Comparison of the Crystal and Solution Structures of Two RNA Oligonucleotides

Jason P. Rife,* Sarah C. Stallings,§ Carl C. Correll,#§ Anne Dallas,* Thomas A. Steitz,*#§ and Peter B. Moore*#
Departments of *Chemistry and #Molecular Biophysics and Biochemistry, §Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520-8107 USA

ABSTRACT Until recently, there were no examples of RNAs whose structures had been determined by both NMR and x-ray crystallography, and thus there was no experimental basis for assessing the accuracy of RNA solution structures. A comparison of the solution and the crystal structures of two RNAs is presented, which demonstrates that NMR can produce solution structures that resemble crystal structures and thus validates the application to RNA of a methodology developed initially for the determination of protein conformations. Models for RNA solution structures are appreciably affected by the parameters used for their refinement that describe intramolecular interactions. For the RNAs of interest here, the more realistic those parameters, the greater the similarity between solution structures and crystal structures.

INTRODUCTION

It is common knowledge that protein conformations can be determined by NMR spectroscopy, and experimental studies that proved that the solution structures of proteins are often closely similar to their crystal structures were important in winning protein NMR the credibility it now enjoys (Braun et al., 1989; Wüthrich, 1990). Solution structures are routinely determined for RNAs using the same techniques, and, curiously, they are widely accepted, even though only computational validation studies have been done (e.g., Allain and Varani, 1997).

Some might argue that the comparative studies the protein spectroscopists did suffice for RNA, but RNA and protein differ enough to give one pause. A significant fraction of the information used in solution structure determinations are proton-proton distance estimates derived from nuclear Overhauser effects (NOEs), and the larger the number of distances estimated per unit volume of structure, the more accurate the result. By comparison with what spectroscopists harvest from proteins, the distance sets obtained by RNA spectroscopists are sparse because the number of protons per unit volume in RNA is about half that in protein. In addition, because the average nucleotide is about twice as big as the average amino acid, more of them are relatively uninformative, intraresidue distances. Furthermore, protein spectra are rich in easily harvested information about backbone conformation and relatively poor in information about side-chain structure. Nucleic acid spectra are just the reverse: they are full of information about base positions, but relatively silent with respect to the conformations of their backbones, which have many more degrees of freedom than protein backbones. Finally, most small RNAs are much less globular than proteins of comparable molecular weight, and errors in residue placement, which are inevitable in solution structures, tend to propagate along their lengths. It is not obvious that usefully accurate RNA solution structures can be obtained by spectroscopic methods.

The solution structure of E73, an RNA oligonucleotide that contains the sarcin/ricin loop (SRL) sequence from rat 28S rRNA, was determined several years ago (Szewczak et al., 1993; Szewczak and Moore, 1995; PDB 1SCL), and recently, the crystal structure of the same molecule was solved (Correll et. al, 1998; NDB UR0002) (Fig. 1). In addition, a solution structure has been obtained for AD3, an oligonucleotide that contains the loop E region of *Escherichia coli* 5S rRNA (Dallas and Moore, 1997; PDB 1a4d), and crystal structures have been determined for fragment 1, a 62-nucleotide domain from *E. coli* 5S rRNA that includes most of AD3 (PDB 356D), and a dodecamer that contains AD3's loop E region (Correll et al., 1997; NDB URL064). Thus, for the first time, meaningful solution structure–crystal structure comparisons can be made for RNAs.

In this instance, direct comparison of published structures was not appropriate. The way we interpret spectroscopic data has evolved considerably since the solution structure of E73 was published (compare Szewczak et al., 1993, with Dallas and Moore, 1997). For example, torsion angle restraints and base pair planarity restraints are imposed much less aggressively than in the past, and because the solution structure of AD3 was completed, we have been experimenting with a new set of refinement parameters, which appears to produce more accurate results. For these reasons, this paper consists of two parts. The first describes the optimization of this new parameter set, and in the second, the best solution structures we are currently able to compute for E73

Received for publication 10 August 1998 and in final form 25 September 1998.

Address reprint requests to Dr. Peter B. Moore, Department of Chemistry, Yale University, P.O. Box 208107, 225 Prospect St., New Haven, CT 06520-8107. Tel.: 203-432-3995; Fax: 203-432-5781; E-mail: moore@proton.chem.yale.edu.

Dr. Dallas's present address is Department of Natural Sciences, University of California, Santa Cruz, Santa Cruz, CA.

Dr. Correll's present address is Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL.

0006-3495/99/01/65/11 $2.00

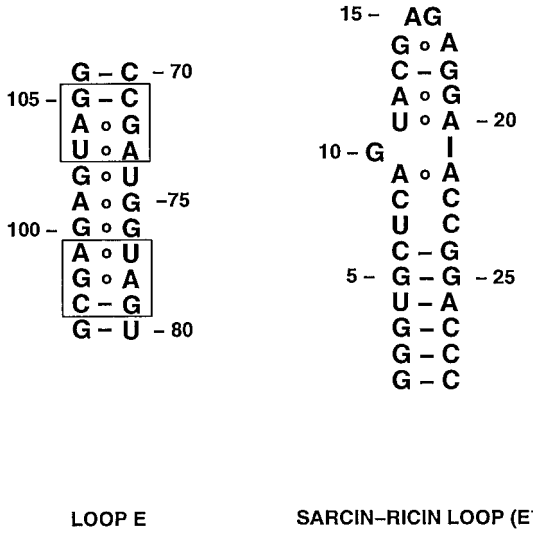

FIGURE 1 Sequences examined in this study. The E73 sequences surrounded by a box are approximately twofold related in the three-dimensional structure.

and the loop E region of AD3 are compared with the corresponding crystal structures.

METHODS

Computations

The structures were computed using either the distance geometry and simulating annealing (DGSA) algorithms in XPLOR (Brünger, 1992) or the torsion angle molecular dynamics (TAMD) algorithms in CNS (Brünger et al., 1998).

Four different parameter/topology sets were used: 1) parallhdg.dna (Brünger, 1992); 2) DNA-RNA.PARAM (Parkinson et al., 1996); 3) RNA-DNA-ALLATOM.PARAM, which is a version of DNA-RNA.PARAM suitable for NMR (J. Rife, unpublished results); and 4) RNA_DNA_amber.par. Parallhdg.dna, which derives from parnah1e.dna (Nilsson and Karplus, 1986) and was distributed with XPLOR for many years, was used to obtain the solution structure reported for the SRL (Szewczak et al., 1993; Szewczak and Moore, 1995). Subsequently, Berman and colleagues discovered that high-resolution nucleic acid crystal structure cannot be refined properly using parallhdg.dna, because it does not specify nucleotide geometries accurately enough. Consequently, they developed DNA-RNA.PARAM, which is based on an analysis of high-resolution nucleotide crystal structures (Parkinson et al., 1996). DNA-RNA.PARAM was used to compute the crystal structures for the sarcin/ricin loop (Correll and Steitz, 1998), the loop E dodecamer, and fragment 1 (Correll et al., 1997), which are discussed below. DNA-RNA-ALLATOM.PARAM was the parameter set used to obtain the published solution structure for AD3 (Dallas and Moore, 1997). The fourth parameter set, RNA_DNA_amber.par, is described below. (RNA_DNA_amber.par and its companion file RNA_DNA_amber.top can be obtained by anonymous ftp from electron.chem.yale.edu. They are to be found in /pub/toppar_files.)

Structures were compared and illustrations prepared using INSIGHT (Molecular Simulations). Two technical points should be noted. First, conventions for designating the prochiral, nonbridging phosphate oxygens in nucleic acids vary from program to program, and during simulated annealing, phosphate groups are likely to invert unless explicitly constrained from doing so. Unless structures being compared designate these oxygens the same way, invalid superpositions will result. Second, some graphics programs superimpose molecules, assuming that their PDB files list atomic coordinates in the same order. Problems can arise when this is not the case.

RESULTS

Structure refinement

The computer programs used to derive solution structures from NMR data search for molecular conformations that are as consistent as possible with both the spectroscopic data available and the information supplied about covalent geometry and intramolecular interactions. The spectroscopic data usually enter these computations as lists of distance and torsion angle estimates abstracted from spectra, but may include measured NOE cross-peak intensities. Crystal structures are refined in fundamentally the same way, but the data considered are measured diffraction amplitudes. Of particular concern here are the parameters fed into these computations that describe nonbonded interactions and the energies associated with distortions of bond lengths and angles. In principle, if these parameters correctly represented the behavior of macromolecules in solution, macromolecular conformations could be computed ab initio, and experimental data could be dispensed with entirely.

The impact of interaction parameters on crystal structures diminishes as resolution increases. At atomic resolution, i.e., at resolutions below 2 Å, the ratio of observations to coordinates to be determined is so large that interaction parameters can be and, indeed, should be dispensed with during final refinement. The impact of interaction parameters on RNA solution structures is much larger because the data are usually only barely adequate to determine conformations. Thus to an extent that is hard to quantify, the details of low-resolution RNA crystal structures and of all RNA solution structures are determined by the energy parameters used during their refinement, and this is why RNA spectroscopists must attend to them.

Parameterization of forces and energies

Berman and colleagues have shown that DNA-RNA.PARAM is superior to parallhdg.dna for crystallographic purposes because its equilibrium bond lengths and angles are more accurate (Parkinson et al., 1996). It is less obvious

that its representation of the interactions responsible for nucleic acid conformation is better. Whereas the force constants in parallhdg.dna that characterize distortions of covalent structure are derived from spectroscopic data and quantum mechanical calculations, the corresponding force constants in DNA-RNA.PARAM derive from the standard deviations of bond lengths and angles observed in the high-resolution nucleotide-related structures deposited in the Cambridge Structural Database and the Nucleic Acid Database. The intent was to ensure that the standard deviations of bond lengths and bond angles that emerge from nucleic acid crystal structures refined using DNA-RNA.PARAM replicate those found in the databases in question (Parkinson et al., 1996). Thus the bond length, bond angle, dihedral, and improper "energies" computed when structures are refined using DNA-RNA.PARAM are figures of merit, not energies, and it is far from obvious how the interactions they represent should be weighted during refinement relative to the more physically based energies ascribed to nonbonded interactions.

RNA_DNA_amber.par

An alternative parameter file has been devised, which is called RNA_DNA_amber.par. The equilibrium geometries it ascribes to nucleotides are identical to those specified in DNA-RNA.PARAM (Parkinson et al., 1996). Its force and energy terms are all taken from the 1995 version of Kollman's AMBER force field (Cornell et al., 1995). The AMBER force field is one of several available that has been optimized for the simulation of molecular dynamics in solution. Because all of its force constants derive from spectroscopy and quantum mechanics and are known to produce accurate results when tested in small molecule simulations, we believe the AMBER force field is likely to represent physical reality more accurately than the force field implicit in DNA_RNA.PARAM.

It is interesting to note that the van der Waals radii implicit in the AMBER force field are all systematically larger than those specified in parallhdg.dna and the parameter sets derived from it. The difference is on the order of 0.1 Å, and it makes a difference. When the first superpositions were done between the crystallographic structure of loop E and solution structures computed using DNA_RNA.PARAM, it became obvious that the interval between adjacent base pairs was somewhat smaller than it should be in the solution structure. This failure of the base pairs in the solution structure to maintain register with the base pairs in the crystal structure disappeared when structures were computed using RNA_DNA_amber.par.

RNA_DNA_amber.par has two other, less obvious advantages over DNA_RNA.PARAM. First, in DNA-RNA.PARAM, some of the energy functions describing torsional rotations have single minima, even though a function with two or three minima is required chemically. In RNA_DNA_amber.par, all chemically plausible rotamers are allowed. Second, in DNA-RNA.PARAM, ribose puckers other than C2′-endo or C3′-endo have unfavorable energies, and during refinement, riboses cannot easily be made to switch between C2′-endo and C3′-endo because the bond lengths and angles ascribed to C2′-endo riboses differ slightly from those attributed to C3′-endo riboses. In RNA_DNA_amber.par, it is assumed that ribose bond lengths and angles are independent of pucker, which is almost if not precisely true, and pucker can vary continuously.

Although we still have much to learn about the mechanics of computing solution structures with RNA_DNA_amber.par, two important facts have emerged. First, by the end of an AMBER refinement, all interactions must be assigned equal weight. Second, by the end of an AMBER refinement, electrostatic interactions must be active. Because hydrogen bonds in AMBER result from Coulombic interactions between electropositive donor hydrogens and electronegative acceptor groups, hydrogen bond donors and acceptors will not interact properly in an AMBER simulation unless electrostatic interactions are "turned on." In this connection, it is interesting that the partial charges assigned nucleotide atoms in parallhdg.dna and DNA-RNA.PARAM differ significantly from those in AMBER.

Charges on phosphate groups

The solution behavior of nucleic acids is notoriously hard to simulate because of the difficulties involved in accounting for the effects of long-range electrostatic interactions between phosphate groups. Consequently, many structural biologists refine nucleic acid structures with the energies ascribed to all electrostatic interactions set to zero or with phosphate groups assigned net charges of zero. This problem could be solved rigorously by including solvent molecules in nucleic acid refinement simulations, but the computational cost would be prohibitive for the typical NMR or crystallographic laboratory, which may do hundreds of cycles of refinement for every RNA structure published. Thus for us to use AMBER force fields at all, a compromise had to be found that avoided the inclusion of solvent molecules.

Because the cations that are always present in nucleic acid solutions tend to neutralize phosphate charges, one is likely to obtain better results from "solvent-free" computations when phosphate groups are assigned charges greater than 1.0e, but less than zero. Similarly, because the dielectric effect of water reduces the magnitudes of all intramolecular Coulombic interactions, the dielectric constant probably should be assigned a value greater than 1.0. What charge should be assigned to phosphate groups, and what should the dielectric constant be?

A practical solution to this problem was sought by computing a series of solution structures for the loop E region of AD3 in which phosphate charges were varied, and comparing the results with the corresponding region of the dodecamer crystal structure. The phosphate charges tested were 0.0e, 0.3e, 0.7e, and 0.9e. The NMR data were held con-

stant, the dielectric constant was fixed at 4.0, and the cutoff for nonbonded interactions was set at 11 Å. The most obvious difference between the four families of structures that emerged was the width of the major groove in the loop E region. Fig. 2 shows a superposition of the backbones of three structures: the dodecamer crystal structure (*red*), the average zero-charge solution structure (*yellow*), and the average solution structure computed with the phosphate charge set to 0.7e (*blue*). Clearly, the 0.7e structure is much closer to the dodecamer crystal structure than the structure obtained when phosphate charges were set to 0, an impression that is supported by data in Table 1. (The structure computed with phosphate charges set at 0.3e is hardly distinguishable from the one computed with phosphate charges set to zero; the heavy atom root mean square deviation (RMSD) between the two (bases 71–79, 97–105) was only 0.342 Å.)

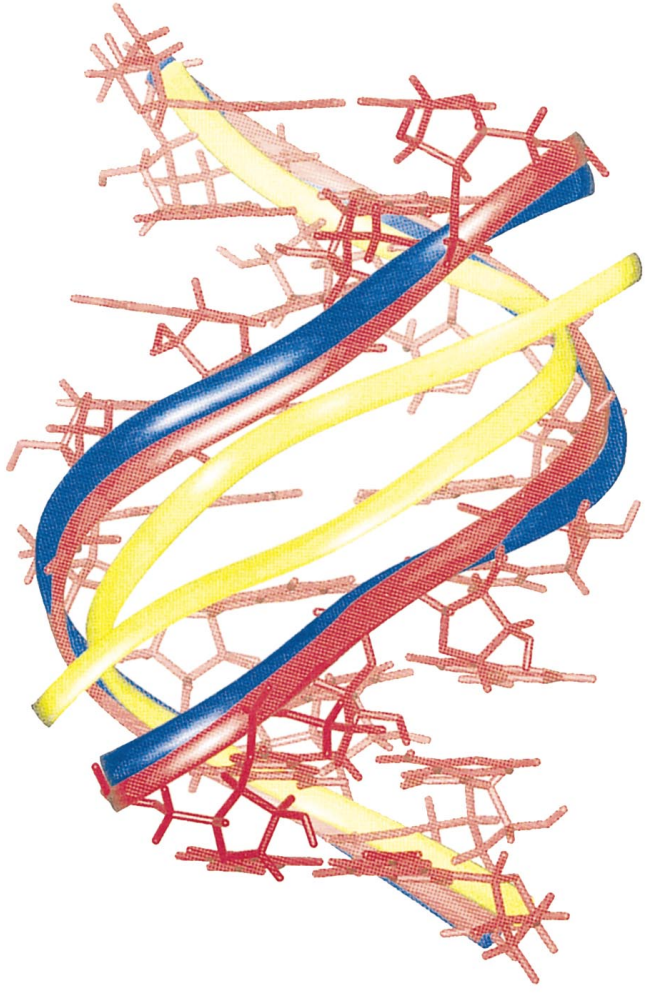

FIGURE 2 A superposition of loop E structures. The x-ray structure of the loop E region of the dodecamer is shown in *red*, with its backbone represented as a continuous ribbon. The backbone of the 0.7e amber solution structure for loop E is shown in *blue*, and the backbone of the 0.0e amber solution structure of loop E is shown in *yellow*. Both the 0.7e and the 0.0e structures were superimposed on the x-ray structure, using all heavy atoms of bases 71–79 and 97–105. The molecule is oriented so that the major groove of loop E faces the viewer.

TABLE 1 Accuracy of solution structures computed using RNA__DNA__amber.par as a function of phosphate charge

Phosphate charge	Heavy atom RMSD (Å)	Torsion angle RMSD (°)
0.0e	2.497	21.2
0.3e	2.342	20.8
0.7e	0.840	13.0
0.9e	1.070	20.0

For each choice of phosphate charge, a family of solution structures was computed for the loop E region of AD3, using the NMR data reported previously (Dallas and Moore, 1997) and RNA__DNA__amber.par. The structures that emerged that had no NOE violations greater than 0.5 Å or torsion angle violations greater than 5° constituted a family, the average member of which was computed and superimposed on the crystal structure of the dodecamer (Correll et al., 1997). The RMSDs between heavy atom positions in bases 71–79, 97–105 in the dodecamer and each average solution structure are tabulated as are the RMSDs in torsion angles.

Surprisingly, the huge difference in major groove width that distinguishes the 0.7e from the 0.0e structure is not associated with an obvious conformational discontinuity. Except for a crankshaft difference in torsion angles in A73 involving β and γ, which has little net effect on the backbone trajectory on either side of A73, the rotamers of all of the torsion angles in the 0.0e and 0.7e solution structures are the same, and the RMS difference in their torsion angles is only 15.9°. Thus the difference in major groove width results from the additive effect of many small torsion angle differences.

The crystals used for the structure determination of fragment 1 contained 1.5 M $MgSO_4$; it is hard to imagine a solute that would neutralize phosphate charges more effectively. Interestingly, the major groove of loop E is much narrower in that structure than it is in the dodecamer structure, which was obtained from crystals grown from solutions of much lower ionic strength (Correll et al., 1997). The central nine base pairs of the loop E region of fragment 1 superimpose on the corresponding base pairs of the zero-charge solution structure of loop E much better than the dodecamer does: 1.6 Å RMSD versus 2.5 Å RMSD. Thus there is reason to believe that the electrostatic effects on major groove width observed in these computations may reflect phenomena that occur in real molecules in solution. It is possible that the charge assigned to phosphate groups during structure refinement should be varied, depending on ionic conditions to which the data refer.

RESULTS

Structure comparisons

The solution structures and crystal structures of E73 should be directly comparable because the same molecule was studied by NMR and crystallography. It is less obvious how loop E comparisons should be done. The conformation of fragment 1 revealed by its crystal structure is distinctly different from that of AD3 both because of the collapse of its major groove in the loop E region, which has already

been noted, and because it lacks an intact terminal loop (loop D). In addition, the crystal structure of fragment 1 is based on a 3-Å resolution electron density map, and conformational details, like backbone torsion angles, cannot be visualized unambiguously at that resolution (e.g., see Schweisguth and Moore, 1997). For this reason, the only part of the AD3 solution structure that can reasonably be compared with a crystal structure is its loop E region, for which the dodecamer crystal structure is available. Its resolution is 1.5 Å.

Loop E structures compared

For these purposes, loop E is defined as the seven consecutive noncanonical base pairs at its center (bases 72–83, 98–104), plus the single Watson-Crick GCs (C71-G79, C97-G105) that flank it on both sides. As Fig. 1 shows, the sequence of loop E has a twofold character. C71-G72-A73 pairs with U103-A104-G105 at the "upstream" end of loop E the same way C97-G98-A99 pairs with U77-A78-G99 at its "downstream" end (Correll et al., 1997). The GA/AU pairs at both ends of loop E are cross-strand A stacks. The six bases in the center of loop E form unusual GG, GA, and GU pairs, which are stabilized both by bifurcated base-base hydrogen bonds and by water-mediated hydrogen bonds (Correll et al., 1997).

Both the solution structure for AD3, which was computed using RNA-DNA-ALLATOM.PARAM (PBD 1a4d), and the 0.7e solution structure described above, which was calculated using the same data and which we will refer to as "the amber structure," imply exactly the same base pairings as found in the dodecamer crystal structure, even though the waters involved in the central base pairs are not visualized (Fig. 3, *bottom*) (Dallas and Moore, 1997). The amber structure is closer to the crystal structure than the earlier solution structure, however. As Table 2 shows, the nine base pairs of PDB 1a4d superimpose on the dodecamer crystal structure with a heavy-atom RMSD of 1.37 Å, but when the amber structure is superimposed on the crystal structure the same way, the RMSD is 0.84 Å. Superpositions done with the upstream cross-strand A stack, the downstream cross-strand A stack, and the three central base pairs reveal that local geometries in amber structure are only slightly closer to the x-ray structure than the PDB structure (Table 2). Thus the amber structure superimposes more accurately on the crystal structure not because it represents local geometries better, but because its representation of the relationship between distant parts of its structure is closer to that in the crystal structure.

The backbone of the loop E region of the dodecamer is A-form-like everywhere except at residues A73 and A93, the two reversed-Hoogsteen A's. The β and γ rotamers in A-form double helix are t and g+, respectively, but for A73 and A99, they are g+ and t, respectively. This departure from A-form geometry causes a distinctive kink in the loop E backbone (Correll et al., 1997). In both solution structures, the β and γ torsion angles of A73 and A99 are the only ones that deviate appreciably from A-form values, but in neither structure have they fully switched to g+ and t. Fig. 3 compares the conformations of the amber structure and the PDB# structure for bases 71–74 and for bases 97–100. The two molecules have been superimposed using bases (71, 72, 74) (Fig. 3, *top*) and bases (97, 98, 100) (Fig. 3, *center*). It is obvious that the difference between the crystal structure and the solution structure is more pronounced at A99 than it is at A73, but even so, as pointed out earlier, the effects barely extend to flanking nucleotides. The relatively benign impact of this torsion angle difference

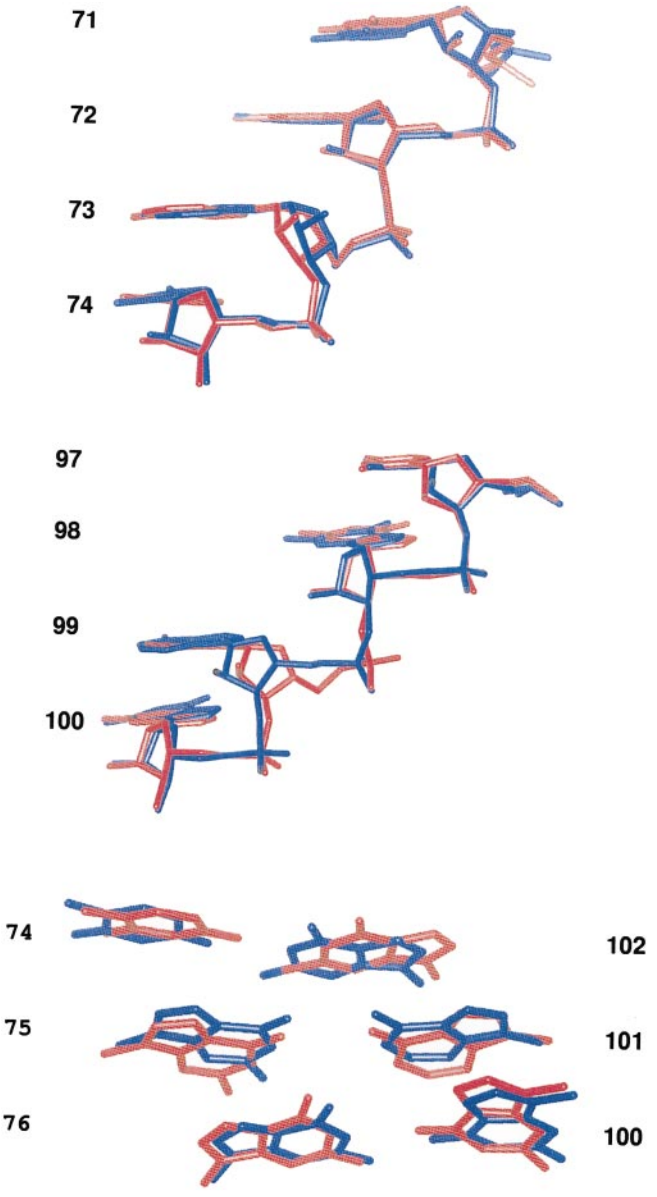

FIGURE 3 Comparisons between the solution structure and the crystal structure of loop E. In all frames, the crystal structure is shown in *red* and the 0.7e amber solution structure is shown in *blue*. (*Top*) Residues 71, 72, and 74 from both structures are superimposed. (*Center*) Residues 97, 98, and 100 are superimposed. (*Bottom*) The positions of the bases in the central three base pairs of loop E are compared.

TABLE 2 Loop E solution structures compared to the dodecamer: RMSDs of selected superpositions

	1SCL (Å)	Amber (Å)	1SCL-73,99 (Å)	Amber-73,79 (Å)
Entire molecule	1.37	0.84	1.40	0.78
Top stacked-A	0.68	0.45	0.65	0.43
Middle pairs	0.85	0.84	—	—
Bottom stacked-A	0.59	0.58	0.54	0.45

The "entire molecule" is bases 71–79, 97–105; the "top stacked-A" is bases 71–73, 103–105; and the "bottom stacked-A" is bases 77–79, 97–99. The "−73,99" superpositions were done with nucleotides 73 and 99 omitted from comparisons.

is supported by superpositions reported in Table 2. When the A99-containing, downstream A-stacks of the two structures are superimposed with A99 omitted, the RMSD of the amber structure falls somewhat, but that of PDB 1a4d hardly changes at all. Upstream A-stack superpositions are virtually indifferent to the omission of A73.

The amber structure discussed above is an average structure computed from a family of nine structures. The average RMSD of the members of that family to the (average) amber structure was 0.71 Å, which is only slightly less than the RMSD of that same average structure to the dodecamer structure, 0.84 Å. Furthermore, because some members of that family have crystal-like torsion angles for A73 and/or A99, it is reasonable to ask whether a structure identical to the dodecamer could have emerged from the computations that produced the amber family. This question was addressed by evaluating the consistency of the dodecamer structure with the NMR data, using the same program and the same data that produced the amber structure. The result is clear. The dodecamer would not have been considered an acceptable member of the amber family because it violates 12 of the 206 distance constraints used to compute solution structures, and the criterion for family membership was no violations. Nevertheless, as you might expect, the nonbonded interaction energies of the dodecamer structure are about the same as those of family members.

Every one of the NOE-derived distances that is inconsistent with the dodecamer structure involves a pair of protons separated by a distance greater than that estimated from the intensity of the corresponding NOE. In every case, there is a third proton lying between the two protons in question that would make a two-step magnetization transfer possible. Clearly, spin diffusion was not adequately allowed for in formulating the distance constraint list used to compute the solution structure of loop E. In this connection, it is interesting to note that the largest distance between protons for which an NOE was observed in AD3 was 6.0 Å, a separation 1.0 Å greater than the maximum allowed for weak NOEs in AD3 computations. Thus with a modest upward adjustment in the distance range assigned to weak NOEs, a distance set could have been generated that is completely consistent with the crystal structure of the dodecamer.

To test the effect of relaxing the upper bound for long distances on AD3 solution structures, the structure of the loop E region of AD3 was computed with the upper bound assigned to long distances raised first to 5.5 Å and then to 6.0 Å. As expected, the families of structures that emerged had wider average RMSDs to their average member, 1.04 Å and 1.11 Å, respectively, instead of 0.71 Å. However, contrary to expectation, the RMSDs of the average members of these two more relaxed families to the crystal structure were somewhat larger than for the amber structure discussed above: 1.04 Å and 1.00 Å, respectively, instead of 0.84 Å. Thus although the crystal structure would have been an acceptable member of both of these families, their average structures were not more crystal-like.

TABLE 3 SRL solution structures

Structure	DG SRL	1SCL	Amber SRL
Refinement method	DGSA	DGSA/full matrix	TAMD
Distance type	Ranges	NOE intensities	Ranges
Param. set used	parallhdg.dna	parallhdg.dna	RNA__DNA__amber.par
Number of exptl. constraints (bases 7–23)			
NOEs/distances	110	110	110
Base pair defs.	8	8	8
Dihedrals	89	77	61
Rotamer exceptions (bases 7–23)	7	2	0
Rotamer errors (bases 7–23)	22	24	19
Torsion angle deviation (RMSD, bases 7–23, backbone and χ)	39.6°	40.2°	35.7°
RMSD to x-ray (in Å)			
Whole molecule	3.06	5.20	1.62
Terminal stem	1.24	1.12	0.89
Entire loop	1.45	1.64	1.46
Tetraloop	1.19	1.14	0.87
Bulged G motif	1.27	1.51	1.31

DGSA implies that the structure in question was determined using distance geometry, simulated annealing. DGSA/full matrix refers to a structure computed using DGSA and then refined by a full matrix relaxation (Nilges et al., 1991). TAMD is a structure obtained using torsion angle molecular dynamics. A "rotamer exception" exists when a torsion angle is restrained to a rotamer different from that found in the E73 crystal structure. A "rotamer error" exists when the rotamer of a torsion angle in the solution structure is not the same as that in the crystal structure.

Superpositions done on the cross-strand A stacks and the middle three base pairs of these new families demonstrated that local geometry had not improved either. It is interesting to note, however, that the backbone torsion angle rotamers in the average structure derived from the family computed with the upper bound for weak NOEs set to 6.0 Å are identical to those in the crystal structure, which is not the case for the amber structure (see above). The RMSD difference in torsion angles between that structure and the crystal structure was only 8.1°.

Comparisons of SRL structures

E73 can be divided into four regions: a stem (bases 1–6, 24–29), a hinge (bases 7–8, 22–23), a bulged G motif (bases 9–12, 19–21), which contains a cross-strand A stack, and a GNRA tetraloop (bases 13–18) (Fig. 1). Three different solution structures have been obtained for E73 with the same NMR data: 1) DG SRL, which was calculated by distance geometry methods using NOE intensities interpreted as distance ranges and parallhdg.dna (Szewczak et al., 1993; Szewczak and Moore, 1995); 2) 1SCL, which is the family of six structures deposited in the Protein Data Bank (PDB# 1SCL) that was refined using full matrix relaxation methods, NOE intensities, and parallhdg.dna (Nilges et al., 1991; White et al., 1992); and 3) amber SRL, which was computed using TAMD methods, NOE data interpreted as distance ranges, and the RNA_DNA_amber-.par.

The torsion angle restraints used for the computation of DG SRL included α and ζ restraints that were set on the basis of phosphorus chemical shifts. The purpose was to find that conformation for E73 that is as A-form-like as possible but does not violate any of the spectroscopic data. The 1SRL family of structures was computed using a set of torsion angle restraints that permitted a wider range of torsion angles in the hinge and the bulged G regions of the SRL than the DG set. In the amber computation, only χ, δ, and ϵ were restrained from bases 7–23, which is to say, the entire 13-nucleotide loop. χ and δ, of course, can be deduced directly from spectroscopic data, and ϵ must be *trans*, unless riboses have C2′-endo puckers, in which case they can be either *trans* or *gauche* (Altona, 1982).

As Table 3 shows, these structures have been superimposed on the crystal structure, and RMSDs have been computed in several different ways. Whole molecules have been compared (bases 2–28), as have stems (bases 2–6, 24–28), the full loop (bases 9–21), the terminal tetraloop (bases 13–18), and the bulged-G motif (bases 9–13, 18–21). The message is clear: the E73 structure computed using RNA_DNA_amber.par is the closest to the crystal structure, even though the input data set used included fewer torsion angle restraints, and many of those eliminated are consistent with the crystal structure (see below). Nevertheless, as was the case with loop E, the amber structure for E73 is closer to its crystal structure, not so much because its representations of local structures are closer, although it is certainly as good as the others in that regard, but because its representation of the relative locations of distant parts is closer.

The SRL loop is linked to the terminal stem of the E73 by a "hinge" that consists of two pyrimidine-pyrimidine juxtapositions. Very few NMR constraints were obtained from that part of the molecule, and in the E73 structures computed using parallhdg.dna there is a huge variation in angle between the axis of the loop and the axis of the stem (Fig. 4 *A*) (Szewczak and Moore, 1995). The members of the RNA_DNA_amber.par family of structures are much less variable in this regard (Fig. 4 *B*) and, on average, are much closer to the crystallographic result (Fig. 5).

The three sets of solution structures are also compared with the crystal structure on the basis of the similarity of their torsion angles to those in the crystal structure (Table 3). In none of the solution structures obtained for E73 do torsion angles replicate those found in its crystal structure very accurately, but it is clear that the amber structure is somewhat closer, as judged both by the average deviation of its torsion angles from the crystallographic values and by the number of torsion angles it contains that have rotamers different from those in the crystal structure. It also appears that the fewer the number of torsion angles restrained to noncrystallographic rotamers, the closer the solution structure to the crystal structure, even if the rule used to eliminate noncrystallographic rotamer assignments removes many entries from the torsion restraint list that are compatible with the crystal structure.

The E73 crystal structure was tested for its compatibility with the NMR data the same way the crystal structure for loop E was, and the result was similar. The crystal structure is not a member of the solution structure family. Its nonbonded energy is low, but it violates a significant number of the NOE-derived distances used to compute the solution structure family. In this instance, the distance range assigned to weak NOEs was 3–6 Å and was consistent with the conclusion drawn from the loop E experience that none of the distances in that class are violated by the crystal structure. Distances in the medium range, 2–4 Å, caused problems; some of the proton pairs assigned to this distance class are separated in the crystal structure by distances outside that range. Multiple step transfers of magnetization are the likely cause of their misassignment. Another source of discrepancies was the distances used to fix the geometry of the noncanonical base pairs. The geometries of these pairs were known only approximately at the time the solution structure of E73 was solved, and the bounds placed on base-pair-defining distances were too tight.

Is there any reason to believe that the true solution structure for E73 differs significantly from its crystal structure? The answer is that they probably do differ a little. In the crystal structure, the only riboses that are C2′-endo are those of A9 and G10. The DQF-COSY data for E73 support that conclusion for A9 and G10, but indicate that the riboses of A15, G16, and A17, which are tetraloop residues, should also be partially C2′-endo. The H1′-H2′ couplings observed

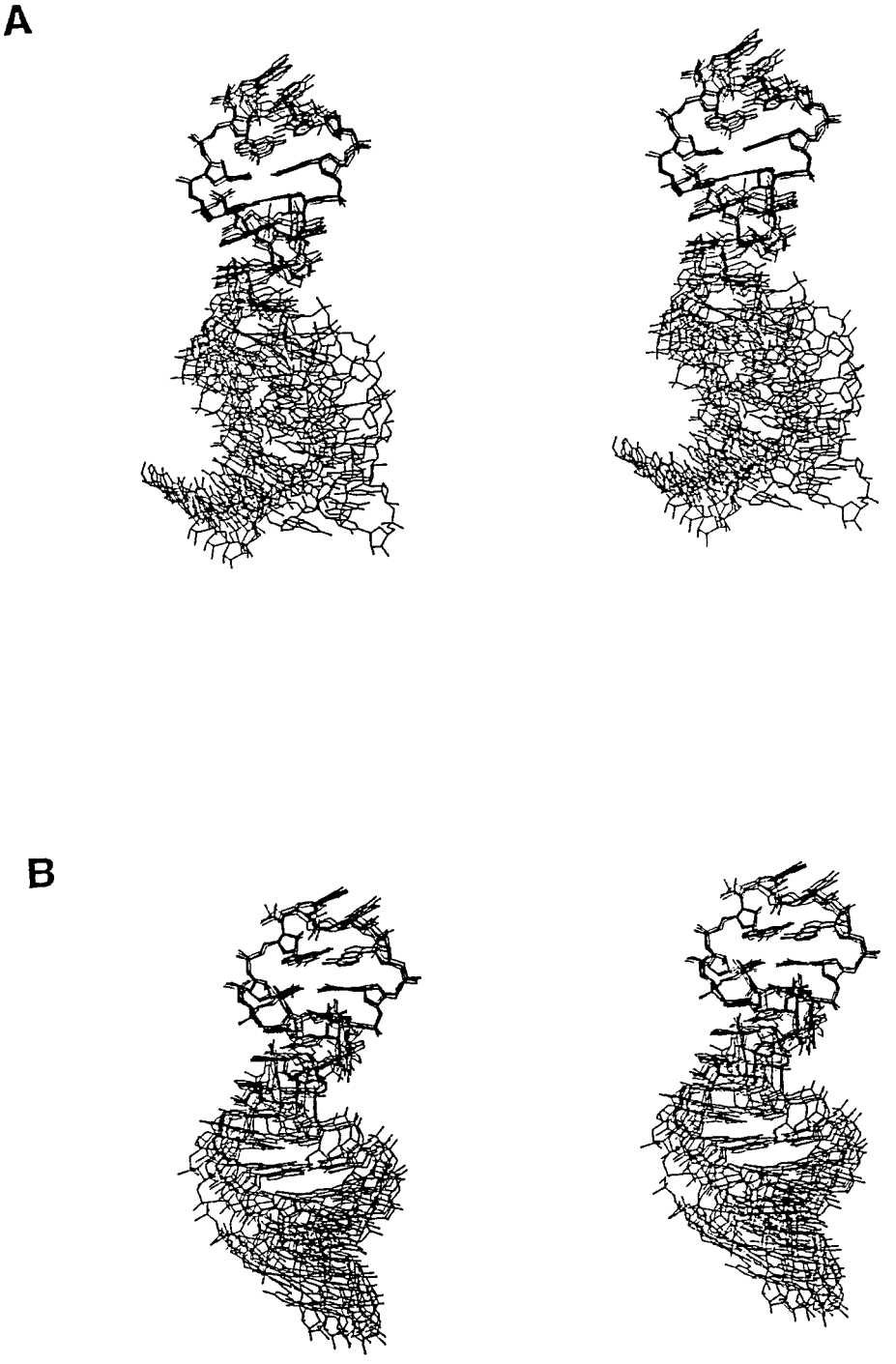

FIGURE 4 Stereo pairs of superpositions of E73 structure families. (*A*) The family of full matrix-refined E73 structures deposited in the Protein Data Bank are shown superimposed on their loop bases, 9–21. (*B*) The family of E73 structures that resulted from AMBER computations is shown, again superimposed on loop bases 9–21.

for these residues suggest that these three residues are C2′-endo ~40% of the time and C3′-endo 60% of the time. Thus the crystal structure of E73 could be identical to one of the conformers that is averaged when E73 is in solution.

Phosphorus chemical shifts

In the past, we have often computed solution structures using phosphorus chemical shifts as indicators of α and ζ torsion angles. When the ^{31}P chemical shift associated with some phosphate lies in the range typical of phosphates in A-form helix, the corresponding α and ζ torsion angles have been constrained to the g- rotamer, which is normal for A-form helices (Gorenstein, 1984). α and ζ have been left unrestrained for phosphate groups with unusual phosphorus chemical shifts. Do the data discussed here support that practice? The loop E structures cast little light on this problem, because all of the phosphate groups in that molecule have A-form like chemical shifts, and the α and ζ torsion angles in the corresponding crystal structure are all (g-, g-). E73, on the other hand, is quite illuminating. The phosphorus spectrum of E73 is well dispersed and fully

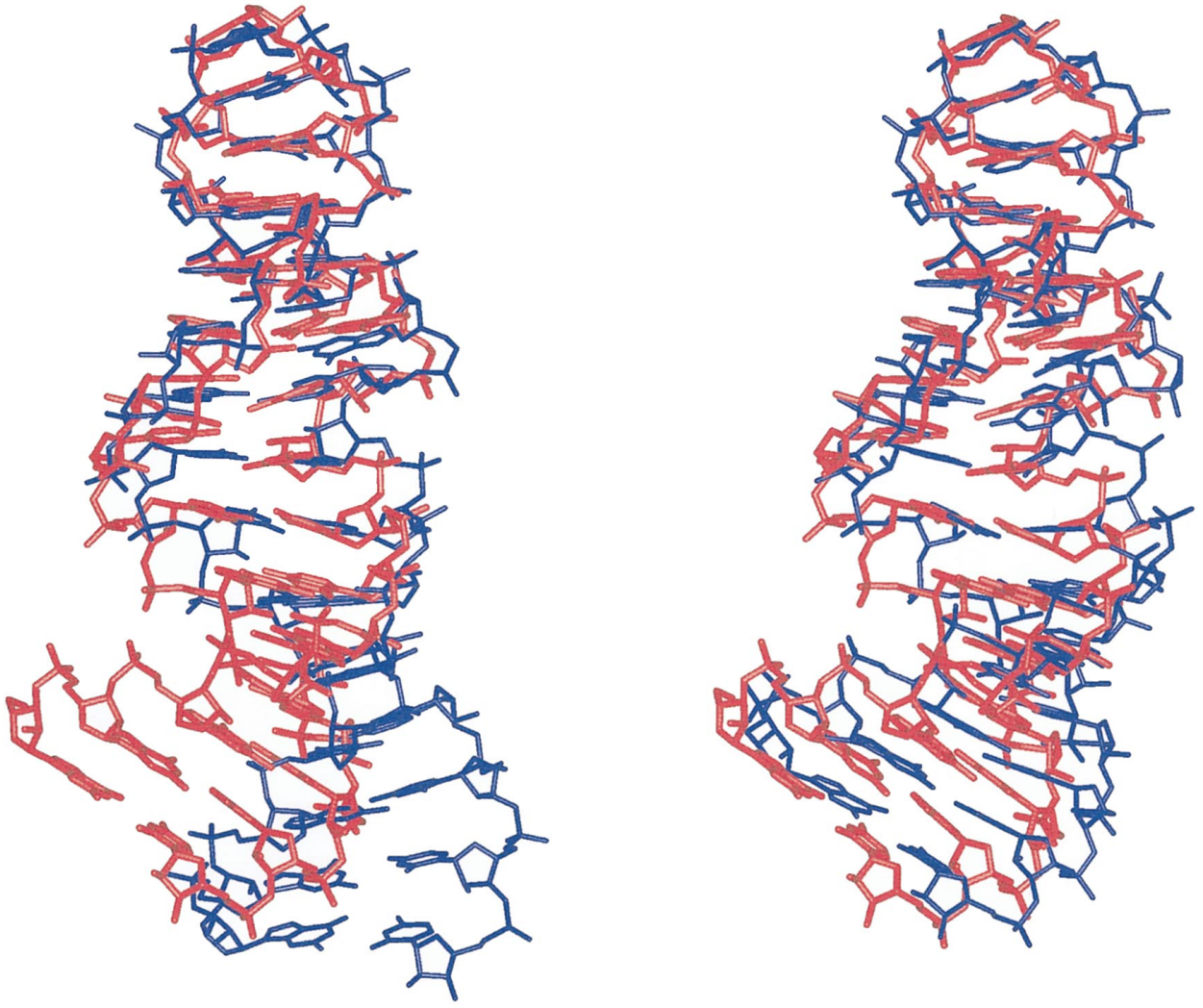

FIGURE 5 Superpositions of E73 solution structure on the E73 crystal structure. The left-hand superposition shows the DG E73 structure (*blue*) superimposed on the E73 crystal structure (*red*), using all bases. The right-hand superposition is an all-base superposition of the E73 crystal structure (*red*) on the average member of the E73 amber family (*blue*).

assigned (Szewczak and Moore, 1995), and its crystal structure contains several phosphate groups with unusual α and ζ torsion angles.

In our initial computations of E73 structures, α and ζ torsion angles were left unrestrained for seven phosphate groups: two because their phosphorus atoms had chemical shifts up-field of the A-form region, and five because their phosphorus atoms were down-shifted. The crystal structure reveals that α and ζ torsion angles of both of the up-shifted phosphate groups are (g-, g-). Apparently up-field shifts of phosphorus atoms do not correlate with departures of α and ζ from A-form values. Of the five phosphate groups that have down-field phosphorus chemical shifts, two have A-form values for α and ζ, and three do not. The mistaken inclusion of two phosphate groups in the "abnormal" set should not have been particularly damaging, because during structure computations α's and ζ's associated with "abnormal" phosphate groups were not restrained. Far more serious is the observation that two phosphate groups that have phosphorus chemical shifts in the A-form range have unusual α's and/or ζ's in the crystal structure. Clearly, no hard and fast rules can be made about correlations between phosphorus chemical shifts and α and ζ values, as others have concluded in the past (Varani et al., 1996).

Granted that phosphorus chemical shifts cannot be interpreted rigorously, can anything useful be done with them? A cautiously affirmative answer may be appropriate. First, up-field phosphorus chemical shifts can probably be ignored; the phosphates with which they are associated are likely to be (g-, g-). Second, the two problematic phosphate groups in E73 are adjacent to phosphates that have both unusual chemical shifts and unusual α's and ζ's in the crystal structure. In addition, they are components of the parts of E73 where its conformation deviates most markedly from A-form helix: its bulged G and its capping tetraloop. In both regions, unusual NOE connectivities and unusual phosphorus chemical shifts signaled the presence of unusual conformation long before any computations were made.

Had we decided to restrain the rotamer ranges of all α's and ζ's to their A-form values, except for those associated with down-shifted phosphate groups and their immediate neighbors on either side, our torsion restraint list would have been error-free.

As pointed out earlier, the amber family of solution structures of E73 was computed using no α and ζ restraints for bases 7–23. Had the rule just described been applied, α and ζ could have been set correctly for six of those residues. This would surely have improved the convergence of that family, and it would have constrained one torsion angle that has a noncrystallographic value in the current average structure to its crystallographic rotamer.

DISCUSSION

Implicit in much of the preceding discussions is the assumption that if the spectroscopic methods used to analyze loop E and E73 were of unerring accuracy, solution structures would have emerged that are identical to their crystal structures. This need not have been true, of course. Crystal packing interactions could have stabilized conformers that are rarely encountered in solution. In addition, conformational differences could also have arisen because the crystal solvents were not the same as those used for NMR data collection. For loop E and E73, however, it appears that both effects are small and that faults in our solution structure methodology account for most of the difference observed between crystal structures and solution structures. Obviously, the parameter sets used to compute solution structures are important, but if the number of restraints derived from NMR spectra could be increased, they would be less critical. In addition, it is clear that errors were made in the geometric interpretation of some of the NMR data and that in some cases restraints were used during computations that in retrospect cannot be justified. Taking all of these factors into account, one concludes that the crystal structures of E73 and loop E are very good models for their solution structures.

As everyone knows, NMR data speak directly to the local geometry of macromolecules and only indirectly to their overall conformations. Crystallographic data are complementary. They speak first to overall conformation and determine the local geometries of macromolecules accurately only if high-resolution data are available. It is not surprising, therefore, that the local geometries of the RNAs discussed above are so well determined by the NMR data that they are nearly independent of the parameter sets used to compute them. As expected, the properties of RNA solution structure models most affected by parameter choice are the spatial relationships of their more distant parts. The inclusion of terms representing electrostatic interactions in parameter sets is extremely important in this regard, because even in the truncated representation of electrostatic interactions used here, they have a far longer range than the NOEs that constitute the bulk of the experimental data available (11 Å versus 5 Å).

The AMBER force field used for these investigations is almost certainly not ideally suited for the simulated annealing computations discussed here, and better force fields may become available in the future. Nevertheless, the results described above suggest that the force fields implicit in the parameter sets used for computing nucleic acid solution structures should be as realistic as possible, and that better results are obtained when electrostatic interactions are taken into account.

In this case, a plausible treatment of electrostatic interactions was found by "tuning" the solution structure of loop E so that it would match its crystal structure. Remarkably, when that treatment of electrostatic interactions was applied to E73, the structure of its hinge region was stabilized, which suggests that the treatment used has some validity. The average RMSD of members of the E73 amber family to the family average is slightly less than 1 Å, but the corresponding RMSD for the family computed with electrostatic interactions turned off (PBD 1SCL) is greater than 4 Å, primarily because of variation in the angle between the stem and the loop. Furthermore, the hinge angle in the E73 amber family is almost the same as in the E73 crystal structure, which again suggests that the structure obtained is better than those obtained when electrostatic interactions are ignored. Nevertheless, it should be recognized that the solution found here for the electrostatic problem is ad hoc. It lacks a sound theoretical justification, and for that reason, there is no assurance it will lead to good results when applied to other RNAs.

Of the several structure computation and refinement methods explored, the best appears to be torsion angle molecular dynamics. In our experience, the results it produces are not markedly different from those generated by distance geometry methods, but the yield of acceptable structures is higher (Dallas and Moore, 1997; Stallings and Moore, 1997). It appears that accurate solution structures should emerge from such computations, provided the upper bound allowed for NOE-determined distances is ~6.0 Å, the maximum distance separating any pair of protons for which an NOE was observed in the spectra of AD3 and E73. Disappointingly, in our hands, the refinement of structures by full matrix methods, which ought to take care of spin diffusion effects better than distance classification methods, led to less accurate structures, but this is an issue that should be revisited.

The structures compared also offer some insight into the question of whether α and ζ torsion angles can be restrained on the basis of phosphorus chemical shifts. It is clear that the hypothesis that down-field phosphorus shifts prove non-A form values for α and ζ is not valid, but a somewhat more relaxed approach to this problem may prove useful. In the RNAs examined here, it would have been satisfactory to constrain all α and ζ angles to A-form values, except for those associated with phosphate groups having down-field chemical shifts, and their neighbors on either side.

Finally, the computational study of Allain and Varani (1997) referred to earlier indicated that solution structures

should resemble crystal structures, just as the experimental results discussed above do, but their study cannot be regarded as definitive. Many of their conclusions were based on computations the objective of which was to reproduce the structure of an RNA using NMR-like data derived from its crystal structure. Proton pairs that might reasonably have resulted in an observable NOE were assigned to specific distance classes on the basis of their separations in the crystal structure. The set of distance restraints that emerged is far superior to what an experimentalist is likely to obtain. The distance ranges used were nonoverlapping, and experimentalists must use overlapping distance ranges because of the uncertainties that surround the interpretation of NOE cross-peak intensities. Furthermore, the model distance set was accurate; no distances were assigned to the wrong distance range, which is unlikely to happen when real data are analyzed. In defense of Allain and Varani, it must be pointed out that it is all but impossible to compute a test, distance restraint set starting from a known structure that faithfully replicates all of the perversities of distance sets derived from real NOE data. It follows that, as we contended at the outset, there was/is a real need to validate RNA solution structures experimentally.

It is now clear that using 1998 methodology, solution structures of useful accuracy can be determined by NMR, and they are likely to be closely similar to crystal structures of the same molecules. There is every reason to anticipate that in the future, spectroscopists will harvest more conformational information from RNA spectra than they do today, and that more accurate solution structures will result.

We thank Dr. Julian Tirado-Rives and Prof. William Jorgensen for advice about parameter sets and Prof. Axel Brünger for counsel about refinement procedures. S.C.S. was supported by Howard Hughes Medical Institute, and C.C.C. is the recipient of American Cancer Society Postdoctoral Fellowship 3840.

This work was supported by Grants GM41651 (to P.B.M.), GM54216 (to P.B.M.), and GM22778 (to T.A.S.) from the National Institutes of Health.

REFERENCES

Allain, F. H.-T., and G. Varani. 1997. How accurately and precisely can RNA structures be determined by NMR? *J. Mol. Biol.* 267:338–351.

Altona, C. 1982. Conformational analysis of nucleic acids. Determination of backbone geometry of single-helical RNA and DNA in aqueous solution. *Recueil. J. R. Netherl. Chem. Soc.* 101:413–433.

Braun, W., O. Epp, K. Wuthrich, and R. Huber. 1989. Solution of the phase problem with the x-ray diffraction method for proteins with the nuclear magnetic resonance solution structure as initial model. Patterson search and refinement for the alpha-amylase inhibitor tendamistat. *J. Mol. Biol.* 206:669–676.

Brünger, A. 1992. X-LOR Version 3.1. System for X-Ray Crystallography and NMR. Yale University Press, New Haven, CT.

Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr.* D54:905–921.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmayer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of protein, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Correll, C. C., B. Freeborn, P. B. Moore, and T. A. Steitz. 1997. Metals. motifs and recognition in the crystal structure of a 5S rRNA domain. *Cell.* 91:705–712.

Correll, C. C., A. Munishkin, Y. Chan, Z. Ren, I. G. Wool, and T. A. Steitz. 1998. Crystal structure of the ribosomal RNA loop essential for binding both elongation factors. *Proc. Natl. Acad. Sci. USA.* (in press).

Dallas, A., and P. B. Moore. 1997. The solution structure of the loop E/loop D region of *E. coli* 5S rRNA. *Structure.* 5:1639–1653.

Gorenstein, D. G. 1984. Phosphorus-31 NMR: Principles and Applications. Academic Press, New York.

Nilges, M., J. Habazattl, A. T. Brünger, and T. A. Holak. 1991. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* 219:499–510.

Nilsson, L., and M. Karplus. 1986. Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comp. Chem.* 7:591–616.

Parkinson, G., G. Vojechnovsky, L. Clowney, A. T. Brunger, and H. M. Berman. 1996. New parameters for refinement of nucleic-acid-containing structures. *Acta Crystallogr. D.* 52:57–64.

Schweisguth, D. C., and P. B. Moore. 1997. On the conformation of the anticodon loops of initiator and elongator methionine tRNAs. *J. Mol. Biol.* 267:505–519.

Stallings, S. C., and P. B. Moore. 1997. The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA. *Structure.* 5:1173–1185.

Szewczak, A. A., and P. B. Moore. 1995. The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* 247:81–98.

Szewczak, A. A., P. B. Moore, Y.-L. Chan, and I. G. Wool. 1993. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl. Acad. Sci. USA.* 90:9581–9585.

Varani, G., F. Aboul-ela, and F. H.-T. Allain. 1996. NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.* 29:51–127.

White, S. A., M. Nilges, A. Huang, A. T. Brünger, and P. B. Moore. 1992. NMR analysis of helix I from the 5S RNA of *Escherichia coli*. *Biochemistry.* 31:1610–1620.

Wüthrich, K. 1990. Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* 265:22059–22062.