, **292**, 275–287.

Endo, Y. and Wool, I.G. (1982) *J. Biol. Chem.*, **257**, 9054–9060.

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

Gayathri, C., Bothner-By, A.A., van Zijl, P.C.M. and MacLean, C. (1982) *Chem. Phys. Lett.*, **87**, 192–196.

Hansen, M.R., Hanson, P. and Pardi, A. (2000) *Methods Enzymol.*, **317**, 220–240.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.

Hausner, T.P., Atmadja, J. and Nierhaus, K.H. (1987) *Biochimie*, **69**, 911–923.

Holbrook, S.R., Cheong, C., Tinoco, I., Jr. and Kim, S.H. (1991) *Nature*, **353**, 579–581.

Huang, C.C., Pettersen, E.F., Klein, T.E., Ferrin, T.E. and Langridge, R. (1991) *J. Mol. Graphics*, **9**, 230–236.

Lippens, G., Dhalluin, C. and Wieruszeski, J.M. (1995) *J. Biomol. NMR*, **5**, 327–331.

Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 11561–11562.

Piotto, M., Saudek, V. and Sklenar, V. (1992) *J. Biomol. NMR*, **2**, 661–665.

Rice, L.M. and Brunger, A.T. (1994) *Proteins*, **19**, 277–290.

Rife, J.P., Stallings, S.C., Correll, C.C., Dallas, A., Steitz, T.A. and Moore, P.B. (1999) *Biophys. J.*, **76**, 65–75.

Skrynnikov, N.R., Goto, N.K., Yang, D., Choy, W., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **295**, 1265–1273.

Stein, E.G., Rice, L.M. and Brunger, A.T. (1997) *J. Magn. Reson.*, **B124**, 154–164.

Szewczak, A.A. and Moore, P.B. (1995) *J. Mol. Biol.*, **247**, 81–98.

Szewczak, A.A., Moore, P.B., Chang, Y.L. and Wool, I.G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9581–9585.

Tjandra, N. and Bax, A. (1997a) *Science*, **278**, 1111–1114.

Tjandra, N. and Bax, A. (1997b) *J. Magn. Reson.*, **124**, 512–515.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

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

Vermeulen, A., Zhou, H. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 9638–9647.

Warren, J.W. and Moore, P.B. (2001) *J. Magn. Reson.*, **149**, 271–275.

Zhu, L., Reid, B.R. and Drobny, G.P. (1995) *J. Magn. Reson.*, **A115**, 206–212.

Zhu, L., Reid, B.R., Kennedy, M. and Drobny, G.P. (1994) *J. Magn. Reson.*, **A111**, 195–202.

Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.