A Distinctive RNA Fold: The Solution Structure of an Analogue of the Yeast tRNA^{Phe} TΨC Domain[†,‡]

Karl M. Koshlap,[§] Richard Guenther,[§] Elzbieta Sochacka,[||] Andrzej Malkiewicz,[||] and Paul F. Agris*[,§]

Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695-7622, and Institute of Organic Chemistry, Technical University, Lodz 90-924, Poland

Received January 19, 1999; Revised Manuscript Received May 11, 1999

ABSTRACT: The structure of an analogue of the yeast tRNA^{Phe} TΨC stem−loop has been determined by NMR spectroscopy and restrained molecular dynamics. The molecule contained the highly conserved modification ribothymidine at its naturally occurring position. The ribothymidine-modified TΨC stem−loop is the product of the m^{5}U$_{54}$-tRNA methyltransferase, but is not a substrate for the m^{1}A$_{58}$-tRNA methyltransferase. Site-specific substitutions and ^{15}N labels were used to confirm the assignment of NOESY cross-peaks critical in defining the global fold of the molecule. The structure is unusual in that the loop folds far over into the major groove of the curved stem. This conformation is stabilized by both stacking interactions and hydrogen bond formation. Furthermore, this conformation appears to be unique among RNA hairpins of similar size. There is, however, a considerable resemblance to the analogous domain in the crystal structure of the full-length yeast tRNA^{Phe}. We believe, therefore, that the structure we have determined may represent an intermediate in the folding pathway during the maturation of tRNA.

Transfer RNA (tRNA) has been studied extensively, and a large amount of information has been obtained about its structure and function. A tRNA molecule was first sequenced in 1965 by Holley et al., who also noted that the nucleotide sequence could be written in the now-familiar cloverleaf representation (*1*). The cloverleaf is composed of the D,[1] anticodon, and TΨC stem−loop domains; the duplex amino acyl acceptor stem; and a variable region (Figure 1). The folding of these domains into the three-dimensional structure was described in 1974 when the first tRNA crystal structures were solved (*2, 3*). These, and the few other tRNA structures that have followed, share a general L-shaped motif. One arm of the L-shaped molecule is an extended helix formed from the coaxial stacking of the TΨC and acceptor stems, while the other is formed from the stacking of the D stem on the anticodon stem. The two arms are held together by a complex network of conserved tertiary interactions between the TΨC stem−loop, the D stem−loop, and the variable region.

Since tRNA was the only large RNA molecule whose structure was known for many years, it served as a model for other RNAs as well. Many of the structural features found in tRNA—such as noncanonical base pairs, base triples, and base zippers—have since been observed in other RNA molecules. However, a striking difference between tRNA and other classes of RNA is the number and variety of modified nucleosides present in tRNA. To date, approximately 100 different modifications have been identified, all of which are introduced posttranscriptionally (*4*). They range from relatively simple methyl additions to quite elaborate modifications, such as is seen in the tricyclic guanine derivative wybutosine. These modifications vary not only in kind and degree of complexity, but also in the frequency of their occurrence. Some modified nucleosides are specific to particular tRNA species or classes of organisms, while others—notably ribothymidine at position 54 (T$_{54}$) and pseudouridine at position 55 (Ψ$_{55}$)—are present in most tRNAs of most organisms (*5, 6*). Indeed, they are so conserved that they lend their name to the domain in which they occur, the TΨC stem−loop.

T$_{54}$ and Ψ$_{55}$ not only are extremely common modifications but also are among the earliest to appear during the complex maturation process of tRNA (*6*). T$_{54}$ and Ψ$_{55}$ have been found, for example, in primary transcripts that still had an intron and unprocessed 3′ and 5′ extensions (*7*). This implies that m^{5}U$_{54}$-tRNA methyltransferase and Ψ$_{55}$-tRNA synthase—which are responsible for the methylation of U$_{54}$ to produce T$_{54}$, and the conversion of U$_{55}$ to Ψ$_{55}$, respectively—recognize features that are common to many tRNAs, and are not necessarily dependent on the overall folded structure of the

[†] This research was supported by USPHS Grant GM-23027 (P.F.A.), Cambridge Isotope Laboratories, Inc. (P.F.A.), and the Polish Committee for Scientific Research Grant PBO506/P3/93/05 (A.M.).

[‡] Coordinates and conformational restraints have been deposited in the Brookhaven Protein Data Bank under the accession code 1kos.

* To whom correspondence should be addressed at the Department of Biochemistry, 128 Polk Hall, Box 7622, North Carolina State University, Raleigh, NC 27695-7622. Phone: 919-515-6188. FAX: 919-515-2047. E-mail: Agris@bchserver.bch.ncsu.edu.

[§] North Carolina State University.

[||] Technical University.

[1] Abbreviations: T, ribothymidine; Ψ, pseudouridine; D, dihydrouridine; m^{5}C, 5-methylcytidine; Gm, 2′-*O*-methylguanosine; TΨC, TΨC stem and loop domain; RUMT, m^{5}U$_{54}$-tRNA methyltransferase; RAMT, m^{1}A$_{58}$-tRNA methyltransferase; NMR, nuclear magnetic resonance; HPLC, high-performance liquid chromatography; DEAE, diethylaminoethyl; EDTA, ethylenediaminetetraacetic acid; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; τ_m, mixing time; ROESY, rotating-frame Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; HETCOR, heteronuclear correlation; HSQC, heteronuclear single quantum coherence; DQ, double quantum; DQFC, double quantum filtered correlation spectroscopy; rmsd, root-mean-square deviation.

10.1021/bi990118w CCC: $18.00
Published on Web 06/16/1999

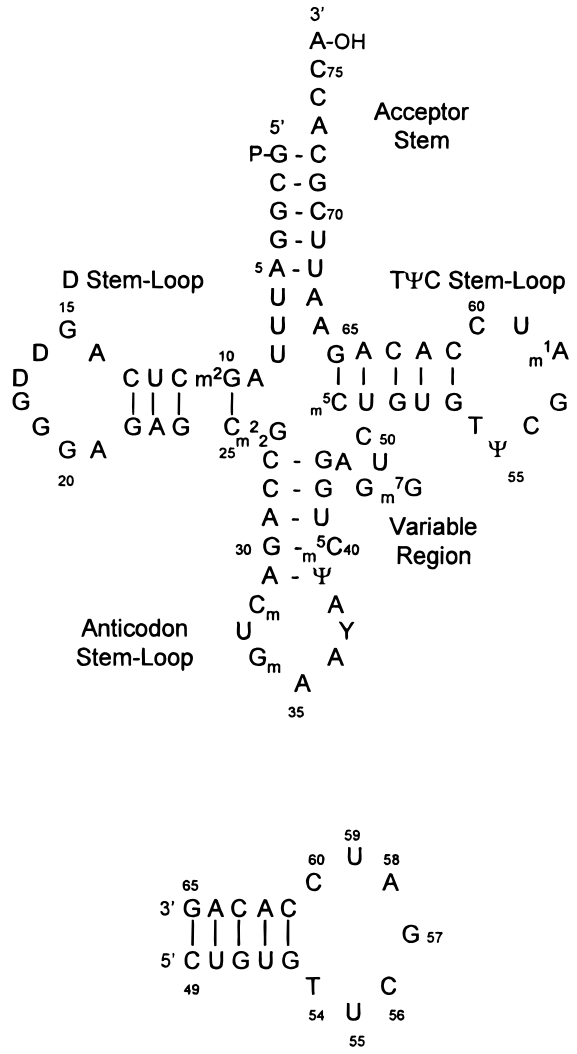

FIGURE 1: Primary sequence and secondary structure of yeast $tRNA^{Phe}$ and its TΨC stem and loop. (Top) Nucleoside sequence of native yeast $tRNA^{Phe}$ including all modified nucleosides shown in the cloverleaf representation of secondary structure. The modified nucleosides are (in order from the 5′ to 3′ termini): m^2G, *N*2-methylguanosine; D, dihydrouridine; m^2_2G, *N*2,2-dimethylguanosine; Cm, 2′*O*-methylcytidine; Gm, 2′*O*-methylguanosine; Y, wybutosine; Ψ, pseudouridine; m^5C, 5-methylcytidine; m^7G, *N*7-methylguanosine; T or m^5U, ribothymidine or 5-methyluridine; m^1A, *N*1-methyladenosine. (Bottom) Sequence and secondary structure of the yeast $tRNA^{Phe}$ TΨC stem and loop domain with T_{54}, TSL-T_{54}. The structure of this singly modified domain was determined in this study.

mature molecules. In fact, a number of recent studies have demonstrated that an RNA hairpin corresponding in sequence to that of the unmodified TΨC stem and loop domain (TSL) is a substrate for the m^5U_{54}-tRNA methyltransferase (RUMT) (*8*−*11*) and the Ψ_{55}-tRNA synthase (*11, 12*). Recently, we have found that the TSL is also a minimal substrate for the m^1A_{58}-tRNA methyltransferase (RAMT), the enzyme responsible for methylation of A_{58} (unpublished data). Elements of the TSL primary structure were identified as determinants for the *E. coli* (*12*) and *S. cerevisiae* (*11*) Ψ_{55}-tRNA synthase. While there appear to be some primary sequence elements responsible for the functioning of the m^5U_{54}-tRNA methyltransferase in yeast (*11*) and in *E. coli* (*10*), it is probable this enzyme primarily recognizes the three-dimensional structure of tRNA's TΨC stem−loop domain. Results from our recent studies of the *E. coli* RUMT and the *T. pyriformis* RAMT demonstrate that the enzymes require conformational constraints on their substrates for efficient function (unpublished data).

Thus, results from studies of modifying-enzyme recognition of substrate TSL raise several issues regarding the structure of the isolated TΨC stem−loop, the influence of modified nucleosides on the structure, and the relationship of any intermediate structure(s) during maturation to its final conformation in the mature, folded tRNA. The relatively small size of the heptadecamer TSL makes it amenable to both chemical synthesis and NMR spectroscopy. The fact that it can clearly be considered an independent domain for enzyme recognition also makes it a useful model for our studies on the effect of nucleoside modifications on stability and structure. We present here the solution structure of an analogue of the yeast $tRNA^{Phe}$ TΨC stem−loop containing T_{54} as the only modification (TSL-T_{54})(Figure 1). Using standard NMR techniques in concert with site-specific labels and nucleoside substitutions, we were able to assign most of the proton resonances. Our results reveal a hairpin with an unusual conformation: the locally structured loop folds over to make contact with the curved stem. While surprising, this structure bears considerable resemblance to that of the TΨC stem−loop in the crystal structure of the mature tRNA.

MATERIALS AND METHODS

Sample Preparation. The oligoribonucleotide 5′ CUGUGTUCGAUCCACAG, corresponding to the yeast $tRNA^{Phe}$ TΨC stem and loop domain with T_{54} (TSL-T_{54}), and related sequences used in this study were chemically synthesized on an Applied Biosystems 394 DNA/RNA Synthesizer (*13*). The ^{15}N-labeled phosphoramidites were prepared in our laboratory (*14*−*16*). The other nucleoside phosphoramidites were obtained from commercial sources: T and m^5C phosphoramidites from ChemGenes Corp. (Waltham, MA); and A, U, C, G, and dI from Glen Research (Sterling, VA). TSLs were HPLC-purified as previously described (*17*) using a Nucleogen 60-7 DEAE (250 × 10 mm) column; desalting was accomplished with Waters Corp. Sep-pak columns. The samples were prepared for NMR spectroscopy by extensive dialysis with buffer (10 mM sodium phosphate, pH 6.0, 0.1 mM EDTA, 94% H_2O/6% D_2O) using Amicon Centricon 3 concentrators. The solvent was exchanged to D_2O by repeatedly drying the sample in the NMR tube with a gentle stream of N_2 (g), and then resuspending the sample in 99.96% D_2O. The samples used for structural analysis had a final RNA concentration of ∼1.6 mM.

NMR Spectroscopy. All NMR spectra were collected on a Bruker DRX500 spectrometer and processed using either XWINNMR (Bruker Inc., Rheinstetten, Germany) or FELIX (Biosym/MSI, San Diego, CA). Solvent suppression for samples in 94% H_2O/6% D_2O was accomplished using either 11 spin−echo (*18*) or WATERGATE (*19*) pulse sequences. The residual water peak in D_2O samples was suppressed using low-power presaturation.

Exchangeable proton resonances of the TSL-T_{54} sample were observed and analyzed using two-dimensional NOESY spectra (*20, 21*) obtained at 1 and 10 °C with $\tau_m = 150$ ms. Spectra were acquired with sweep widths of 12 019 Hz in both dimensions, 1024 points in t_2, and a minimum of 300 points in t_1 with 96−128 scans per increment. A ROESY

experiment (*22*, *23*) also was performed at 1 °C. One-dimensional spectra were obtained as a function of temperature (1−50 °C).

NOESY spectra of the sample in D_2O were collected at 10, 15, 20, and 25 °C with mixing times varying from 50 to 400 ms. At least 256 points were collected in t_1, with 32−64 scans per increment. For these and the following homonuclear experiments (except where noted), the spectral width was 5000 Hz in both dimensions, and 1024 points were collected in t_2. TOCSY (*24*) or clean TOCSY (*25*) experiments using the MLEV17 mixing sequence were performed at 15, 20, and 25 °C with mixing times of 74−149 ms. DQF-COSY (*26*) and double quantum (DQ) (*27*, *28*) experiments were obtained at 25 °C. The two-dimensional heteronuclear spectra that were collected included a ^{1}H−^{31}P HETCOR (*29*), a ^{1}H−^{31}P heteroTOCSY (*30*), and a heteroTOCSY-NOESY (*31*). These spectra were all acquired at 25 °C. Natural-abundance ^{1}H−^{13}C HSQC experiments (*32*−*34*) were performed at both 10 and 25 °C. For some samples with site-specific ^{15}N labels, ^{1}H−^{15}N HSQC spectra were obtained in H_2O at 1 °C.

Data were obtained for structure calculations from a series of phase-sensitive NOESY spectra that were collected at 25 °C with mixing times of 50, 75, 100, 200, 300, and 400 ms. The sample was not removed from the magnet during the process of collecting these spectra. The spectra were acquired with spectral widths of 5000 Hz in both dimensions, 1024 points in t_2, 360 points with 64 scans per block in t_1, and a recycle delay of 2.5 s. Using FELIX, the spectra were processed with 60° phase-shifted sine bell apodization functions and third degree polynomial base line flattening in both dimensions. The base line in t_2 was further treated using the FLATT algorithm (*35*). To provide suitable digital resolution for cross-peak integration, the spectra were zero-filled to 2048 by 2048 points.

Structure Determination. Distance restraints between nonexchangeable protons were obtained from the NOESY spectra of the mixing time study. The spectra were processed and integrated using FELIX and were calibrated by setting pyrimidine H5−H6 cross-peaks to a distance of 2.44 Å. All lower bounds were set to 1.8 Å; upper bounds were set to the calculated distance +20%. In the case of methyl protons, an additional 1.5 Å was added to the upper bounds. The distance restraints involving exchangeable resonances were obtained from NOESY spectra of the sample in 94% H_2O/6% D_2O. The cross-peaks were qualitatively classified as strong (1.8−3.5 Å), medium (1.8−4.5 Å), or weak (1.8−5.5 Å).

A high-resolution DQFC spectrum with ^{31}P decoupling in both dimensions was collected in order to characterize the sugar conformations. The spectrum was acquired with a spectral width of 2500 Hz in both dimensions, 2048 points in t_2, and 360 points in t_1 with 64 scans per block. Residues with no observable H1′−H2′ coupling were constrained to the C3′-*endo* conformation. Those sugars with coupling constants >7 Hz were constrained to the C2′-*endo* range. Those residues with either intermediate coupling constants (i.e., <7 Hz but >3 Hz) or whose H1′−H2′ cross-peaks were too close to the diagonal were left unconstrained. The glycosidic torsion angle χ was loosely constrained for most residues on the basis of the intensity of their intraresidue H1′-aromatic NOESY cross-peak. The α and ξ torsion angles were loosely constrained to exclude the trans conformation for those residues whose ^{31}P chemical shifts fell within the narrow range commonly seen for regular A-form structures (*36*).

Structure calculations were performed using X-PLOR version 3.851 (*37*), and were based on published protocols (*36*−*38*) with minor modifications. Initially, 50 starting structures were obtained by metric matrix distance geometry embedding of all atoms (dg_full_embed.inp). These structures were regularized by subjecting them to the dgsa.inp protocol that involves template fitting and simulated annealing. To obtain the global fold of the TSL, NOE and hydrogen-bonding restraints were used, but experimental dihedral angle restraints were omitted (*36*, *38*). However, to enforce proper chirality, dihedral angle restraints involving the four ligands of each chiral center were introduced. As previously described (*39*), to obtain a higher percentage of converging structures, the bond angle potential was increased 8-fold while the NOE ceiling was decreased by a factor of 10 from the published X-PLOR values. During the simulated annealing procedure, the structures were heated to 2000 K for 15 ps, and then cooled to 100 K over 22 ps with a 0.5 fs time-step. These folded molecules were then further refined (refine.inp), with the modified parameters returned to their standard values. All experimental dihedral angle constraints were introduced at this stage. During the second round of simulated annealing, the structures from the previous step were cooled from 2000 to 100 K over 40 ps with a time-step of 0.5 fs. Finally, a 1000-step minimization using the full van der Waals term was performed. The chirality of the final structures was analyzed with chiralcheck.awk (*40*). Helix parameters were calculated using Curves 5.2 (*41*).

RESULTS

Assignment of Exchangeable Resonances. The two-dimensional NOESY spectrum of the sample in H_2O revealed that the TSL was stable and had a single predominant conformation (Figure 2). Sequential imino−imino connectivities for the stem were observed, indicating that the RNA adopted a folded structure. As is often the case, however, due to fraying these sequential connectivities did not extend to the terminal base pair, C_{49}-G_{65}. While most of the spectrum could be assigned using standard protocols, site-specific substitutions or ^{15}N labels were used for confirmation or to identify proton resonances of loop residues. For example, we were able to positively identify the imino protons of loop residues U_{55} and U_{59} and to confirm the assignments of the stem imino proton resonances using samples which were site-specifically ^{15}N-labeled (data not shown).

Cross-peaks observed in the imino to imino, and imino to amino, aromatic, and H1′ regions of the spectrum were those expected for right-handed, helical nucleic acids (*42*, *43*). In contrast, cross-peaks (indicated in Figure 2) involving the loop nucleoside C_{60} amino protons and residues from the distal region of the stem are quite unusual. The C_{60}H5, U_{50}-H5, U_{50}H6, and G_{51}H8 assignments were initially proposed based on the comparison of NOESY spectra in H_2O and D_2O at 10 °C. Since these cross-peaks suggested an unexpected and unusual fold of the molecule, additional evidence for these assignments was sought. Therefore, two sequences with site-specific nucleoside substitutions were synthesized and

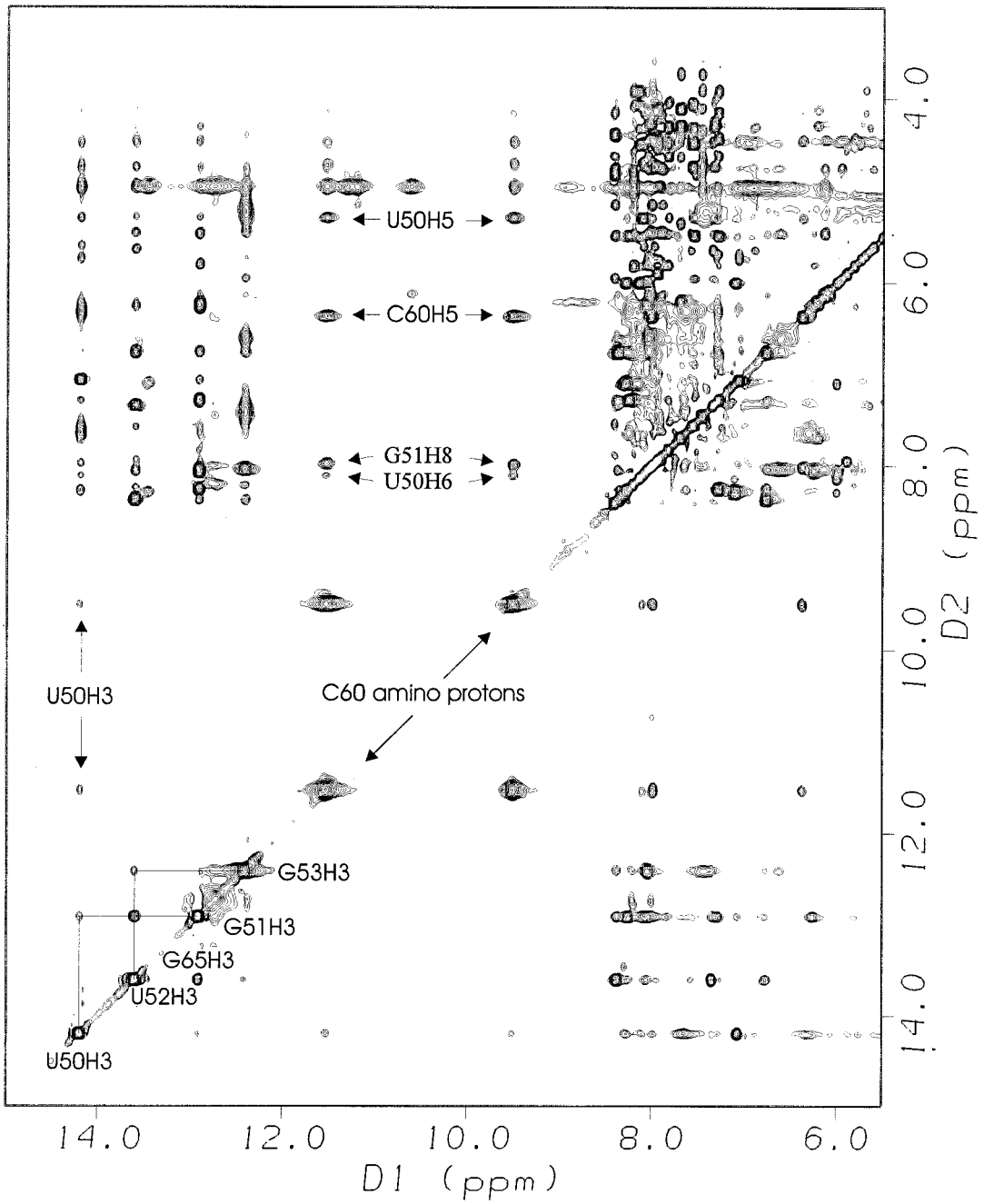

FIGURE 2: Portion of a two-dimensional NOESY spectrum of the TSL-T_{54} in H_2O at 1 °C. The sequential imino−imino connectivities of the stem are traced in the lower left corner. The NOE cross-peaks between protons of the loop residue C_{60} and those from stem residues are indicated.

studied. A sample of identical sequence, but with U_{50} replaced by T_{50} (TSL-$T_{50}T_{54}$), confirmed the original assignment of U_{50}H5; i.e., the cross-peaks attributed to U_{50}H5 were replaced by T_{50} methyl cross-peaks (Supporting Information). Likewise, a sample with m^5C_{60} replacing C_{60} (TSL-$T_{54}m^5C_{60}$) confirmed the assignment of C_{60}H5 (Supporting Information). The critical assignments of the C_{60} amino protons were unambiguously identified with a sample in which the amino group of C_{60} was ^{15}N-labeled (TSL-$T_{54}{}^{15}$N-C_{60}). In the spectra of this sample, the C_{60} amino proton resonances were clearly split by coupling to the ^{15}N-exocylic amino nitrogen (Figure 3 and Supporting Information). Several lines of evidence indicated that the C_{60} amino protons are involved in hydrogen-bonding interactions. For example, the ROESY spectrum revealed that the strong cross-peak between the two C_{60} amino protons was an NOE and not an exchange cross-peak (data not shown). Furthermore, a set of one-dimensional spectra taken at temperatures between 1 and 50 °C demonstrated that these peaks—like those of the hydrogen-bonded imino protons of the stem—persisted to high temperature, unlike the imino proton peaks of the loop U_{55} and U_{59} residues (Figure 3). As has been observed in DNA triplexes, the downfield-shifted positions of the C_{60} amino proton peaks indicated that C_{60} was protonated (*44*). Protonation of C_{60} was confirmed by the ^{15}N chemical shift of the C_{60} amino nitrogen (*45*) as determined from the ^{1}H−^{15}N HSQC of the TSL-$T_{54}{}^{15}$N-C_{60} [91.8 ppm using $^{15}NH_4Cl$ in 10% HCl as an external reference; following the recent joint IUPAC−IUBMB−IUPAB recommendations (*46*) concerning the referencing of heteronuclei, this corresponds to a

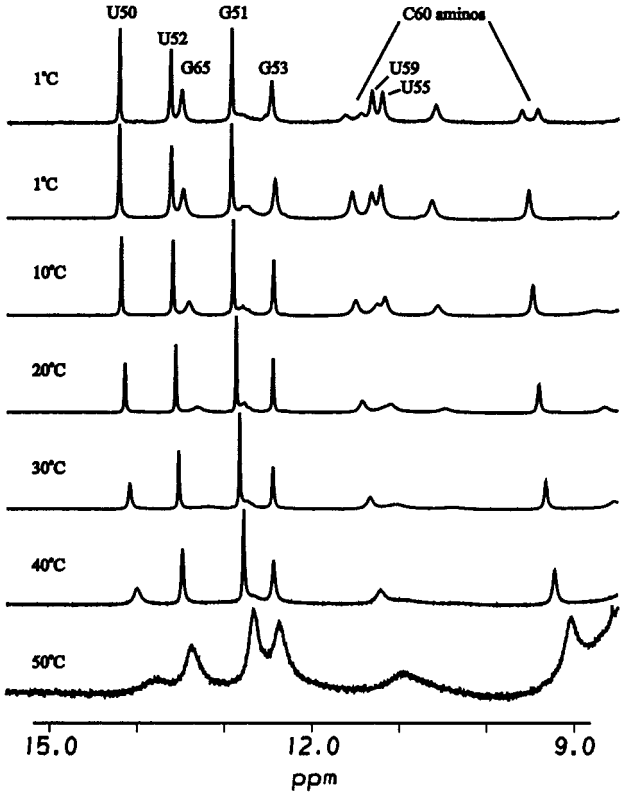

FIGURE 3: Temperature study of the TSL-T_{54}. One-dimensional spectra of the TSL-T_{54} in H_2O were collected at temperatures between 1 and 50 °C. The portion of the spectra shown includes the hydrogen-bonded imino protons of the stem and some of the exchangeable resonances of the loop. At the top is the coupled spectrum of the TSL-T_{54}^{15}N-C_{60} at 1 °C, showing the splitting of the C_{60} amino protons due to the ^{15}N label.

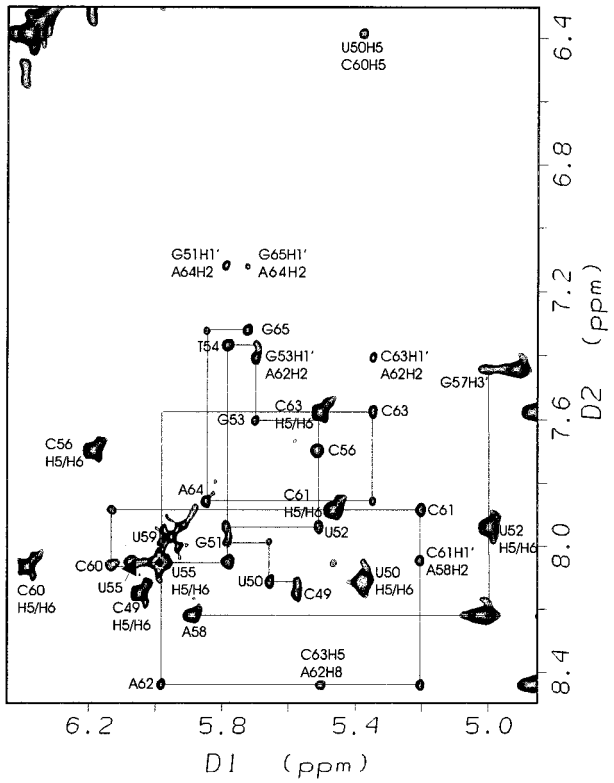

FIGURE 4: H1′ and H5 to aromatic proton region of a NOESY spectrum of the TSL-T_{54}. The NOESY spectrum was acquired at 25 °C with the TSL in D_2O. The sequential H1′−aromatic connectivities are indicated. Near the top of the figure is the cross-peak between U_{50}H5 and C_{60}H5.

chemical shift of 113.8 ppm]. Cross-peaks of the C_{60} amino protons had essentially the same ^{1}H and ^{15}N chemical shifts at pH 5.0 as at pH 6.0, but disappeared at pH 7.0. Although it made little difference to the final structure determination, a protonated C_{60} was used in the calculations for the structures presented here.

Assignment of Nonexchangeable Resonances. Assignment of the nonexchangeable resonances relied on a variety of homo- and heteronuclear experiments. The presence of T_{54} in our sequence provided a convenient starting point for assignments. The T_{54} H6−T_{54} methyl cross-peak was observed in both DQFC and TOCSY experiments, which also readily identified the pyrimidine H5−H6 cross-peaks. Particularly helpful were natural-abundance ^{1}H−^{13}C HSQC spectra which not only differentiated the CH5 from the UH5 resonances and indicated the H1′ protons, but also distinguished the AH2 from the purine H8 and pyrimidine H6 resonances. Comparison to NOESY spectra of the sample in H_2O enabled us to assign the AH2 resonances. Both stem AH2 protons had NOE cross-peaks to the H1′ proton of their 3′ neighbor, as well as an interstrand cross-peak to the H1′ proton of the 5′-adjacent base pair. These were helpful guideposts in the assignment procedure, and sequential aromatic−H1′ connectivities could be traced for most of the molecule (see below).

Cross-peaks between H1′ and H2′ protons for G_{65} and the loop residues (excluding G_{57}) were observed in DQFC and TOCSY experiments. A short mixing-time (50 ms) NOESY spectrum and a DQ spectrum were also helpful in assigning the H2′ resonances. In addition to DQ, DQFC, and TOCSY experiments, ^{1}H−^{31}P heteronuclear experiments were used in assigning the H3′ resonances. The H4′ and H5′, H5″ resonances that could be assigned were identified on the basis of these experiments.

Sequential aromatic−H1′ connectivities can be traced from the 5′ and the 3′ termini of the stem to the second residues of the loop (Figure 4). At those points, there are breaks in the sequential connectivities. However, a sequential connectivity between A_{58}H8 and G_{57}H3′ was present, and is indicated in Figure 4. The assignments for these two residues were confirmed using site-specific substitutions, one with dI at position 57 (TSL-T_{54}dI_{57}), and the other with dI at position 58 (TSL-T_{54}dI_{58}). In addition, C_{56} had a few interresidue cross-peaks, all to G_{57}. In contrast to these fairly typical cross-peaks, the cross-peak between C_{60}H5 of the loop and U_{50}H5 of the stem was unexpected, and indicated that this sequence has an unusual conformation. The assignment of this cross-peak was confirmed by NOESY spectra of the $T_{50}T_{54}$ and $T_{54}m^5C_{60}$ TSLs in D_2O.

Hairpin Characterization. A general problem in studies of nucleic acid hairpins is the question as to whether the sample is indeed a monomer. Given the unusual cross-peaks we observed in our NOESY spectra between the loop nucleoside C_{60} and residues from the distal portion of the stem, we felt compelled to show that our RNA was in fact a monomer under the conditions of this study. To demonstrate this, we measured the translational diffusion coefficient of our hairpin, and compared it to that of other RNA sequences of various lengths. This was accomplished with

Table 1: Structure Determination Statistics

distance restraints	278
intranucleotide NOEs	128
internucleotide NOEs	124
hydrogen-bonded stem base pairs	26
torsion angle restraints	39
refinement statistics (8 final lowest-energy structures)	
NOE violations (>0.5 Å)	0
dihedral angle violations (>5°)	0
average pairwise rmsd (all heavy atoms) (Å)	
all residues	1.97 ± 0.27
stem residues (49−53, 61−65)	0.98 ± 0.14
loop residues (54−60)	2.54 ± 0.48
average rmsd from ideal covalent geometry	
bond lengths (Å)	0.015
bond angles (deg)	5.01
impropers (deg)	0.416

the pulsed field-gradient spin−echo technique (*47, 48*). The method is performed on the NMR sample itself, and thereby obviates any complications or ambiguities of interpretation that may arise from approaches other than that by NMR methods (*49*). As expected for a monomer of this size, the translational diffusion coefficient of the heptadecamer was bracketed on one side by those measured for a hexamer and a dodecamer, and on the other side by that of a 28-mer. Our results are in approximate agreement with published values of the translational diffusion coefficients of nucleic acids of similar sizes (*49, 50*).

Further evidence for the monomeric nature of the sample resulted from analysis of spectra obtained from the sample at high dilution. Upon 5-fold dilution (i.e., from ~1.6 to ~0.3 mM), the spectrum obtained is nearly identical to that of the original sample. In particular, the similarity of chemical shifts of the C_{60} amino proton resonances at both concentrations indicates that their NOE cross-peaks with the stem residues do indeed arise from a fold-back of the loop, and not from intermolecular contacts. In addition, at neither concentration is there any evidence of a second set of cross-peaks, as would be expected if a monomer−dimer equilibrium really did exist.

NOE and Torsion Angle Restraints. A total of 278 distance constraints were used in the structure calculations. These restraints included 252 restraints from NOEs and 26 hydrogen-bonding constraints for the 5 base pairs of the stem (Table 1). Of the NOE-derived restraints, 192 were obtained from the mixing time study of the nonexchangeable protons, while 60 constraints came from spectra of exchangeable resonances. There were 124 and 128 inter- and intraresidue distance constraints, respectively. However, the long-range constraints illustrated in Figure 5 were more informative and certainly more important in determining the global fold of the molecule. In particular, the distance constraints between the loop nucleoside C_{60} and both U_{50} and G_{51} of the middle/distal region of the stem and those between the loop nucleoside A_{58} and the proximal stem base pair of C_{61} and G_{53} were critical in defining the structure of the TSL. Distance constraints between C_{60} and U_{50} or G_{51} were defined by NOEs from C_{60}H41 to G_{51}H8, U_{50}H3, U_{50}H5, and U_{50}H6; from C_{60}H42 to G_{51}H8, U_{50}H3, U_{50}H5, and U_{50}H6; and from C_{60}H5 to U_{50}H5 and U_{50}H6. Constraints between A_{58} and the base pair C_{61}-G_{53} were defined by NOEs from A_{58}-H1′ to C_{61}H42, C_{61}H5, and C_{61}H6; from A_{58}H2 to C_{61}H1′ and C_{61}H6; and from A_{58}H8 to G_{53}H1.

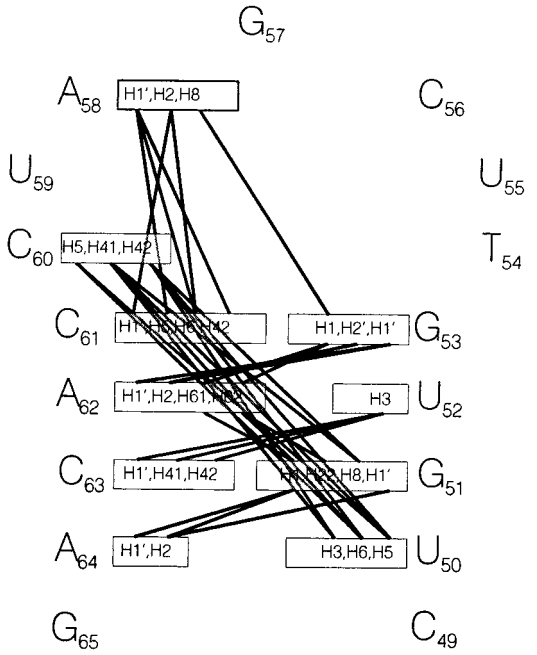

FIGURE 5: Long-range NOEs used in determination of the TSL-T_{54} structure. The primary sequence and secondary structure of the TSL-T_{54} are depicted with some of the interresidue, long-distance NOEs most important to constraining the global structure of the molecule. Intraresidue, intra-base pair, and sequential NOEs are omitted.

The δ torsion angles for 13 residues were constrained based on the analysis of a high-resolution DQFC spectrum. Residues U_{55}, C_{56}, A_{58}, and C_{60} had H1′−H2′ coupling constants greater than 7 Hz and therefore were constrained to the C2′-*endo* range. In contrast, the stem residues (except for G_{65}) had no observable H1′−H2′ cross-peak, and were constrained to the C3′-*endo* range. Residues T_{54}, G_{57}, U_{59}, and G_{65} were left unconstrained. Based on the ^{31}P chemical shifts of residues U_{50}, U_{52}, T_{54}, G_{57}, and A_{64} their corresponding α and ξ torsion angles were loosely constrained to exclude the trans conformation. Using the relative intensities of the intraresidue H1′−H6/H8 NOE cross-peaks as a criterion, the glycosidic χ torsion angle was constrained to the anti range for all residues except A_{58} and U_{59}. The intense A_{58} H1′−H8 cross-peak clearly indicated that it was syn, and in fact was found to be so in early calculations in which no torsion angle constraints were included. The intraresidue H1′−H6 cross-peak of U_{59} overlapped its H5−H6 cross-peak, and so the χ torsion angle was left unconstrained.

DISCUSSION

RNA hairpins can be problematic for NMR structural studies. The loops can be flexible and unstructured, particularly when as large as the seven-residue loop of the TSL. Even if the loop is structured and does have a single conformation, it is often difficult to determine the spatial relationship of the loop relative to the stem. We are fortunate that neither problem arose in the current study. The large number of long-range NOE cross-peaks observed between A_{58} of the loop and the G_{53}−C_{61} stem base pair indicates that the loop is at least locally structured (Figure 5). There were even more NOE cross-peaks between the loop residue C_{60} and both G_{51} and U_{50} of the stem. This leads to the inescapable conclusion that the hairpin bends in such a

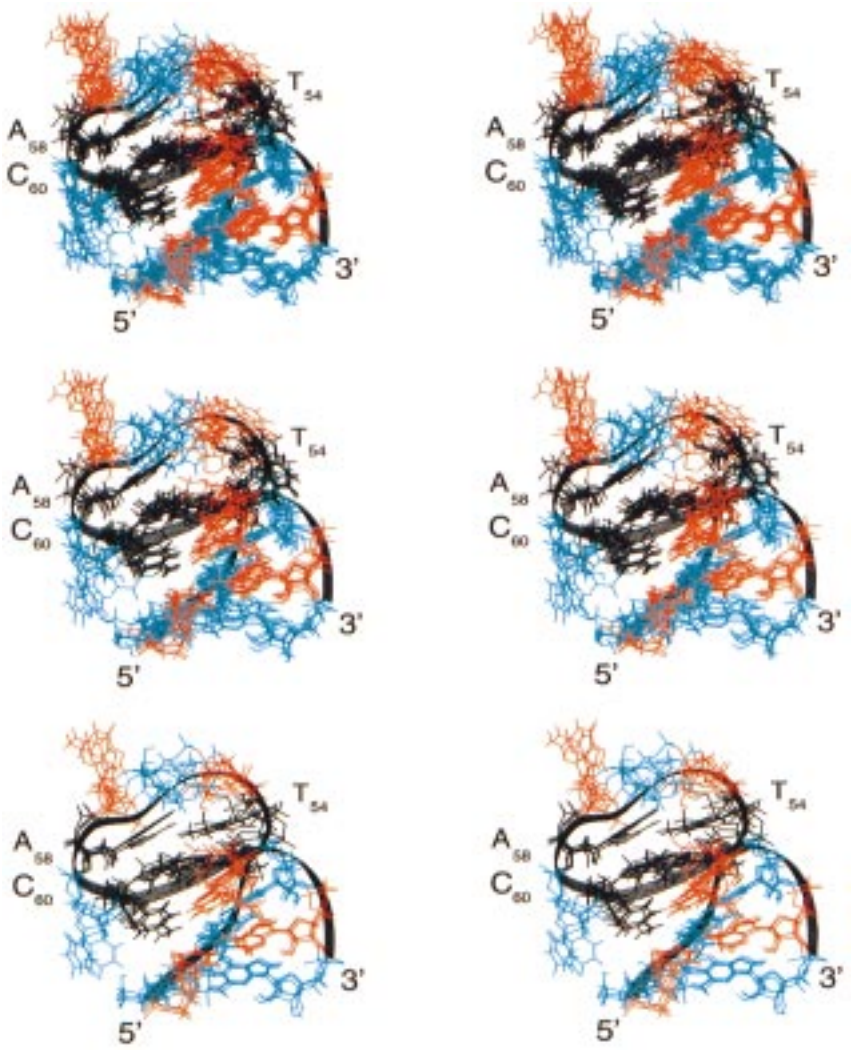

FIGURE 6: Defocused stereo representations of the superimposed, calculated TSL-T_{54} structures. The TSL-T_{54} structures are oriented in order to present an oblique view into the minor groove of the stem. The $G_{65}-C_{49}$ and $C_{63}-G_{51}$ base pairs, and residues C_{56} and U_{59} are in blue; the $A_{64}-U_{50}$ and $A_{62}-U_{52}$ base pairs, and residues U_{55} and G_{57} are in red; and the $C_{61}-G_{53}$ base pair, and residues T_{54}, A_{58}, and C_{60} are in black. (Top) Superposition of the eight lowest-energy structures. (Middle) Family A, with T_{54} in the minor groove (5 of 8 structures). (Bottom) Family B, with T_{54} in the interior of the loop (3 of 8 structures).

fashion so as to bring that portion of the loop into the major groove of the stem. Together, these two sets of cross-peaks dictate the global fold of the molecule.

A superposition of the eight lowest-energy structures is shown in Figure 6. The average pairwise rmsd for these eight structures is 1.97 ± 0.27 Å for all heavy atoms of the stem−loop. Relative to the mean structure, the non-hydrogen atom average rmsd was 1.31 Å. The precision of the stem region is higher than that of the loop; the heavy-atom average pairwise rmsd for the stem and loop are 0.98 ± 0.14 and 2.54 ± 0.48 Å, respectively. Statistics for the refinement of the structures are summarized in Table 1.

The compact appearance of the structure is fairly unusual, and could not have been anticipated *a priori*. The base-paired stem is decidedly curved; the total angle between the first and last helical segments is approximately 35 ± 7°, and results in an end-to-end shortening of the helical axis path of ∼4.1 ± 1.0%. The loop folds over to position C_{60} in the major groove midway down the stem, well below the $G_{53}-C_{61}$ base pair adjacent to the loop. This fold also draws U_{59} into the major groove, where it extends down from the loop toward the terminus of the stem, thereby shielding C_{60} to some extent. The curvature of the stem is such that C_{60} is in the proximity of all three of the central base pairs of the stem (Figure 7). This leads to some uncertainty in determining potential hydrogen-bonding partners of C_{60}. Using criteria that the distance between hydrogen-bond donor and acceptor heteroatoms is no more than 3.5 Å and that the angle is greater than 120°, the C_{60} amino proton H41 could form hydrogen bonds with U_{52} O4, G_{51} O6, or G_{51} N7. Although the evidence indicates that C_{60} H42 is also involved in a hydrogen bond, the acceptor is not obvious; one might postulate a water-mediated hydrogen bond. C_{60} is protonated, and the resultant H3 is positioned such that G_{51} O6 is a likely hydrogen bond acceptor. Hydrogen bond formation would explain the apparently higher pK_a of C_{60} than is observed for mononucleoside cytidine (*51*). It should be noted that these possibilities arise from an analysis of the average structure, and that no hydrogen-bonding constraints involving C_{60} were introduced in the calculations.

In addition to hydrogen bonds, of course, stacking interactions are also important in determining the stability of the molecule. In all structures of all calculations, A_{58} stacks on C_{61}, which is in the final stem base pair adjacent to the loop.

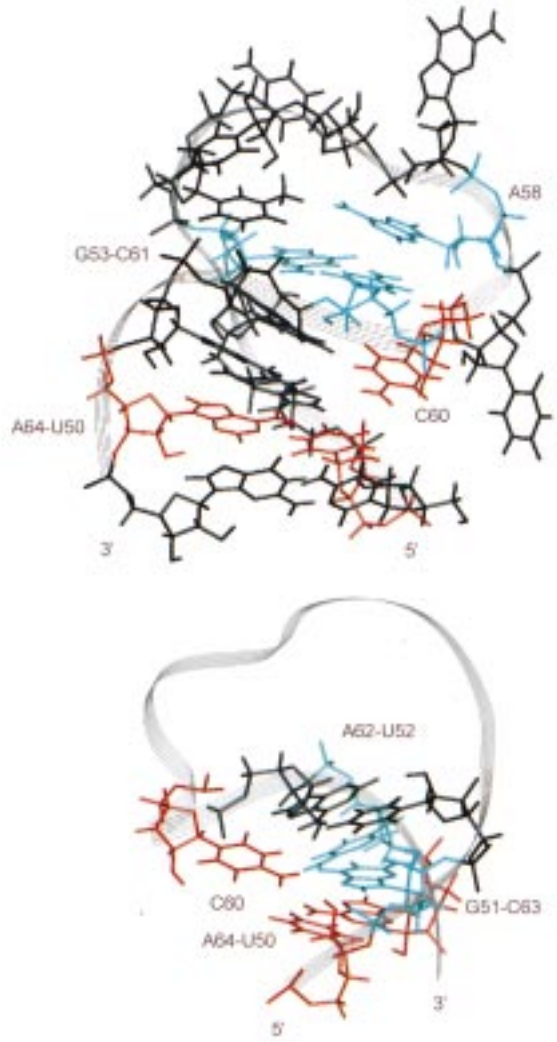

FIGURE 7: Lowest energy TSL-T_{54} structure. (Top) A view into the major groove. The $G_{53}-C_{61}$ base pair and A_{58} are in blue, and the $A_{64}-U_{50}$ base pair and C_{60} are in red. (Bottom) A view of the structure rotated with respect to the above orientation. For clarity, only C_{60} and the central three base pairs of the stem are shown. C_{60} and the $A_{64}-U_{50}$ base pair are in red; the $G_{51}-C_{63}$ base pair is in blue; and the $A_{62}-U_{52}$ base pair is in black.

The relatively protected location of A_{58} may explain our recent observation that the m^1A_{58}-tRNA methyltransferase is unable to use this particular hairpin as a substrate (unpublished data). The stacking of A_{58} on C_{61} is also seen in the crystal structure of the mature $tRNA^{Phe}$. In the native yeast $tRNA^{Phe}$, A_{58} is modified to m^1A_{58}, and participates in a reverse Hoogsteen base pair with T_{54}. However, we see little evidence of a $T_{54}-A_{58}$ base pair. In fact, the structures can be divided into two families that are distinguished by the orientation of T_{54}. In one family, T_{54} is excluded from the loop and is positioned in the minor groove of the stem (Figure 6). In the other family, T_{54} is within the loop and is approximately parallel to the stem base pairs, so that the T_{54} methyl group stacks on G_{53} (Figure 6). In the study of a related, but totally unmodified hairpin which otherwise has the same loop sequence and proximal stem base pair, Yao et al. (*52*) also saw no evidence of a $T_{54}-A_{58}$ base pair. However, their molecule appeared to have a very dynamic loop, as evidenced by the fact that A_{58} H2 had NOE cross-peaks to five different H1′ protons. In a subsequent paper (*53*), they reanalyzed their data and were able to reduce their large time-averaged ensemble to one major and five or six minor conformational families. It is interesting that for all members resulting from this latter analysis, C_{60} slides into

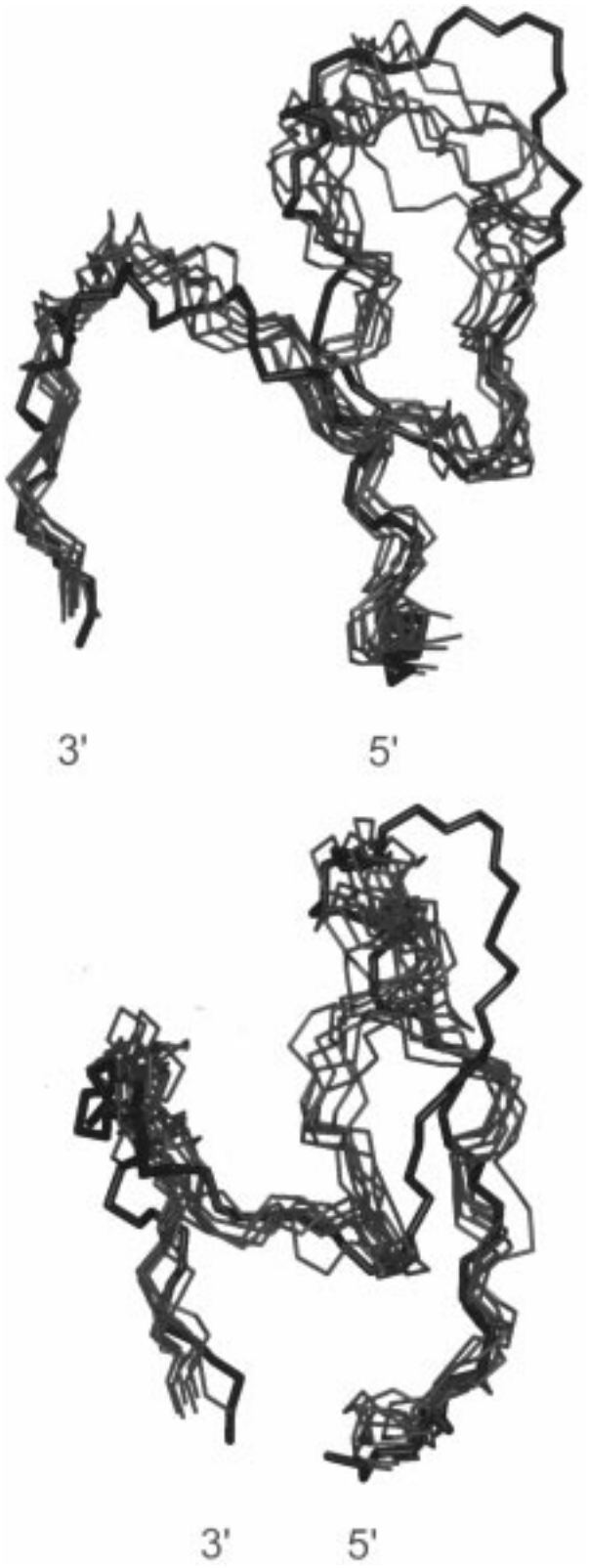

FIGURE 8: Comparison of the solution structure of the TSL-T_{54} and the analogous region from the crystal structure of yeast $tRNA^{Phe}$ (*56*). The eight lowest-energy structures of the TSL-T_{54} calculated in this study are shown in gray. The crystal structure in black is superimposed on the solution structures. (Top) A view of the comparison looking into the major groove. (Bottom) A view of the comparison rotated 35° with respect to the above orientation.

the major groove to some degree, although certainly not to the extent that we observe with our sequence. In our structures, G_{57} is directed away from the loop and is extended into the solvent. C_{56} and U_{55}, both of which form base pairs with residues of the D stem−loop in the mature wild-type tRNA^{Phe}, are fairly disordered in our structures.

The striking contrast between the results we report here and those obtained by Yao et al. could arise from the differences in pH employed in the two studies, i.e., theirs at pH 6.4, ours at pH 6.0. It is conceivable that this slight difference is significant, perhaps by affecting the protonation state of C_{60}. However, apparently this is not the case; at pH 6.4, we still observe the crucial, diagnostic cross-peaks between the C_{60} amino protons and both G_{51} and U_{50} of the stem (Supporting Information). It would appear, therefore, that the decisive factor is instead the particular TΨC stem and loop sequence they investigated. It is probably significant that their sequence was totally unmodified, and/or that it does not correspond to any naturally occurring TΨC domain. The sequence they employed in the NMR structural studies is not the Val TSL used earlier (*8*, *10*), and is not found in the tRNA database (*5*), nor in the *E. coli* or yeast (*S. cerevisiae*) databases (*54*, *55*). Thus, it might be of interest to study other, authentic sequences in order to assess the generality of our results.

When our solution structure is compared to the TΨC stem−loop portion of the crystal structure of the mature tRNA, a strong resemblance is apparent (Figure 8). The distinguishing geometric feature of the isolated domain from the crystal structure is the marked folding of the loop region into the major groove of the stem. This is mimicked to a large extent by our solution structure. The greatest divergence in the superposition of the backbones appears to be in the region of the loop from C56 to U59. In the intact tRNA, intrusive interactions occur between residues of the D loop and this region of the TΨC loop. G_{18} intercalates between m^1A_{58} and G_{57}, while G_{57} in turn intercalates between G_{18} and G_{19}, thereby forming a $m^1A_{58}-G_{18}-G_{57}-G_{19}$ base zipper. G_{18} and G_{19} are also involved in base pairs with Ψ_{55} and C_{56}, respectively. In light of these tertiary interactions which are necessarily absent in our construct, the observed similarity is all the more remarkable. Furthermore, we believe that it is highly unlikely that this similarity is coincidental. Rather, in agreement with a hierarchical view of RNA folding, we conclude that the TΨC stem−loop is largely preassembled, and that our structure represents an intermediate in the folding pathway of tRNA.

ACKNOWLEDGMENT

We thank Winnell Newman and Guihua Liu of the NCSU Nucleic Acids Facility for the synthesis and purification of the RNAs used in this study. We also thank Robert J. Cain, Hanna Sierzputowska-Gracz, and Salman Ashraf of NCSU; Peter Schultze of UCLA; and Thorsten Dieckmann of UCD for their generous advice and assistance.

SUPPORTING INFORMATION AVAILABLE

Four figures of spectra of the TSL-T_{54} with ^{15}N-amino-labeled C_{60} (TSL-$T_{54}{}^{15}$N-C_{60}), methylated U_{50} or C_{60} (TSL-$T_{50}T_{54}$ and TSL-$T_{54}m^5C_{60}$, respectively), and TSL-T_{54} at pH 6.4; and two figures of graphs of data from translational diffusion experiments in H_2O at 1 °C and in D_2O at 25 °C (6 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penwick, J. R., and Zamir, A. (1965) *Science 147*, 1462−1465.
2. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., and Rich, A. (1974) *Science 185*, 435−440.
3. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., and Klug, A. (1974) *Nature 250*, 546−551.
4. Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) *Nucleic Acids Res. 27*, 196−197.
5. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
6. Björk, G. R. (1995) in *tRNA Structure, Biosynthesis, and Function* (Söll, D., and RajBhandary, U. L., Eds.) pp 165−206, ASM Press, Washington, D.C.
7. Nishikura, L., and DeRobertis, E. M. (1981) *J. Mol. Biol. 154*, 405−420.
8. Gu, X. R., and Santi, D. V. (1991) *Biochemistry 30*, 2999−3002.
9. Guenther, R. H., Bakal, R. S., Forrest, B., Chen, Y., Sengupta, R., Nawrot, B., Sochacka, E., Jankowska, J., Kraszewski, A., Malkiewicz, A., and Agris, P. F. (1994) *Biochimie 76*, 1143−1151.
10. Gu, X., Ivanetich, K. M., and Santi, D. V. (1996) *Biochemistry 35*, 11652−11659.
11. Becker, H. F., Motorin, Y., Sissler, M., Florentz, C., and Grosjean, H. (1997) *J. Mol. Biol. 274*, 505−518.
12. Gu, X., Yu, M., Ivanetich, K. M., and Santi, D. V. (1998) *Biochemistry 37*, 339−343.
13. Usman, N., Ogilvie, K. K., Jaing, M. Y., and Cedergren, R. J. (1987) *J. Am. Chem. Soc. 109*, 7845−7854.
14. Kamaike, K., Takahashi, M., Utsugi, K., Tomizuka, K., Okazaki, Y., Tamada, Y., Kinoshita, K., Masuda, H., and Ishido, Y. (1996) *Nucleosides Nucleotides 15*, 749−769.
15. Ti, G. S., Gaffney, B. L., and Jones, R. A. (1982) *J. Am. Chem. Soc. 104*, 1316.
16. Wu, T., Ogilvie, K. K., and Pon, R. T. (1989) *Nucleic Acids Res. 17*, 3501−3517.
17. Agris, P. F., Malkiewicz, A., Kraszewski, A., Everett, K., Nawrot, B., Sochacka, E., Jankowska, J., and Guenther, R. (1995) *Biochimie 77*, 125−134.
18. Sklenár, V., and Bax, A. (1987) *J. Magn. Reson. 74*, 469−479.
19. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J. Biomol. NMR 2*, 661−665.
20. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
21. Macura, S., and Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
22. Bothner-By, A. A., Stephens, R. L., Lee, J.-M., Warren, C. D., and Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811−813.
23. Griesinger, C., and Ernst, R. R. (1987) *J. Magn. Reson. 75*, 261−271.
24. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
25. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
26. Piantini, U., Sørensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
27. Braunschweiler, L., Bodenhausen, G., and Ernst, R. R. (1983) *Mol. Phys. 48*, 535−560.
28. Mareci, T. H., and Freeman, R. (1983) *J. Magn. Reson. 51*, 531−535.
29. Sklenár, V., Miyashiro, H., Zon, G., Miles, H. T., and Bax, A. (1986) *FEBS Lett. 208*, 94−98.
30. Kellogg, G. W. (1992) *J. Magn. Reson. 98*, 176−182.

31. Kellogg, G. W., Szewczak, A. A., and Moore, P. B. (1992) *J. Am. Chem. Soc. 114*, 2727−2728.
32. Varani, G., and Tinoco, J. I. (1991) *J. Am. Chem. Soc. 113*, 9349−9354.
33. Palmer, A. G., III, Cavanagh, J., Wright, P. E., and Rance, J. (1991) *J. Magn. Reson. 93*, 151−170.
34. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
35. Güntert, P., and Wüthrich, K. (1992) *J. Magn. Reson. 96*, 403−407.
36. Varani, G., Aboul-ela, F., and Allain, F. H.-T. (1996) *Prog. Nucl. Magn. Reson. Spectrosc. 29*, 51−127.
37. Brünger, A. T. (1992) *X-PLOR Version 3.1: A System for Crystallography and NMR*, Yale University Press, New Haven, CT.
38. Wimberly, B. T. (1992) Ph.D. Thesis, University of California.
39. Koshlap, K. M., Schultze, P., Brunar, H., Dervan, P. B., and Feigon, J. (1997) *Biochemistry 36*, 2659−2668.
40. Schultze, P., and Feigon, J. (1997) *Nature 387*, 668.
41. Lavery, R., and Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
42. Varani, G., and Tinoco, J. I. (1991) *Q. Rev. Biophys. 24*, 479−532.
43. Wijmenga, S. S., and van Buuren, B. N. M. (1998) *Prog. Nucl. Magn. Reson. Spectrosc. 32*, 287−387.
44. Feigon, J., Koshlap, K. M., and Smith, F. W. (1995) *Methods Enzymol. 261*, 225−255.
45. Leitner, D., Schroder, W., and Weisz, K. (1998) *J. Am. Chem. Soc. 120*, 7123−7124.
46. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998) *Pure Appl. Chem. 70*, 117−142.
47. Stejskal, E. O., and Tanner, J. E. (1965) *J. Chem. Phys. 42*, 288−292.
48. Tanner, J. E. (1970) *J. Chem. Phys. 52*, 2523−2526.
49. Lapham, J., Rife, J. P., Moore, P. B., and Crothers, D. M. (1997) *J. Biomol. NMR 10*, 255−262.
50. Yang, X., Sanghvi, Y. S., and Gao, X. (1997) *J. Biomol. NMR 10*, 383−388.
51. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
52. Yao, L. J., James, T. L., Kealey, J. T., Santi, D. V., and Schmitz, U. (1997) *J. Biomol. NMR 9*, 229−244.
53. Schmitz, U., Donati, A., James, T. L., Ulyanov, N. B., and Yao, L. (1998) *Biopolymers 46*, 329−342.
54. Blattner, F. R., Plunkett, G. I., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science 277*, 1453−1474.
55. Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T. Y., Schroeder, M., Weng, S., and Botstein, D. (1998) *Nucleic Acids Res. 26*, 73−79.
56. Westhof, E., and Sundaralingam, M. (1986) *Biochemistry 25*, 4868−4878.

BI990118W