Journal of Biomolecular NMR, 9 (1997) 229–244
ESCOM

J-Bio NMR 423

The dynamic NMR structure of the TΨC-loop: Implications for the specificity of tRNA methylation

Letitia J. Yao, Thomas L. James, James T. Kealey, Daniel V. Santi and Uli Schmitz*

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.

Received 16 July 1996
Accepted 8 November 1996

Keywords: Dynamic NMR refinement; MD with time-averaged restraints; TΨC-loop; RUMT; RNA/protein recognition

Summary

tRNA (m^5U54)-methyltransferase (RUMT) catalyzes the *S*-adenosylmethionine-dependent methylation of uridine-54 in the TΨC-loop of all transfer RNAs in *E. coli* to form the 54-ribosylthymine residue. However, in all tRNA structures, residue 54 is completely buried and the question arises as to how RUMT gains access to the methylation site. A 17-mer RNA hairpin consisting of nucleotides 49–65 of the TΨ-loop is a substrate for RUMT. Homonuclear NMR methods in conjunction with restrained molecular dynamics (MD) methods were used to determine the solution structure of the 17-mer T-arm fragment. The loop of the hairpin exhibits enhanced flexibility which renders the conventional NMR average structure less useful compared to the more commonly found situation where a molecule exists in predominantly one major conformation. However, when resorting to softer refinement methods such as MD with time-averaged restraints, the conflicting restraints in the loop can be satisfied much better. The dynamic structure of the T-arm is represented as an ensemble of 10 time-clusters. In all of these, U54 is completely exposed. The flexibility of the TΨ-loop in solution in conjunction with extensive binding studies of RUMT with the TΨC-loop and tRNA suggest that the specificity of the RUMT/tRNA recognition is associated with tRNA tertiary structure elements. For the methylation, RUMT would simply have to break the tertiary interactions between the D- and T-loops, leading to a melting of the T-arm structure and making U54 available for methylation.

Introduction

tRNA (m^5U54)-methyltransferase (RUMT) catalyzes the AdoMet-dependent methylation of uridine-54 in the TΨ-loop of all transfer RNAs in *Escherichia coli* to form the m^5U54 ribosylthymine residue. RUMT methylation involves the initial formation of a covalent Michael adduct between an enzyme nucleophile (Cys^{324}) and the 6-carbon of the target U54 residue activating the 5-position for subsequent methyl transfer (Kealey et al., 1994). The proposed cis mechanism involves precise positioning of the catalytic cysteine and AdoMet with respect to C5 and C6 of U54.

Crystal structures (Quigley and Rich, 1976; Holbrook et al., 1978) of various tRNAs show that C6 of U54 is completely solvent inaccessible and C5 is almost completely buried. NMR studies of imino protons (Heerschap et al., 1983a,b; Roy and Redfield, 1983; Hyde and Reid, 1985; Hall et al., 1989; Amano and Kawakami, 1992) confirm this picture since some of the loop imino protons are fairly protected from the exchange with bulk solvent. U54 is stacked between G53 and U55 and is involved in a reverse Hoogsteen hydrogen bond with residue A58. Additionally, four of the seven loop residues are actually outside of the T-loop, in particular residues 59 and 60, such that A58 stacks directly onto the T-stem base pairs.

*To whom correspondence should be addressed.

Supplementary material available from the authors: six pages and four figures, describing the details of the assignment process.

Abbreviations: tRNA, transfer ribonucleic acid; RUMT, tRNA (m^5U54)-methyltransferase; AdoMet, *S*-adenosylmethionine; OD, optical density at 260 nm at 25 °C; MWCO, molecular weight cutoff; NOESY, nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; DQF-COSY, double-quantum filtered correlation spectroscopy; rMD, restrained molecular dynamics; MDtar, molecular dynamics with time-averaged restraints; rmsd, root-mean-square deviation; KP, potassium phosphate; T_m, UV melting point; T_1, longitudinal relaxation time; T_2, transverse relaxation time.

U55, which is post-transcriptionally modified to pseudouridine, Ψ, and C56 are hydrogen-bonded to G18 and G19, respectively, of the D-loop. The U54-A58 base pair stacks over the G53-C61 base pair to stabilize the conformation of the T-arm (Romby et al., 1987).

Extensive mutation studies (Gu et al., 1996) of the T-loop nucleotides in tRNA and a 17-oligonucleotide T-arm mimic revealed that the primary sequence requirements for interaction with RUMT are far less stringent than it would appear from the tRNA consensus sequence. No particular base composition is needed for the stem as long as base pairing is maintained. A loop of seven nucleotides is essential for binding and catalysis. For intact tRNA, U54 is the only base in the T-loop that is essential for RUMT substrate activity. With the exception of C56, any single base of the loop can be substituted by any other base without a significant decrease in K_m or complete loss of activity.

These results indicate that the structural requirements for RUMT substrates are not based upon the primary sequence of the T-loop. Furthermore, it seems very unlikely that all of these sequences give rise to a similar, stable structure. On the other hand, the absence of base sequence specificity makes the question of how RUMT gains access to the methylation site even more important. It is clear that RUMT must disrupt the local RNA structure and there are strong indications that RUMT plays an essential role in disrupting tertiary interactions between the D-loop and the T-arm (Kealey et al., 1994). Structural details of the RUMT/T-arm interaction can only be derived from high-resolution structures of the complex, whose molecular weight of 48 kDa makes NMR structural studies difficult. However, first insights can be gained by comparing the solution structure of the 17-mer T-arm with that in the complete tRNA crystal structure. Here we present the structure determination of a 17-mer T-arm RNA fragment using high-resolution NMR methods.

In our previous work (James, 1991; Schmitz and James, 1995), we have demonstrated that more accurate NMR structures can be determined when high-precision restraints are employed, available via complete relaxation matrix methods. Since our NMR work indicated early on that the loop of the 17-mer T-arm construct exhibits typical signs of pronounced conformational flexibility, we also explored the dynamic refinement of the structure using time-averaged restraints (Torda et al., 1990) in our rMD calculations. This is the first application of this soft refinement approach to an RNA molecule.

Materials and Methods

RNA synthesis and purification

The 17-mer was synthesized by in vitro transcription using T7 RNA polymerase on a synthetic DNA template according to the method of Milligan et al. (1987). The transcription reaction was optimized to yield 8 OD/ml reaction mixture using final conditions of 0.4 M Tris-HCl (pH 8.1), 36 mM $MgCl_2$, 50 mM dithiothreitol, 10 mM spermidine, 6 mM each of ATP, CTP, GTP, and UTP, 0.1% Triton X-100, 800 nM DNA template, and 5 μg/ml T7 RNA polymerase. The mixture was incubated at 36 °C for 4 h in a total volume of 45 ml.

After precipitation with ethanol and centrifugation, the RNA-containing pellet was dissolved in 8 ml of 7 M urea and purified on five 20% polyacrylamide / 7 M urea gels ($43 \times 35 \times 0.15$ cm). Excised bands were electroeluted, precipitated, and finally dialyzed against a cascade of buffers (10 mM KP, 5 mM EDTA, 100 mM NaCl; 10 mM KP, 0.1 mM EDTA, 100 mM NaCl; 10 mM KP, 50 mM NaCl; water (2×)) for 24 h each at 4 °C, pH 6.4, using a 1000 MWCO membrane. The final RNA material still contained ~10% of the 'n + 1' transcription product.

NMR sample preparation

The 17-mer RNA was lyophilized several times from 0.5 ml of D_2O for analysis of nonexchangeable protons. The lyophilized powder of the 17-mer RNA was dissolved in 3 ml of D_2O with TSP as the internal standard and 10 mM KP at pH 6.4 for a final sample concentration of 0.3 mM. The sample was annealed immediately prior to use by heating at 75 °C for 10 min and snap-cooling on ice for 30 min. For analysis of exchangeable protons, the sample was lyophilized and dissolved in 9:1 H_2O:D_2O containing TSP and 10 mM KP at pH 6.4.

Thermodynamic measurements

Thermal denaturation profiles were measured on a Cary 3E UV–visible spectrophotometer equipped with a temperature controller, at a heating rate of 1 °C/min with detection at 260 nm. Data were collected on 1 ml samples in three buffers at pH 6.4 (10 mM KP; 10 mM KP, 50 mM NaCl; 10 mM sodium cacodylate, 5 mM $MgCl_2$) over the temperature range 5–97 °C.

Gel filtration HPLC was performed on a Superose 12 column, running at 0.5 ml/min at a pressure of ~28 bar with detection at 260 nm. Samples (1–5 mg/ml) were prepared in the running buffers at pH 6.4 (10 mM KP; 10 mM KP, 50 mM NaCl; 10 mM KP, 10 mM $MgCl_2$), heated to 75 °C for 7 min, and then either cooled quickly (snap-cooled on ice) or slowly (overnight at RT).

NMR experiments

^{1}H NMR experiments were acquired at 600 MHz on a Varian Unityplus spectrometer. All 2D NOE spectra (τ_m = 50, 150, 200, 250, and 400 ms) in D_2O were acquired in the hypercomplex mode (States et al., 1982) at 25 °C, using a spectral width of 5999 Hz in both dimensions with the carrier frequency set to the HDO resonance frequency. A total of 512 t_1 values were recorded for each FID. Thirty-two scans with a repetition time of 2.8 s were

recorded for each t_1 increment with 4K data points in F2. Double-quantum filtered COSY spectra were acquired under the same conditions, but with a 2.0 s relaxation delay. Total coherence transfer spectroscopy (TOCSY) experiments were run with MLEV-17 mixing and cycling (Bax and Davis, 1985) with mixing times of 30 and 75 ms.

All 2D NOE experiments in H_2O were collected at 10 °C using the SSNOESY pulse sequence (Smallcombe, 1993) with a spectral width of 12 000 Hz using a symmetrically shifted S-pulse with a pulse width of 88.8 μs for water suppression. Excitation maxima occur at ±7576 Hz from the carrier frequency. The mixing times were 150 and 400 ms.

1D NMR experiments in H_2O were acquired with a 1 $\bar{3}$ 3 $\bar{1}$ pulse for solvent suppression on a GE GN-500 MHz spectrometer. Longitudinal relaxation experiments were carried out using the inversion recovery method.

2D data sets were transferred to a SUN SPARCstation 2 and were processed using Striker and Sparky (Kneller, 1992). A Gaussian window function was used for resolution enhancement (Gaussian multiplier of 0.2 and line broadening of −5.0) in both dimensions for TOCSY and SSNOESY spectra. For DQF-COSY, the data were apodized with a 30–45° shifted sine-bell function. NOESY spectra were processed with a 60° shifted sine-bell function. Prior to Fourier transformation, the FIDs were zero-filled to give a final 2K × 2K data set. The base planes of all spectra were subsequently corrected using polynomial fitting.

The intensities of nonexchangeable proton cross peaks from the 250 ms NOESY data set were integrated with Sparky. Overlapping peaks were integrated by performing a multiple line-fitting of the cross peaks to a Gaussian function and subsequent calculation of the area below the theoretical curve.

Restraint generation

Preliminary distance restraints were obtained with the program MARDIGRAS (Liu et al., 1990,1994), which uses a complete relaxation matrix approach assuming overall isotropic molecular motion. Hybrid relaxation matrices were built using the cross-peak volumes of the 250 ms 2D NOE data set in combination with either an extended RNA chain or the T-arm coordinates of the crystal structure of $tRNA^{Phe}$ (PDB entry 1tra). Correlation times of 2, 3, and 4 ns, which were estimated from T_1 and T_2 measurements (Schmitz and James, 1995), were utilized along with an intensity error of twice the volume of the smallest peak. Distance restraints were generated by averaging the results of the six MARDIGRAS runs using twice the standard deviation as bounds.

A second generation of distance restraints was generated via MARDIGRAS using a preliminary rMD-refined NMR average structure of the T-arm as the starting model. Here, restraint bounds were set in a more conservative and adequate way, utilizing the randomardi procedure (Liu et al., 1995). This latest feature in MARDIGRAS lets the user define bounds for the actual intensities according to a user-definable percentage error on top of an absolute noise error. MARDIGRAS calculations are then repeated many times for the slightly different intensity sets, arising from choosing random values of the intensities from within the intensity bounds. Here, 50 random intensity sets were created for the 250 ms 2D NOE data set using 10% as the minimum percentage error for all intensities in conjunction with an absolute noise error of the size of the smallest peak. Besides repeating these MARDIGRAS calculations for correlation times of 2, 3, and 4 ns, we also employed faster correlation times for the loop hydrogens (2, 3 ns) versus the stem hydrogens (4 ns), assuming that the larger flexibility of the loop would influence the relaxation. Bounds were constructed using the average of the minima and maxima of all 5 × 50 MARDIGRAS runs, resulting in an average restraint width of 1.02 Å.

Five torsion angle restraints for the sugar ring heavy atoms were used to restrain all ribose moieties exhibiting $^3J_{H1'H2'} < 3$ Hz to the C3'-endo conformation employing bounds of 5°. To assure proper base pairing during the refinement, Watson–Crick hydrogen bonding was imposed for the five stem base pairs (Schmitz and James, 1995). Since the MARDIGRAS distance restraints for the terminal base pair reflect dynamics, ideal A-form distances were used instead for these residues, with bounds of 0.2–0.5 Å. For some of the MD simulations, we also employed torsion angle restraints to keep the backbone of the hairpin stem in the A-form realm ($\varepsilon = -151°$; $\zeta = -73°$; $\alpha = -62°$; $\beta = -179°$; $\gamma = +47°$; with bounds of ±10°).

For all structures obtained through conventional rMD calculations, theoretical NOE spectra were calculated via CORMA (Keepers and James, 1984) assuming a correlation time of 3 ns. Most of the theoretical spectra of the preliminary structures exhibited NOE cross peaks that were clearly not observed in the experimental data set. Such contacts were ultimately used to build 'non-NOE distance restraints' where only a lower bound of 4.5 Å was employed.

MD refinement

For all conventional rMD refinement calculations, the modified, energy-minimized coordinates of the T-arm in the $tRNA^{Phe}$ crystal structure (PDB entry 1tra) were used, generated with the modeling package SYBYL (Molecular Modeling Software, Tripos Associates, St. Louis, MO, U.S.A., 1994) by nucleotide swapping and removal of nucleotide modifications. All MD calculations were carried out with AMBER 4.1 (Pearlman et al., 1995) on our local Hewlett-Packard workstation cluster, the CRAY C90 at the Pittsburgh Supercomputer Center, and the CRAY Y-MP at the Frederick Biomedical Supercomputing Center of the National Cancer Institute.

To determine a conventional NMR average structure for the 17-mer hairpin, the above restraints were utilized in 10 ps simulated annealing runs under in vacuo conditions with 'fat' counterions (Seibel et al., 1985). SHAKE (Ryckaert et al., 1977) was used to constrain all covalent bond lengths and angles. The temperature of these simulations was kept at the nominal value by rescaling the velocities whenever deviations exceeded 20°. In our rMD protocol, the temperature was ramped up from 100 to 400 K in 2 ps and decreased after another 5 ps to the target temperature of 300 K. Restraint force constants were ramped from 0.5 to 5 times the final values in 2 ps. Different restraint sets were used during the first and second half of the run: the first set included all restraints mentioned above with the exception of the non-NOE restraints and the distance restraints for nucleotides C56-G57-A58; the second set included all restraints with the exception of the A-form backbone torsion angle restraints. The weights for the second set are ramped from 0.5 to 5 times the final values from 5 to 6 ps and scaled down to the target values after 7 ps. Our protocol also involves a 90% reduced weight for van der Waals and electrostatic terms for the first 7 ps. The full force field including the target restraint force constants was employed for the last 2 ps. Target force constants were either 1, 1.5, or 2 times the following values: non-NOE and NOE distance restraints, 10 kcal/mol $Å^2$; pucker torsion angle restraints, 80 kcal/mol rad^2; Watson–Crick hydrogen bonding restraints, 70 kcal/mol $Å^2$ for distances and 80 kcal/mol rad^2 for angles.

Conventional rMD average structures were derived by averaging the coordinates of the last 2 ps of a specific run and 10 000 steps of conjugate gradient energy minimization with only NOE distance and hydrogen bonding restraints (the force constants were 10 kcal/mol $Å^2$ for all distance restraints and 100 kcal/mol $Å^2$ for angle restraints).

One of the rMD structures was solvated with a 6 Å layer of TIP3P water, which was relaxed during 15 ps of MD at 300 K where only the water molecules were free to move. The system was further prepared for the longer MD simulations by bringing the RNA and solvent to 300 K and the target restraint force constants as described above in 5 ps and subsequent equilibration over another 20 ps. Coordinates and velocities of this run were then subjected to several 100 and 200 ps simulations using conventional and time-averaged restraints. For all of these calculations the temperature was controlled by Berendsen coupling (Berendsen et al., 1984) (300 K) utilizing different coupling constants for RNA and the rest of this system which now includes regular sodium counterions.

Additional parameters for the MDtar simulations include a memory decay constant $\tau = 30$ ps and third-root averaging for distances and force constants of 1.5 times the values described above. To avoid local heating of the system, the restraint weights had to be ramped over 10 ps for the MDtar runs. Note that the pucker torsion angle restraints were not switched to the time-averaged mode in those later runs.

All trajectories were analyzed using the mdanal and carnal modules of AMBER 4.1. Visualization of structures and trajectories was done with MidasPlus (Gallo et al., 1985,1989) and MOIL-VIEW (Simmerling et al., 1995). The latter was also used for the calculation of 2D rms maps and for subsequent cluster analysis. Much of the scrutiny of the restraint violations in structures and ensembles utilized the NOESHOW module in MidasPlus, which provides a convenient graphical representation of the fit.

Results and Discussion

Thermodynamics

UV absorption melting curves (data not shown) of the 17-mer RNA under various buffer conditions and concentrations (1–50 μM) were independent of the RNA concentration, indicating the typical unimolecular transition of a hairpin (Puglisi and Tinoco, 1989). The 17-mer T-loop is relatively stable in low salt ($T_m = 61$ °C, 10 mM KP, pH 6.4); the addition of 50 mM NaCl increases the stability somewhat ($T_m \sim 70$ °C). Small amounts of $MgCl_2$ (5 mM), on the other hand, produce a larger increase in stability ($T_m = 82$ °C). This is noteworthy since one of the Mg^{2+}-binding sites in the $tRNA^{Phe}$ crystal structure involves the T-loop (Pan et al., 1993).

However, it quickly became clear that these high salt buffers are not conducive to our NMR studies since they promote the formation of a second species as evidenced by the doubling of peaks in the NMR spectra at higher concentrations and gel filtration studies in various buffers and different annealing protocols. These observations are the typical, notorious behavior for RNA hairpins that are in equilibrium with duplex species.

Eventually, dilution of the 1.3 mM sample to 0.3 mM utilizing a 10 mm NMR probe practically eliminated the formation of the second species, such that we were able to record high-quality 2D NMR spectra on the single monomer species without interference by signals from the second species. All NMR experiments used for the structure determination were collected at this low concentration (25 °C or 10 °C, 10 mM KP, pH 6.4).

NMR assignments

NMR assignments were performed according to standard methods for nucleic acid structure determination (Wüthrich, 1986; Varani and Tinoco, 1991; Wijmenga et al., 1993).

Exchangeable protons In 1D NMR spectra collected in 9:1 H_2O:D_2O at 10 °C, only four sharp resonances were observed in the 12–15 ppm range (Fig. 1a), which we interpreted as belonging to the four nonterminal base

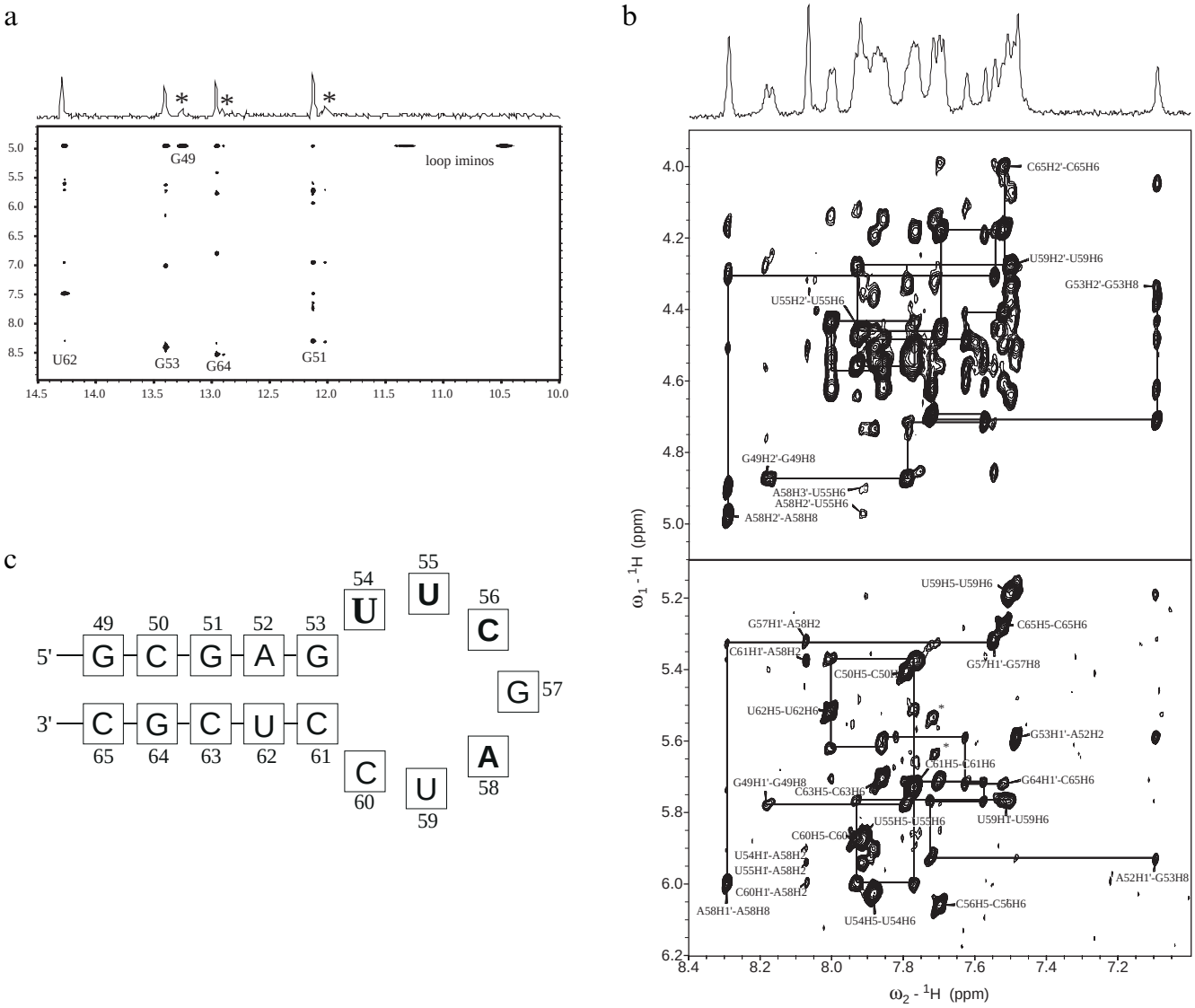

Fig. 1. Portions of the NMR spectra of the 17-mer T-arm fragment at 600 MHz with nucleotide sequence: (a) imino proton-H1', H5, amino proton region of the 400 ms 2D NOE spectrum collected at 10 °C (asterisks in 1D slice indicate cross peaks arising from the minor species); (b) 1D trace and base-ribose and base-H1' proton regions of the 250 ms 2D NOE spectrum collected at 25 °C; (c) RNA sequence used in NMR studies. The residues in bold are those conserved in the $tRNA^{Phe}$ consensus sequence recognized by RUMT. U54 cannot be mutated without a significant loss of binding. NOE H6/8-H1' (lower spectrum) and H6/8-H2' (upper spectrum) walks are traced. The ends of walks are labeled with assignments as well as TOCSY cross peaks (lower spectrum) and NOE connectivities between the A58 H2 and H1' protons (asterisks indicate cross peaks arising from the minor species).

pairs of the stem. A fifth, broader resonance was also observed which was assigned to the terminal GC base pair. The conspicuous absence of broader imino proton resonances in the region of 9–12 ppm, typically observed for loop iminos, indicates that the TΨC-loop imino protons of the 17-mer are not protected from exchange with the solvent. This precludes base pairs in the loop which involve imino protons. Minor resonances belonging to the putative duplex were also observable for certain sample conditions.

2D NOESY spectra collected under the same conditions allowed us to identify hydrogen-bonded base pairs. NOE connectivities were also observed between base-paired G imino protons to C H5 and H1' protons (H1' of the sugar on the same strand, 3' to the G, and H1' of the sugar on the opposite strand, 3' to this base pair), arising from an NOE pathway involving the G amino protons (Heus and Pardi, 1991) (Fig. 1a). This allowed us to confirm peak assignments in the NOESY walks of the nonexchangeable protons (vide infra).

Nonexchangeable protons Spectral characteristics of the sample in H_2O readily suggested that the loop might not be well structured. In such a case, traditional, homonuclear assignment methods must be used with care so as not to rely blindly on through-space connectivity patterns typical for A-form structures. Therefore, we took

TABLE 1
^{1}H NMR RESONANCE ASSIGNMENTS (ppm) FOR THE 17-MER RNA HAIRPIN

	H8	H6	H5	H2	H1'	H2'	H3'	H4'	NH	NH2a	NH2b	$^3J_{H1'H2'}$ (Hz) [a]
G49	8.19	–	–	–	5.78	4.87	4.56	overlap	13.22			8.1
C50	–	7.79	5.41	–	5.71	4.72	4.54	4.33	–	6.79	8.53	<3
G51	7.58	–	–	–	5.77	4.70	4.63	4.55	12.12			<3
A52	7.72	–	–	7.49	5.93	4.71	4.62	overlap	–			<3
G53	7.09	–	–	–	5.59	4.33	4.38	4.43	13.39			<3
U54	–	7.88	6.03	–	5.90	4.51	4.55	overlap	–	–	–	8.6
U55	–	7.92	5.87	–	5.94	4.47	4.74	overlap	–	–	–	7.6
C56	–	7.70	6.06	–	5.71	4.18	4.45	3.99	–			8.3
G57	7.55	–	–	–	5.32	4.31	4.86	4.18	–			8.2
A58	8.29	–	–	8.07	6.00	4.98	4.90	4.51	–			7.8
U59	–	7.50	5.19	–	5.77	4.28	4.64	4.43	–	–	–	8.4
C60	–	7.93	5.88	–	6.00	5.46	4.46	4.30	–			8.6
C61	–	7.77	5.73	–	5.37	4.43	4.51	overlap	–	7.01	8.41	<3
U62	–	8.01	5.52	–	5.62	4.57	4.62	4.47	14.26	–	–	<3
C63	–	7.86	5.71	–	5.59	4.49	4.55	overlap	–	6.95	8.30	<3
G64	7.63	–	–	–	5.72	4.41	4.61	overlap	12.95			<3
C65	–	7.52	5.28	–	5.76	4.00	4.18	overlap	–	7.02	8.34	5.6

[a] Accuracy ±0.2 Hz.

great care to consider all possible alternatives for resonance assignments, calculating rMD structures for the more ambiguous sets. The supplementary material gives a more detailed account of our assignment process.

Adenine H2 protons were readily assigned via non-selective T_1 inversion recovery experiments. The DQF-COSY spectrum allowed us to identify the 10 pyrimidines in our sequence via the H5-H6 resonances of the aromatic rings and the H1'-H2' cross peaks belonging to the C2'-endo and/or dynamic ribose moieties. Ultimately, seven strong peaks in the region of the H1'-ribose NOEs could be assigned to the loop residues. Two weaker peaks in this region arise from the terminal G49 and C65 riboses. A 75 ms TOCSY identified most of the H3' and H4' resonances of the loop residues from their corresponding H1' and H2' cross peaks.

In the region of the 2D NOESY exhibiting the base-H1' connections (Fig. 1b, lower spectrum), continuous H6/8-H1' NOE walks can be traced from residues G49 to G53 on the 5'-side of the stem and from C65 to U59 on the 3'-side of the stem and two residues into the loop. The two purines in the loop also show connectivities to each other. Other cross peaks in this region include those arising between the H2 of A58 in the loop and five H1' protons of other nucleotides, an indication that A58 samples a wide range of conformational space. There are two interresidue H5-H6 cross peaks and one H5-H8 long-range connectivity (G53 H8 to U59 H5). Analogous H6/8-H2' NOE walks can be traced in the base-ribose region of the NOESY spectrum (G49 to G53 and C65 to U59). Within the loop, a H6/8-H2' NOE walk can be traced from A58 to U55. Significantly, U54 shows no connectivities in these walks.

Ambiguities were resolved by rMD refinements for alternate resonance assignments. The structures resulting from the alternative assignments were quite unreasonable. These complications would have been avoided using ^{13}C, ^{15}N-labeled materials. However, since these experiments must be performed at low RNA concentrations, the hardware requirements (8 mm triple resonance probes) are not trivial and were not available at the time.

Proton assignments are listed in Table 1, and a summary of interresidue NOE contacts is depicted in Fig. 2. Even before structure calculations are performed, it is possible to draw some preliminary conclusions based on salient features of these spectra. Breaks in H6/8-H1' and H2' NOESY walks around U54 indicate a disruption of regular stacking; contiguous walks for U59 to C65 suggest that the two residues (U59 and C60) that are looped out in the crystal structure (Holbrook et al., 1978) are actually part of some kind of stacked structure.

There are several indications that we are not dealing with a stable structure but rather a seriously flexible loop: (i) medium-to-strong $^3J_{H1'H2'}$ coupling constants for all loop residues, typical for repuckering or S-type ribose moieties; (ii) no loop imino groups involved in stable base pairs; and (iii) no evidence of the A58:U54 reverse Hoogsteen base pair but instead interresidue contacts between A58 H2 and five different H1' protons (see Fig. 2). At this point, it is readily apparent that the T-loop solution structure is substantially different from that in tRNA.

Structure refinement

The most commonly applied approach in the NMR structure determination of RNA hairpins involves rMD calculations starting from randomized structures where the experimental NOEs are utilized as semiquantitative distance restraints and coupling constants as torsion angle restraints (Jaeger and Tinoco, 1993; Shen et al., 1995). The rMD approach generally yields a single structure that

best satisfies all NMR-derived restraints and the force field parameters simultaneously. In the case of severe conformational dynamics, the rMD average structure might exhibit structural artifacts since not all restraints can be well satisfied in a single, physically meaningful average structure (Pearlman and Kollman, 1991; Pearlman, 1993). Of course, it is also possible that a molecule exists in an equilibrium of several conformers and the rMD-derived average structure looks reasonable and all restraints are satisfied. Here, one must not forget that rMD-derived structures are a compromise between the empirical force field and the experimental restraints. It is clear that with increasing error bounds on the restraints, the potential for these restraints to conflict with each other decreases. However, if one encounters a situation where all restraints cannot be satisfied within the margins of reasonable molecular geometry, as indicated by high conformational energies, significant underlying flexibilities might be the reason. Another consequence of such a situation is the pronounced sensitivity of the rMD structure upon the balance between force field and restraint weights (Schmitz et al., 1996). On the other hand, it is also possible to avoid artifacts by making the restraint bounds for the putative flexible regions deliberately wide, which, however, would leave us with a low precision for these areas. For the T-arm hairpin, where the region of interest is very flexible and the latter approach would not contribute to the elucidation of the structural peculiarities, a more flexible refinement strategy has been employed to obtain structural insights which do not rely on satisfying all restraints in one single structure. To date, two rMD-based methods are available for the refinement of intrinsically dynamic structures: (i) multiple copy refinement, where several MD tracks are simulated in a parallel fashion and the restraint penalty is evaluated for all tracks together (Bonvin and Brünger, 1995; Fennen et al., 1995; Kemmink and Scheek, 1995); and (ii) refinement with time-averaged restraints (MDtar) (Pearlman, 1993; Torda et al., 1993; Nanzer et al., 1995), where the penalty is calculated as a running average over a short period of time, with more recent conformations weighted more strongly. The first method results in a small ensemble satisfying the restraints with as many members as MD tracks generated (Bonvin and Brünger, 1996). This is certainly promising for a situation where a few distinct conformations contribute to the conformational envelope. For the MDtar approach, it is the entire trajectory that satisfies the restraints, leaving us with hundreds of structures. Nevertheless, for DNA duplexes (Schmitz and James, 1993; Schmitz et al., 1993) and RNA/DNA hybrids (González et al., 1995; Schmitz et al., 1996), we have demonstrated that MDtar ensembles provide a more realistic, yet more flexible, picture of the structures. In these cases, local dynamics of relatively small scale, e.g., sugar repuckering and backbone flips, contributed greatly to the dynamics picture. In the present work we have applied the MDtar approach for the T-arm hairpin, where the magnitude of the flexibility is presumably much larger than what we have seen in our earlier work.

A key requisite for obtaining MDtar ensembles that are indeed driven by the internal inconsistencies of the experimental restraints is an rMD-refined average structure with the inherent residual restraint violations (Pearlman and Kollman, 1991). In this respect, it is clear that the NOE distance restraints employed in rMD refinement should be as tight as possible, but accurate as well, making complete relaxation matrix methods for the conversion of NOE intensities into distances the method of choice.

With the program MARDIGRAS, which accounts for all dipole–dipole interactions and provides more accurate distance values than NOE buildup curves or qualitative categorization into few classes of distances, we obtained preliminary distance restraints using either an extended RNA chain or the pertinent coordinates of the $tRNA^{Phe}$ crystal structure as starting models. To minimize the influence of the starting geometry on the resulting distances, a preliminary rMD average structure was determined which was subsequently used for more MARDI-

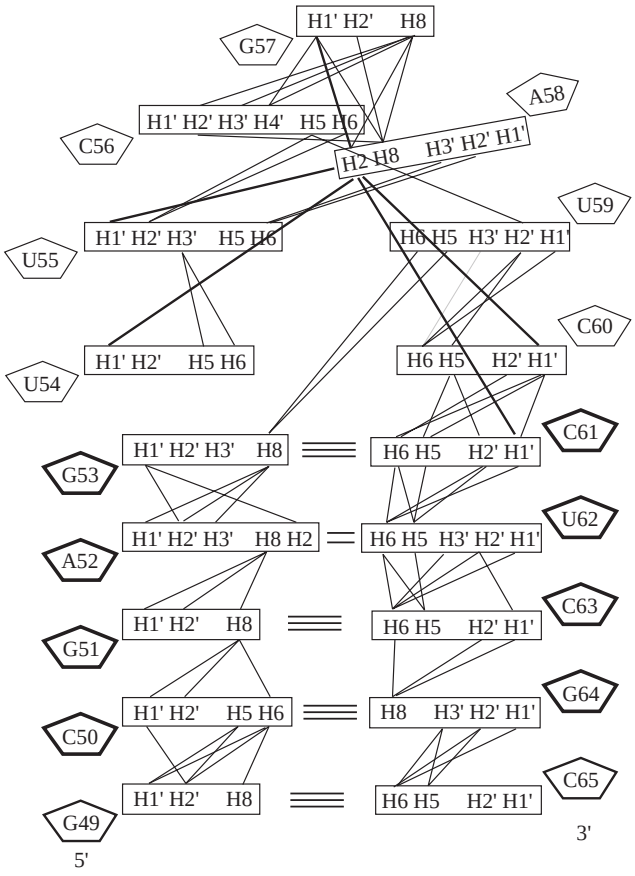

Fig. 2. Interresidue NOEs observed in the 250 ms NOESY spectrum used for structure determination. (The connectivities between the A58 H2 and H1' protons of five residues are highlighted in bold; ribose moieties with C3'-endo conformations are shown in bold lines.)

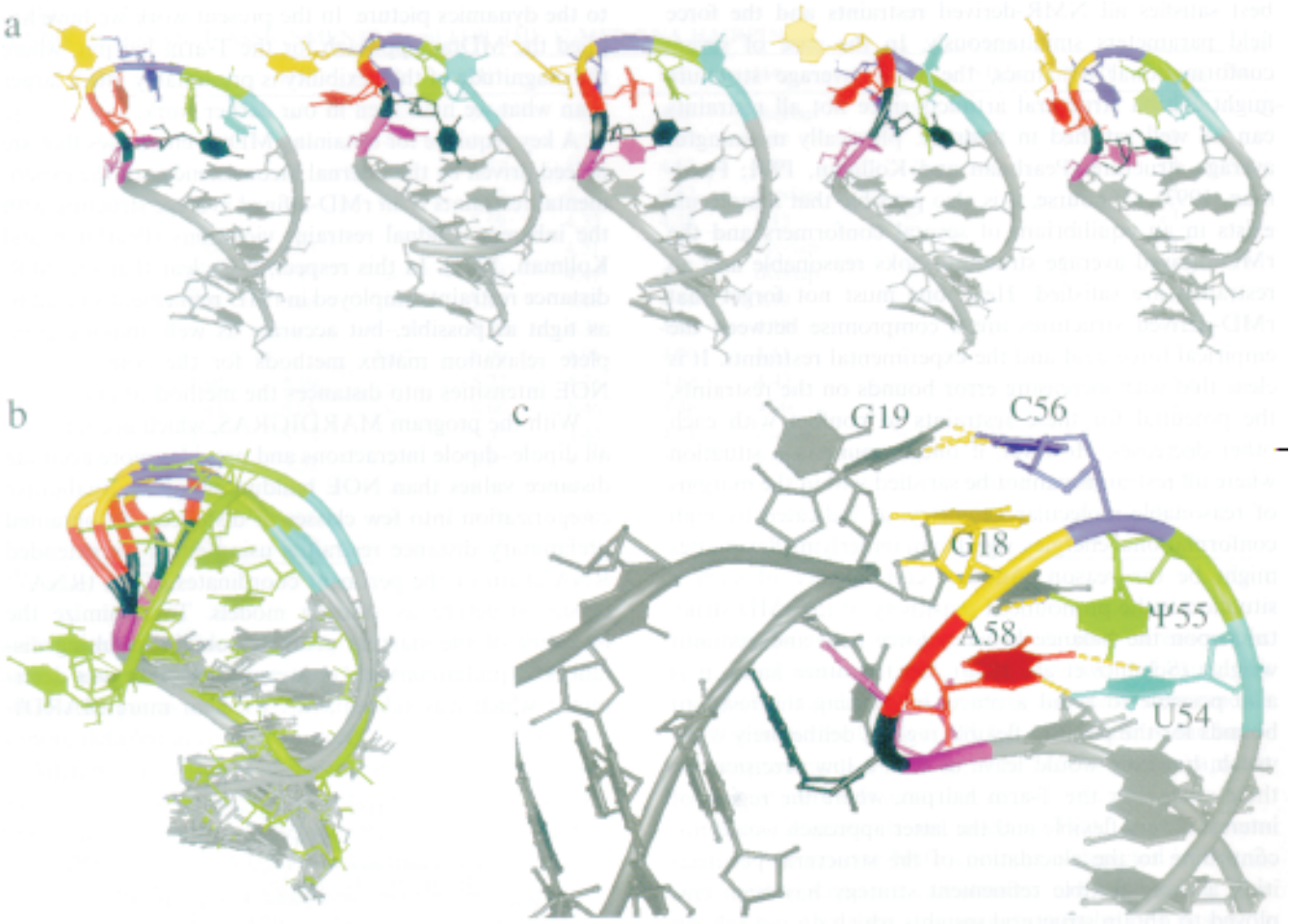

Fig. 3. Average T-arm NMR structures derived by conventional rMD in comparison with the crystal structure of $tRNA^{Phe}$: (a) ribbon diagrams of the five best rMD structures I to V (left to right); stem (gray), U54 (cyan), U55 (green), C56 (purple), G57 (yellow), A58 (red), U59 (black), C60 (magenta); (b) least-squares superposition of stem residues only, for the loop only the backbone is traced; energy-minimized crystal structure coordinates are shown in green; (c) portion of the $tRNA^{Phe}$ crystal structure depicting the tertiary interaction between the D-arm (dark gray) and the T-arm.

GRAS calculations. This procedure helped to resolve ambiguities in peak assignments for crowded regions and it enabled us to compile a list of over 70 'non-NOE' restraints for proton pairs close in the rMD structures, yet not observed in the experimental data. Such indirect distance restraints were extremely helpful in better defining underrestrained regions of the molecule. Since we are very interested in the conformation around U54, a residue which does not exhibit many NOEs, indirect restraints were mostly collected for the loop residues. Practically, non-NOE restraints were defined by assigning the lower bound to 4.5 Å. We note that this procedure should only be used when the two protons involved have normal line widths and the region of interest does not contain other unassigned protons. Besides quantitative distance restraints (total 188; stem 106; loop 82; intraresidue 108; interresidue 80; non-NOE 72), we also utilized torsion angle restraints to enforce C3'-endo puckers for the inner stem residues, which was gleaned from the small J-couplings between H1' and H2'. Furthermore, Watson–Crick hydrogen bonding restraints were utilized for the entire stem.

The second round of MARDIGRAS calculations utilized the new 'randomardi' procedure (Liu et al., 1995) for generating more adequate distance bounds; here, MARDIGRAS runs are repeated many times with slightly different intensity sets for a given starting structure and correlation time. These variations are generated by picking random combinations for the NOE intensities which had initially been assigned individual bounds for each intensity according to a fixed absolute noise error and a relative error defined by the user. The randomardi procedure yields more realistic distance error bounds, especially for weak NOEs. This was demonstrated for a covalently linked drug–DNA complex (Liu et al., 1995), where a number of relatively long fixed distances were reproduced fairly well. A comparison of restraint sets obtained with randomized intensity sets and those used in our earlier work shows a significantly increased average width of the distance restraints; here, the new procedure produces a width of 1.02 Å compared to 0.3–0.4 Å in our earlier DNA work. However, it appears that this difference mostly arises from much larger bounds of the longer

distances (the average widths are 1–3 Å compared to 0.5–1.5 Å without randomization). For the present work, the most common flat well width was still fairly small (0.40–0.45 Å).

Calculation of the rMD average structure

Most commonly, rMD protocols for RNA hairpin structure determination have utilized starting structures with randomized torsion angles. The degree of convergence, typically inferred from the fuzziness of the superimposed structures, reflects how well defined the rMD structures are. In the case of the T-arm, it was clear from the beginning that not all restraints would be satisfied in one structure, especially those for residues A58 and G57 (see Fig. 2) in the loop. Hence, the strategy described above would not yield much useful information. Instead, since we seek to deduce dynamic features of the T-arm, the rMD average structure with the putative residual restraint violations is the best starting point for the ensemble refinement (Pearlman and Kollman, 1991). Therefore, we did not utilize other starting structures for our rMD calculations besides the crystal structure coordinates. Nevertheless, even when the starting structure is much more similar to the target structure, an appropriate rMD protocol still has to be found. To this end, different rMD protocols using AMBER 4.1 were tested by varying the temperature as well as the weight of the restraints and the van der Waals and electrostatic terms of the force field. A protocol and the ensuing structure were deemed satisfactory when the conformational energy was no more than 300 kcal/mol higher than the lowest possible conformational energy that emerges from unrestrained energy minimizations of all structures (including the crystal structure). The most successful protocols utilized partial restraint sets for the first part of the trajectory. In all acceptable structures, residues U59 and C60 are swung back into the loop (see Fig. 3), an adjustment which was often blocked by A58 in protocols using all restraints from the beginning. Therefore, restraints for residues C56 to A58 were blended in only for the second half where the non-NOE restraints were also included. When different random seed numbers were used in satisfactory protocols, similar structures were obtained, yet with larger fluctuations in the loop, especially for residues A58 and G57. However, when comparing the results of different valid protocols using 10–25 kcal/mol Å^2 for the final force constants, the variations become more dramatic; while the average rms deviation for *stems* of 10 acceptable structures remains low (1.1 Å), the *loop* structures are rather different, evident from rms deviations ranging from 2.2 to 3.4 Å (average 2.5 Å).

Our final structures obtained after averaging over the last 2 ps and mild restrained energy minimization (the force constants were 5–10 kcal/mol Å^2 for the distance restraints) still exhibit a fair amount of distorted geometries. Also, the match with the distance target is not very impressive. The overall average distance deviation is 0.18 Å (see Table 2), with the largest violations in the loop. For example, the five largest violations (1–2 Å) are all within the loop. Not unexpectedly, the six restraints involving A58 H2 exhibit an average violation of 1.01 Å. Higher force constants could improve these violations accompanied by a dramatic deterioration of the conformational energy. These results are to be expected for a severely dynamic structure where the average might not be physically meaningful.

Nevertheless, a closer look at the five structures with lowest energy is insightful. Table 2 compares the average values of some figures of merit measuring the deviation from the distance target and, more importantly, the fit with the initial NOE data of the five best rMD structures with those of the crystal structure. For all five conventional rMD structures, depicted in Figs. 3a and b, it is clear that the stem region is well defined and the fit with the NOE data is quite good. R-values compiled for the loop NOEs alone, however, are almost doubled relative to the stem. Note that the fit for interresidue NOEs is also much worse compared to intraresidue data (see Table 2).

TABLE 2
CHARACTERIZATION OF MD STRUCTURES AND ENSEMBLES IN COMPARISON WITH THE T-LOOP CRYSTAL STRUCTURE

Structure(s)	Sixth-root R-values[a]					Average distance deviation (Å)	Energy (kcal)
	Intraresidue	Interresidue	Total	Stem	Loop		
Crystal structure	12.0	26.1	17.1	14.5	32.5	0.841	−1880[b]
Average values for 5 rMD structures ± SD	8.08 (±0.19)	12.62 (±0.44)	9.72 (±0.22)	8.26 (±0.74)	14.76 (±1.40)	0.18 (±0.01)	−1598 (±53)
rMD ensemble[a]	8.2	12.5	9.8	8.1	17.9	0.195	n/a
MDtar ensemble[a]	6.9	10.3	8.1	7.8	9.8	0.142	n/a
Average values for 10 clusters ± SD	11.02 (±0.71)	20.89 (±1.60)	14.57 (±0.76)	13.07 (±1.27)	24.49 (±3.69)	0.62 (±0.09)	−1869 (±12)
Cluster ensemble[c]	9.1	14.8	11.2	11.8	14.6	n/a	n/a

[a] Ensemble R-values reflect 100 snapshots in fast exchange assuming equal probability.
[b] Energies are conformational energies in vacuo after energy minimization; for details see the Materials and Methods section.
[c] Ensemble R-values reflect 10 structures in fast exchange assuming equal probability.

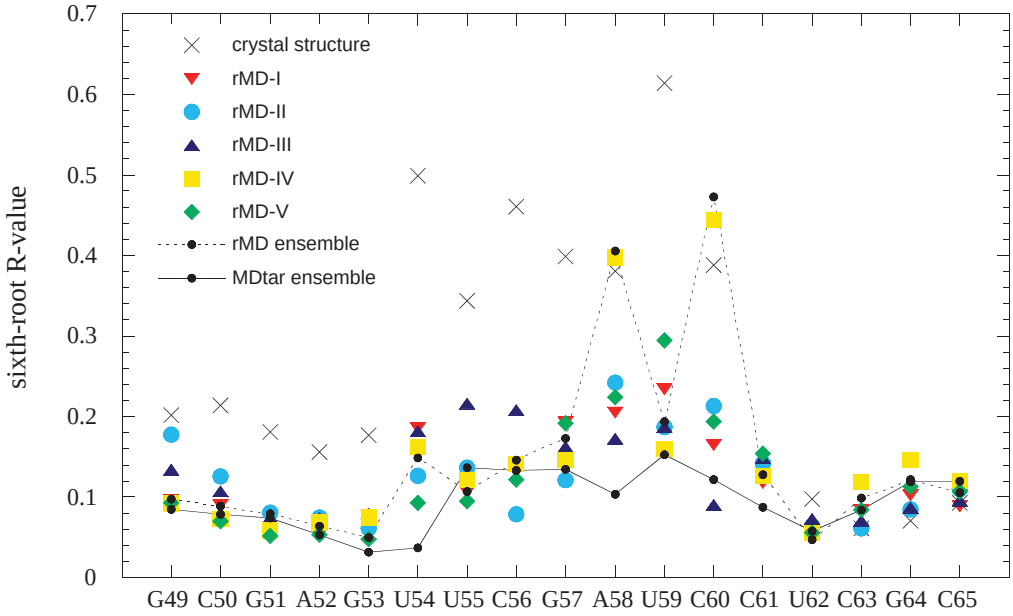

Fig. 4. Interresidue sixth-root R-values per residue for crystal structure and conventional rMD average structure in comparison with the ensemble R-values for rMD and MDtar 100 ps trajectories.

These structures readily reveal that the general fold of the loop is quite different from the crystal structure (Figs. 3b and c); residues C59 and U60 are not bulged and U54 is pushed out with minimal curvature in the backbone. Despite the variability of the base positions, especially residues U55 to A58, the fold of the loop backbone in our ribbon rendering for all structures is quite similar, revealing an interesting turn between A58 and U59.

However, details of the loop structure should not be evaluated from these models, since none of them can satisfy all the unusual interresidue constraints for the loop. R-values for individual residues assume rather large values for all loop residues, A58 and C60 in particular (Fig. 4).

To study whether these structures would exhibit dynamics or interchange between each other in longer conventional rMD trajectories, model rMD-IV (Fig. 3a), which has a relatively low average deviation from the distance restraints but shows the worst interresidue R-values for A58 and C60, was solvated with a 6 Å layer of water and subjected to 100 ps of rMD under conditions equivalent to those of the final period of the initial 10 ps rMD run. In fact, the structure barely changes and the whole loop is locked in a conformation which does not undergo any significant dynamics or interchanges with the other rMD models. When the R-values are calculated for an ensemble of 100 quickly interconverting structures generated every picosecond, the ensuing ensemble R-values are very similar to the average of the R-values obtained for the individual structures (Table 2). (Note that the ensemble R-value is not simply the average of 100 R-values, but is based on the average of 100 relaxation rate matrices, thus reflecting the sixth-root distance weighting presented by the ensemble (Schmitz et al., 1992).) Individual interresidue ensemble R-values in the loop do not improve for the rMD ensemble compared to the individual structures (Fig. 4), which is not surprising because they are all similar. Atomic rms deviations from the starting structure for the rMD trajectory separated for loop and stem (Fig. 5a) reveal that the stem exhibits even larger deviations from the starting structure than the loop, whose deviations barely exceed 0.5 Å. Similar effects were seen in other rMD trajectories as well, where the loop structure is also locked into one conformation. Therefore, we concluded that there is only one way to distribute the restraint violations in the loop. This is probably due to a handful of nonsequential restraints, here related to the conformational dynamics, that tie the loop into a particular fold, similar to long-range restraints in proteins. Any weight change will cause a redistribution of the violations. For comparison, a well-defined DNA duplex ensemble under similar conditions exhibits fluctuations of 0.5–0.7 Å, which suggests that the stem values of over 1 Å are a consequence of the smaller number of restraints per residue for the RNA.

Dynamic structure with time-averaged restraints

Obviously, we need to resort to a method that will create a number of structures satisfying the experimental data as an ensemble. Successful applications of MDtar for flexible refinements have not only been reported for nucleic acid duplexes (Schmitz et al., 1993; González et al., 1995) with small-scale flexibilities, but also for several peptides (Torda et al., 1990; Pearlman, 1993; Nanzer et al., 1995) that were undergoing larger dynamic reorientations, larger than typically investigated for such molecules.

In general, the MDtar approach should produce useful results whenever several conformations associated with local conformational energy minima that reduce the average restraint violations can be sampled efficiently enough during the averaging time interval, τ. The latter typically comprises just a few picoseconds, e.g., 10 ps. If important contributors to the reduction of restraint violations are conformationally far apart, short τ-values might not provide for these states to be reached, despite the fact that the sampling rate of MDtar simulations is artificially enhanced compared to free MD (Pearlman and Kollman, 1991; Schmitz and James, 1993).

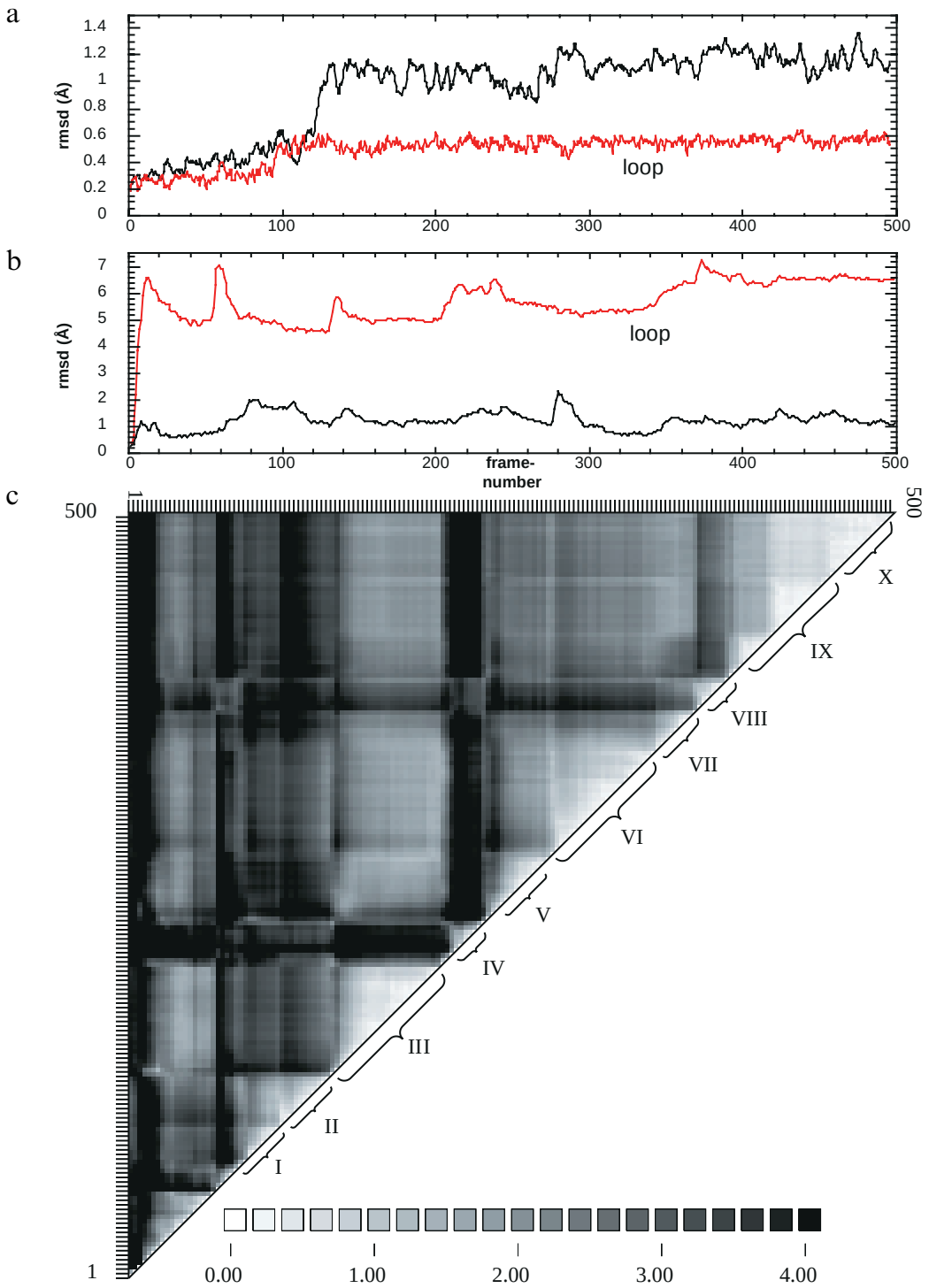

Fig. 5. Characterization of the rMD and MDtar trajectories based on atomic rms deviations: rms deviation of the (a) rMD trajectory and (b) MDtar trajectory from the starting structure (curves for the loop are labeled); (c) 2D rms map of coordinate fluctuations in the 500-snapshot MDtar trajectory (gray scale values indicate the level of rms deviation for the seven loop residues).

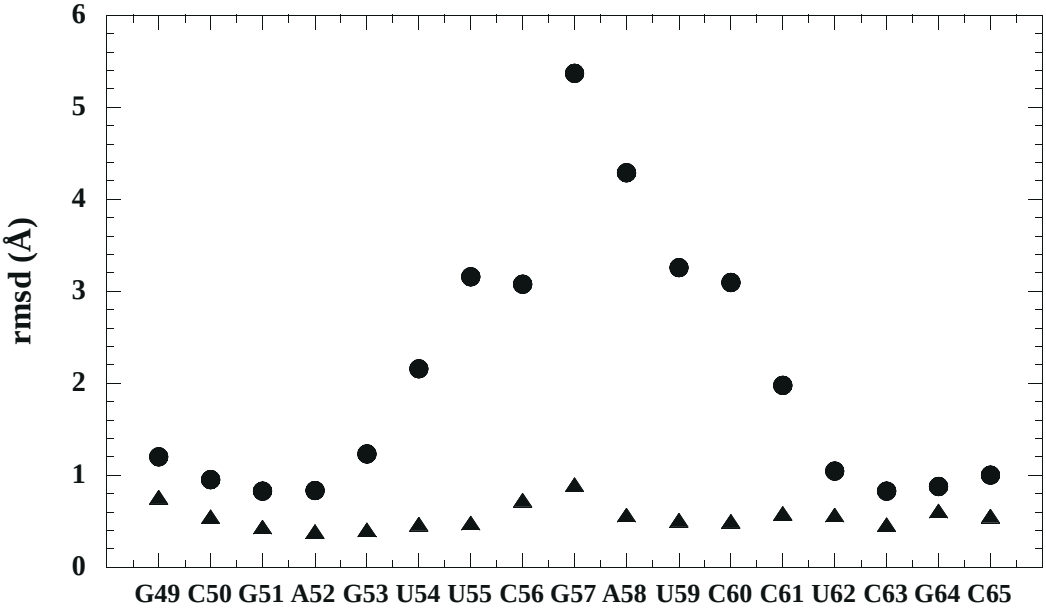

Fig. 6. Average atomic rms fluctuations for individual residues encountered in 100 ps of rMD (▲) and MDtar (●) trajectories.

As a starting point for the MDtar simulations of the T-arm, we chose the same coordinates and velocities of the solvated system that were used for the 100 ps rMD simulation discussed above. As previously pointed out by us (Schmitz et al., 1996) and others (Pearlman and Kollman, 1991; Nanzer et al., 1995), parametrization of MDtar calculations is less than trivial. Here, we generated a number of 100 ps trajectories for different force constants, time constant (τ) values and several ramping protocols for reaching full restraint weights (see the Materials and Methods section). For a fair comparison with the 100 ps rMD simulation, we needed to generate an MDtar trajectory with roughly the same restraint violation energy, the major force driving the calculation, which was achievable with force constants of 15 kcal/mol Å^2. All 100 ps MDtar simulations utilizing 10, 20, 30, and 50 ps for τ exhibit significantly smaller ensemble R-values compared to the 100-snapshot rMD ensemble (Table 2). Using 200 ps simulation time did not significantly improve the ensemble R-values compared to shorter simulation times. Ensembles generated with the smaller τ-values exhibit more distorted geometries, e.g., warped bases, similar to those from conventional rMD with high force constants. For 100 ps simulations, the trajectory for $\tau = 50$ ps shows less conformational reorientations than that for $\tau = 30$ ps, suggesting that the running average of the violations becomes too insensitive to the conformational fluctuations for the given restraint force and overall simulation time. The trajectory obtained for $\tau = 30$ ps, 300 K, and force constants of 15 kcal/mol Å^2 was therefore considered the best data set and will be discussed here in more detail. While the MDtar approach does not produce a unique solution to our problem, it nevertheless will show us some broad conformational themes, revealing possibilities as to what array of conformations together will match the NMR data better than a simple average structure.

Our MDtar trajectory of the T-loop exhibits a significant improvement in all figures of merit, with the largest effect in the loop (Table 2). The average distance deviation decreased to 0.14. The average distance deviation for the restraints involving A58 H2 improved from over 1 Å to 0.26 Å. But even for our MDtar trajectory some larger violations remained; three violations were found over 1 Å (1.05–1.52 Å). A more direct impression of the improvement can be gleaned from R-values, which compare theoretical with experimental NMR data. The dramatic improvement of the individual interresidue R-values for the loop is evident from Fig. 4. Obviously, the MDtar ensemble allowed some conformational averaging to accommodate the dynamic requirements of residues A58 and C60. To understand the structural excursions of the loop in the MDtar trajectory, we analyzed the latter on the basis of atomic rms deviations between all 500 snapshots in conjunction with an inspection of the animated coordinates. The rms deviations for stem and loop residues from the starting geometry (Fig. 5b) reveal a trend opposite to the rms deviations of the rMD trajectory shown (Fig. 5a). Here, both stem and loop make a jump within the first few picoseconds. On top of the thermal fluctuations producing the noise-like modulation of the rmsd trace, the MDtar simulation also exhibits several sudden excursions of up to 2 Å. Note that stem and loop share some of these jolts besides having their own individual ones. The animated MDtar trajectory reveals readily that the conformational jumps are fast transitions between quieter conformational plateaus. This notion becomes clearer when inspecting the associated 2D rms map (Fig. 5c) (Simmerling et al., 1995) depicting rms deviations for

just the loop residues. Dark areas represent periods of larger deviations associated with conformational jumps. These jumps comprise conformations that are very different from the main parts of the MDtar trajectory and reflect high-energy structures with noticeable violations of the bond angle and dihedral terms of the force field. Obviously, when the distance restraint penalty of the system builds up, the necessary conformational jumps will increase the penalty even more before the system can relax into a new conformation.

To captivate the essence of the conformational journey of the T-arm in the MDtar trajectory, we attempted to define clusters of similar conformations assuming that the higher energy transition structures do not contribute significantly to the total conformational envelope. The clustering was done with the program MOIL-VIEW (Simmerling et al., 1995). Since most clusters arose from contiguous time stretches in the trajectories, we chose to represent each cluster by just one contiguous string of coordinates; each time-cluster was subsequently averaged and subjected to 1000 steps of free energy minimization. The conformational energies of all 10 cluster structures are very similar to that of the best free MD structures (compare the average conformational energies in Table 2), indicating that strain is no longer inflicted by the distance restraints.

Before turning to structural details, it is important to realize that each loop residue experiences its own level of flexibility, evident from the average rms fluctuations during the trajectory depicted in Fig. 6. Note that the individual flexibilities reflect the degree to which restraints were inconsistent and should not be mistaken as indicators for real degrees of flexibilities, such as B-factors in crystallography. Since it is not straightforward to grasp the dynamics from the depiction of the 10 cluster structures alone, a description of the salient features of the actual trajectory might be in order. Residues G57 and A58 are clearly the most flexible, a result expected only for A58. Most of the large conformational changes of A58 clearly force other residues to adjust. An interesting

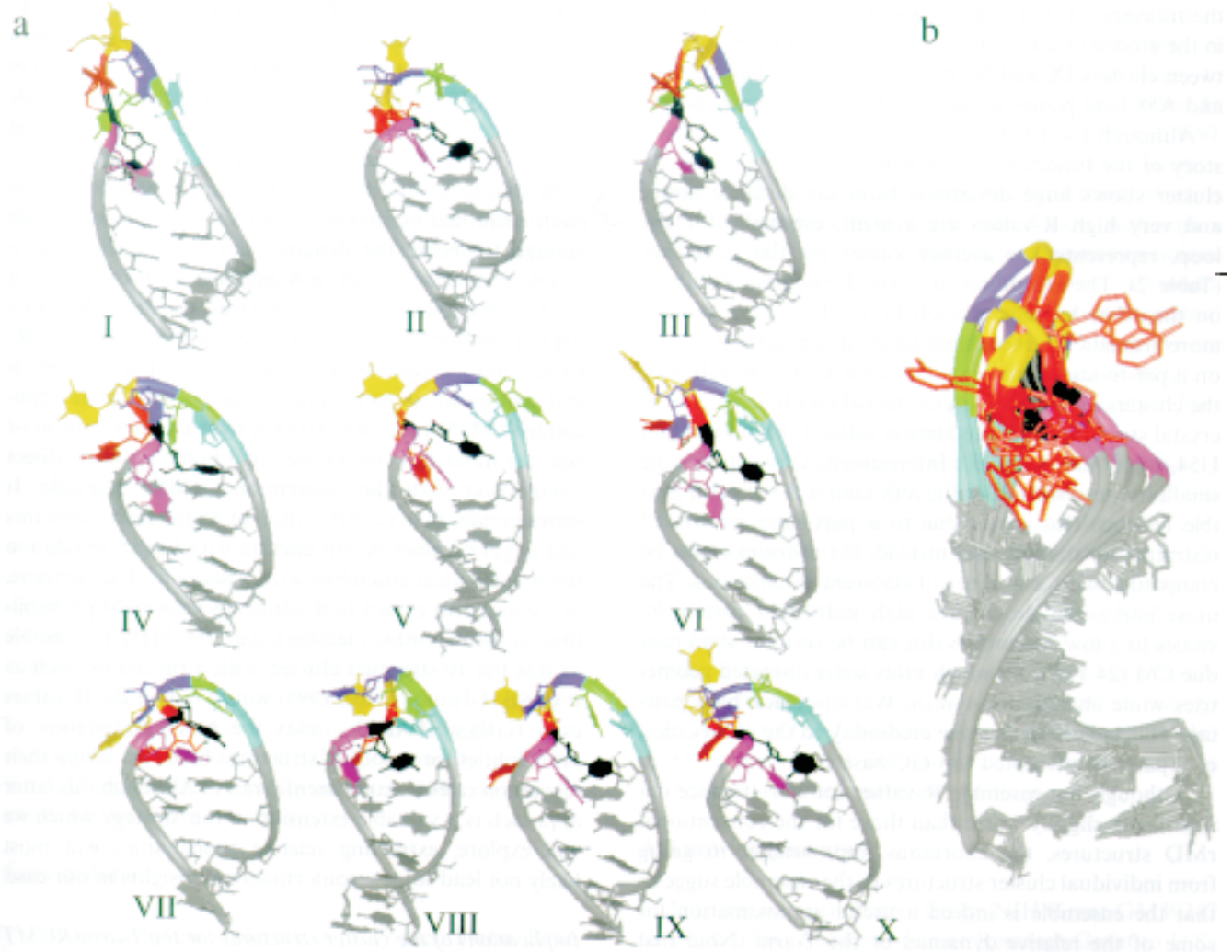

Fig. 7. Ten MDtar time-cluster structures based on the 2D rms map in Fig. 5: (a) individual cluster structures; (b) superposition of all 10 cluster structures based on stem coordinates; only A58 is shown for the loop. (For the color coding of the loop residues see Fig. 3.)

coupling is seen for the two pyrimidines following A58; C60 can swing from one groove into the other, a move that is responsible for the jolt at the beginning of the trajectory inducing a temporary disruption of the top base pair. Remnants of this perturbation are visible in clusters I and II (Fig. 7). The GC base pair later forms again after C60 stacks on C61 to slide off into the major groove (clusters III and IV). A vertical stretch of the entire loop led by G57 takes place, followed by a compression in which A58 dives into the minor groove (cluster V). This seems to propel C61 out of its base pair, creating room for residues A58 to C60 to adjust. During large parts of the periods associated with clusters IV and V, a stack forms between U59, C60, and C61 which is most visible in cluster VII. For some time, U54 assumes the top position of that stack with an almost perfect A-form helical twist (cluster VII). This arrangement breaks and re-forms subtly; the differences between clusters VI and VII mostly arise from roaming purines. For a short period, U59 even stacks on G53 while C60 is dangling into the major groove (cluster VIII). For the remainder of the trajectory, U59 is often stacking on G53 while C60 is in the groove or sometimes on top of C61. Differences between clusters IX and X are associated with flips of G57 and A58 base planes invoking backbone rearrangements.

Although the 10 cluster structures cannot tell the full story of the trajectory, it is remarkable how every single cluster shows huge deviations from the distance target and very high R-values are evident, especially for the loop, represented as average values for the structures (Table 2). The relaxation rate based ensemble R-values, on the other hand, are much lower. This effect is even more dramatic when considering the interresidue R-values on a per-residue basis (data not shown). The R-values of the clusters amount to almost the values obtained for the crystal structure, with the highest values for loop residues U54, U55, A58, and C60. Interestingly, G57 exhibits the smallest interresidue R-value although it is the most flexible residue. This is not due to a particular paucity of restraints for that residue; instead, the restraints must be compatible with a number of different orientations. The most impressive drop from high individual cluster R-values to a low ensemble value can be seen for stem residue C61 (24.1 to 9.8), which visits some distorted geometries while abandoning typical Watson–Crick base pairing. This indeed lends some credibility to the observation of a partially disrupted top GC base pair.

Although the ensemble R-values for our 10-piece ensemble are slightly worse than those for the conventional rMD structures, the enormous improvement in going from individual cluster structures to the ensemble suggests that the ensemble is indeed a useful approximation for some of the relative dynamics of the T-arm. Note that compared with the original MDtar trajectory, not only are all transition states excluded, but also the relative probabilities for the different conformational states are changed, which should account for most of the R-value increase.

A structural comparison of the 10 cluster structures (Fig. 7) with the initial, strained rMD structures (Fig. 3) also yields some interesting similarities despite the large rms deviations (the average rmsd for all 15 structures is 4.31 Å): (i) the turn between A58 and U59 is apparent in all structures and even more pronounced in some of the cluster structures, e.g., clusters V and X; and (ii) U54 is swung out of the loop leaving the backbone with no helical twist, creating a gap at its 3'-side in almost all structures which could explain how RUMT gains access to the methylation site. Figure 7b depicts the superposition of all 10 cluster structures. The ribbon rendering reveals some interesting differences in the loop folding for the T-arm. It appears that the turn between A58 and U59 is more pronounced for those structures where C60 is further out in the major groove, which also leads to a vertical compression of the loop. Two clusters stand out with their elongated loops that require the two 3'-pyrimidines to gravitate more toward the minor groove. With respect to the 5'-half of the loop, two themes emerge, where the backbone either continues with some helical twist or straightens out more to let the loop be more elongated. For the loop residues in Fig. 7b, only A58 is shown. It is remarkable that this residue is found in three very different places in order to roughly satisfy the restraints (vide supra).

With respect to our methodological concerns that the conformational excursions in MDtar would not be wide enough to model the dynamic situation of the loop, a simple conclusion is not at hand. Although the range of the loop flexibilities is relatively large, the fit of the entire MDtar ensemble with the structural information (i.e., R-values and average deviation from the distance target) is still slightly worse compared to our earlier MDtar applications to DNA and RNA/DNA hybrids. Here, one must bear in mind that we chose conditions to allow direct comparison with the conventional rMD ensemble. It seems certain that a combination of higher force constants and larger τ-values in conjunction with longer simulation times could yield ensembles with a better fit. Furthermore, one could also expect that adjusting individual probabilities to the ensemble members, e.g., the MDtar ensemble or just the 10 structure cluster, with a procedure such as PARSE (Ulyanov et al., 1995) would reduce the R-values even further. (PARSE yields the best combination of probabilities for a pool of structures based on fitting their relaxation rates to experimental rates.) Although this latter approach is a valuable extension in our strategy which we will explore, extracting relative probabilities will most likely not lead to new conformational insights in our case.

Implications of the cluster structures for the T-arm/RUMT recognition

The fact that RUMT can methylate almost any hair-

pin with a stable stem and a seven-membered loop with a 5'-leading uracil and a few other sequence requirements (Kealey et al., 1994; Gu et al., 1996) suggests that RUMT does not recognize a specific, stable structure of those loops. The recognition seems to involve a consensus conformation in which the hairpin presents U54 and some tertiary structure elements of the loop for specificity. The flexibility of the loop as seen in the present study suggests that those tertiary structure elements could be as simple as certain spacings of phosphate groups or specific orientations of some base planes. Here we have to ask why the enzyme does not perform more U-methylations in vivo since other seven-membered RNA loops with a 5'U can be found in vivo. The structural flexibility obtained in our MDtar trajectories supports the idea that RUMT acts upon a dynamic T-arm and that the specific recognition of its tRNA substrate hinges on the tertiary structure embodiment of the T-arm in tRNA. Of course, our study cannot claim that the way in which RUMT gains access to the originally inaccessible U54 methylation site in tRNA is the same as in the T-loop alone. The very accessibility of U54 in the free T-arm, however, seems to indicate that a reactive situation could arise easily if the tertiary interactions between the T-arm and the D-loop in tRNA were disrupted. Indeed, the increase in methylation kinetics for D-arm tRNA mutants (G19C; G18A) where one of the three base pairs between the T- and D-loops could not form, demonstrated that part of the RUMT/tRNA interaction involves breaking those base pairs (Kealey et al., 1994). Also, the flexible loop structure does not require adjustments for the stem. Therefore, it seems logical that as soon as the stabilizing interactions between the D- and T-loops are disrupted by RUMT, the reverse Hoogsteen base pair A58:U54 is not strong enough to keep the T-loop together. This notion is also supported by MD studies of tRNA fragments (Louise-May et al., 1996), demonstrating that the A58:U54 pair is much less stable when the T-arm cannot interact with the D-loop.

For the final enzymatic step, after the AdoMet-dependent methylation has occurred, the protein/RNA interaction is probably weakened, causing dissociation, and the tertiary tRNA base pairs can be reestablished as soon as RUMT leaves. It is unlikely that modification of U54 alters the loop structure such that the new structure could be the driving force for the disintegration of the complex; U54 seems too exposed and our structures do not suggest any significant intraloop interaction of a methylated U54. We will enzymatically methylate our NMR sample to check on the above rationale.

Conclusions

Insights into the structure of a 17-mer T-arm RNA fragment which serves as a substrate for U54 methylation by RUMT have been gained through homonuclear NMR and different rMD approaches. The vast majority of the nonexchangeable protons could be assigned rendering a large number of interproton NOEs sufficient for a high-resolution structure determination. The NMR data indicated that the loop portion of the T-arm is flexible and no intraloop base pairs are formed. Interproton NOEs were converted into accurate distance restraints via complete relaxation matrix methods. Restrained MD calculations yielded an NMR average structure that could not satisfy all restraints well, especially in the loop region. However, when the average structure is subjected to MD with time-averaged restraints yielding a loose ensemble of structures, the original NMR data are much better satisfied, especially for the loop. To distill the structural essence of the MDtar ensemble, 10 time-clusters were generated on the basis of atomic rms deviations of the loop and converted into low-energy average structures. Taking these 10 structures as an approximation of the dynamic structure of the T-arm, some structural features emerge that aid in understanding how the tRNA methylation by RUMT might occur. In the dynamic solution structure, both of the bulged pyrimidines involved in tertiary base pairs with the D-arm, U59 and C60, are swung back into the loop. Interestingly, they seem to spend time both on major and minor groove sides. In several cluster structures, stacking between U59 and C60, U59 and G53, or C60 and C61 is observed. Most interestingly, U54 is found in a fairly exposed situation, which is often accompanied by a gap on the 3'-side. Although a particular MDtar ensemble cannot be considered a unique description of the structural details of a molecular system, we can be confident of the aforementioned structural traits as consistent dynamic features of the T-arm. We are currently exploring a systematic comparison of different MDtar ensembles in conjunction with probability refinement using the PARSE program.

The dynamic features of the T-arm solution structure support the idea that RUMT is a rather promiscuous enzyme with minimal sequence and structure requirements for the methylation itself, although the enzymatic mechanism involves a great deal of precision. The specificity for tRNA is most likely due to the ability of RUMT to recognize tertiary structure elements of its substrate. Breaking some of the intrinsic tertiary hydrogen bonds between the D- and T-arm would be a natural way to give access to the U54-base. Our structural results also intimate how other T-arm modification enzymes, such as U55-pseudouridine synthetase, perform.

Acknowledgements

This research was supported by NIH Grant GM39247 to T.L.J., a UC Academic Senate Research Grant to U.S., and NRSA postdoctoral fellowship GM18337 from the National Institutes of Health to L.J.Y. We gratefully

acknowledge the UCSF Computer Graphics laboratory supported by NIH Grant RR01081 and the use of the Pittsburgh Supercomputing Center as well as the Frederick Biomedical Supercomputing Center for project grants to T.L.J. (CHE880090P) and U.S. (162002), respectively.

References

Amano, M. and Kawakami, M. (1992) *Eur. J. Biochem.*, **210**, 671–681.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–372.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Di Nola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Bonvin, A.M.J.J. and Brünger, A.T. (1995) *J. Mol. Biol.*, **250**, 80–93.

Bonvin, A.M.J.J. and Brünger, A.T. (1996) *J. Biomol. NMR*, **7**, 72–76.

Fennen, J., Torda, A.E. and van Gunsteren, W.F. (1995) *J. Biomol. NMR*, **6**, 163–170.

Gallo, K., Huang, C., Ferrin, T.F. and Langridge, R. (1985,1989) Molecular Interactive Display and Simulation (MidasPlus), University of California, San Francisco, CA, U.S.A.

González, C., Stec, W., Reynolds, M. and James, T.L. (1995) *Biochemistry*, **34**, 4969–4982.

Gu, X., Ivanetich, K.M. and Santi, D.V. (1996) *Biochemistry*, **35**, 11652–11659.

Hall, K.B., Sampson, J.R., Uhlenbeck, O.C. and Redfield, A.G. (1989) *Biochemistry*, **28**, 5791–5801.

Heerschap, A., Haasnoot, C.A.G. and Hilbers, C.W. (1983a) *Nucleic Acids Res.*, **11**, 4483–4499.

Heerschap, A., Haasnoot, C.A.G. and Hilbers, C.W. (1983b) *Nucleic Acids Res.*, **11**, 4501–4520.

Heus, H.A. and Pardi, A. (1991) *J. Am. Chem. Soc.*, **113**, 4360–4361.

Holbrook, S.R., Sussman, J.L., Warrant, R.W. and Kim, S.-H. (1978) *J. Mol. Biol.*, **123**, 631–660.

Hyde, E.I. and Reid, B.R. (1985) *Biochemistry*, **24**, 4307–4314.

Jaeger, J.A. and Tinoco Jr., I. (1993) *Biochemistry*, **32**, 12522–12530.

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

Kealey, J.T., Gu, X. and Santi, D.V. (1994) *Biochimie*, **76**, 1133–1142.

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

Kemmink, J. and Scheek, R.M. (1995) *J. Biomol. NMR*, **6**, 33–40.

Kneller, D.G., Sparky, NMR display and processing program, University of California, San Francisco, CA, U.S.A., © 1992.

Liu, H., Borgias, B., Kumar, A. and James, T.L. (1990,1994) MARDIGRAS, University of California, San Francisco, CA, U.S.A.

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

Louise-May, S., Auffinger, P. and Westhof, E. (1996) In *Biological Structure and Dynamics* (Eds., Sarma, R.H. and Sarma, M.H.), Vol. 2, Adenine Press, Schenectady, NY, U.S.A., pp. 73–91.

Milligan, J.F., Groebe, D.R., Wilherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.

Nanzer, A.P., van Gunsteren, W.F. and Torda, A.E. (1995) *J. Biomol. NMR*, **6**, 313–320.

Pan, T., Long, D.M. and Uhlenbeck, O. (1993) In *The RNA World* (Eds., Gesteland, R.F. and Atkins, J.F.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, U.S.A., pp. 271–302.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.*, **220**, 429–457.

Pearlman, D.A. (1994) *J. Biomol. NMR*, **4**, 1–16.

Pearlman, D.A., Case, D.A., Caldwell, J.C., Ross, W.S., Cheatham III, T.E., Ferguson, D.N., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1995) AMBER v. 4.1, University of San Francisco, San Francisco, CA, U.S.A.

Puglisi, J.D. and Tinoco, I.J. (1989) *Methods Enzymol.*, **180**, 304–325.

Quigley, G.J. and Rich, A. (1976) *Science*, **194**, 796–806.

Romby, P., Acabon, P., Westhof, E., Ehresmann, C., Ebel, J.-P., Ehresmann, B. and Giege, R. (1987) *J. Biomol. Struct. Dyn.*, **5**, 669–687.

Roy, S. and Redfield, A.G. (1983) *Biochemistry*, **22**, 1386–1390.

Ryckaert, J.P., Cicotti, G. and Berendsen, H.J.C. (1977) *J. Comput. Phys.*, **23**, 327–341.

Schmitz, U., Kumar, A. and James, T.L. (1992) *J. Am. Chem. Soc.*, **114**, 10564–10566.

Schmitz, U. and James, T.L. (1993) In *Structural Biology: The State of the Art* (Eds., Sarma, R.H. and Sarma, M.H.), Vol. 2, Adenine Press, Schenectady, NY, U.S.A., pp. 251–272.

Schmitz, U., Ulyanov, N.B., Kumar, A. and James, T.L. (1993) *J. Mol. Biol.*, **234**, 373–389.

Schmitz, U. and James, T.L. (1995) *Methods Enzymol.*, **261**, 1–43.

Schmitz, U., González, C., Liu, H., Ulyanov, N.B., Blocker, F. and James, T.L. (1996) In *Biological Structure and Dynamics* (Eds., Sarma, R.H. and Sarma, M.H.), Vol. 2, Adenine Press, Schenectady, NY, U.S.A., pp. 165–189.

Seibel, G.L., Singh, U.C. and Kollman, P.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6537–6540.

Shen, L.X., Cai, Z. and Tinoco, I. (1995) *FASEB J.*, **9**, 1023–1033.

Simmerling, C., Elber, R. and Zhang, J. (1995) In *Modeling of Biomolecular Structures and Mechanisms* (Ed., Pullman, A.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 241–265.

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

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

Torda, A.E., Scheek, R.M. and van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Ulyanov, N.B., Schmitz, U., Kumar, A. and James, T.L. (1995) *Biophys. J.*, **68**, 13–24.

Varani, G. and Tinoco, I.J. (1991) *Q. Rev. Biophys.*, **24**, 479–532.

Wijmenga, S.S., Mooren, M.M.W. and Hilbers, C.W. (1993) In *NMR of Macromolecules* (Ed., Roberts, G.C.K.), Vol. 134, Oxford University Press, New York, NY, U.S.A., pp. 217–288.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY, U.S.A.